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

Tsung-Ren Huang and Stephen Grossberg

Center for Adaptive Systems
Department of Cognitive and Neural 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, steve@cns.bu.edu

Corresponding Author: Stephen Grossberg

Submitted: August 12, 2009
Revised: April 24, 2010

Psychological Review, in press

***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

Technical Report CAS/CNS-TR-09-010

Abstract

How do humans use target-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, humans can learn that a certain combination of objects may define a context for a kitchen and trigger a more efficient search for a typical object, such as a sink, in that context. The ARTSCENE Search model is developed to illustrate the neural mechanisms of such memory-based context learning and guidance, and to explain challenging behavioral data on positive/negative, spatial/object, and local/distant cueing effects during visual search, as well as related neuroanatomical, neurophysiological, and neuroimaging data. 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 as a scene is scanned with saccadic eye movements. The model simulates the interactive dynamics of object and spatial contextual cueing and attention in the cortical What and Where streams starting from early visual areas through medial temporal lobe to prefrontal cortex. After learning, model dorsolateral prefrontal cortex (area 46) primes possible target locations in posterior parietal cortex based on goal-modulated percepts of spatial scene gist that are represented in parahippocampal cortex. Model ventral prefrontal cortex (area 47/12) primes possible target identities in inferior temporal cortex based on the history of viewed objects represented in perirhinal cortex.

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 (Neider & Zelinsky, 2006). 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 (Chun, 2000).

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). After 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 through 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 (Gold & Shadlen, 2007; Grossberg & Pilly, 2008; Heekeren, Marrett, & Ungerleider, 2008; Irwin, 1991; Jonides, Irwin, & Yantis, 1982). 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, gaze locations and thus eye scanning paths do not solely depend on structural statistics or embedded contexts of an external scene, but also reflect internal drives and task-dependent goals (Ballard & Hayhoe, 2009; Hayhoe & Ballard, 2005; Rothkopf, Ballard, & Hayhoe, 2007). 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. Yarbus (1967) has provided a classic 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 (Chen & Zelinsky, 2006; Leber & Egeth, 2006) or emotional primes (Armony & Dolan, 2002; Öhman, Flykt, & Esteves, 2001).

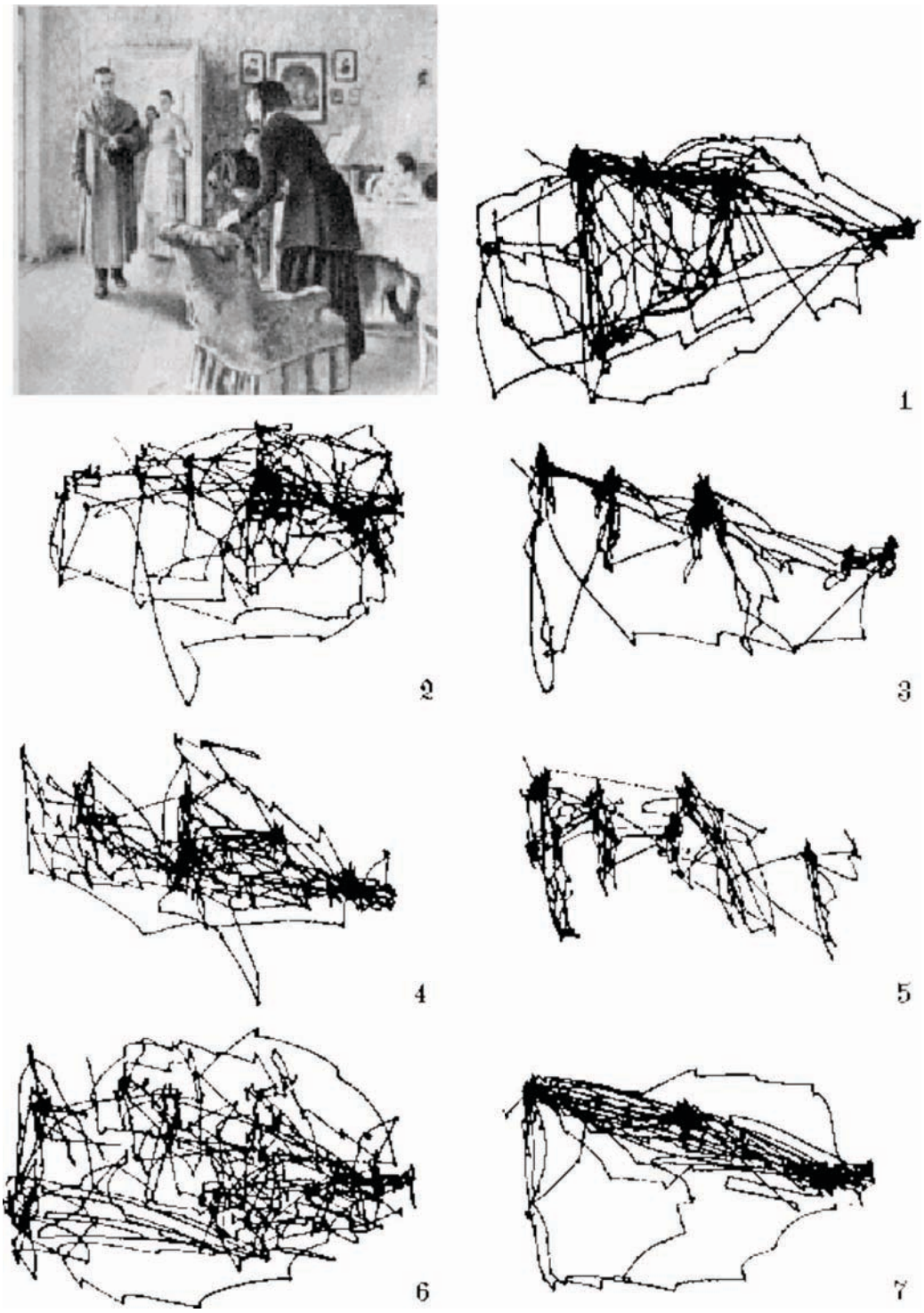


Figure 1. Eye-scan paths and fixations of an observer varied given different task instructions. Conditions in each panel are (1) free viewing, (2) estimating the wealth of the family, (3) judging their ages, (4) guessing what they had been doing before the arrival of the unexpected visitor, (5) remembering 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 scanning eye movements. 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. The ARTSCENE Search model is here developed to illustrate how global-to-local evidence accumulation, combined with learned contextual information acquired from multiple objects and positions in a scene during a sequence of attention shifts and eye movements, can quantitatively explain a large set of visual search data. In what follows, Sections 2 and 3 review psychological data and models of visual search and contextual cueing. Section 4 summarizes related brain regions and their functional roles. Section 5 provides a heuristic summary of the ARTSCENE Search model. Section 6 summarizes model explanations and simulations of contextual cueing search data. The Section 7 discusses a variety of issues, including effects of model lesions on search behavior, comparisons with other models of attention and search, and model extensions. The Appendix provides a complete mathematical description of the model.

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.

To quantify search performance in psychophysical studies, reaction time (RT), the elapsed time between presentation of a search screen and discovery of a task target, is usually recorded. 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 the set size, producing a non-zero search slope (Figure 2b).

Based on the seeming dichotomy of efficient feature search versus inefficient conjunction search, Treisman and Gelade (1980) proposed a two-stage visual attention model named Feature Integration Theory (FIT) in which primitive features are first processed within their own feature map in a rapid, pre-attentive, and parallel manner, followed by a 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 reported 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).

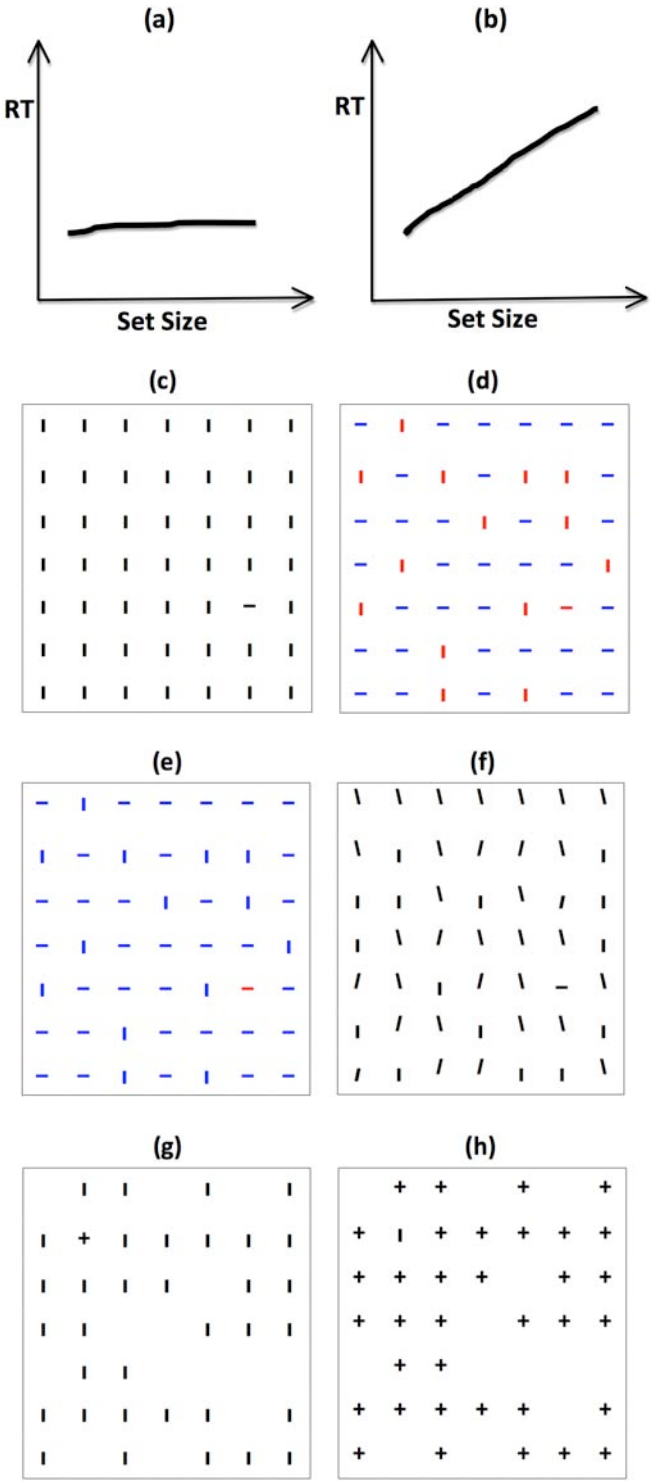


Figure 2. Summary of basic search properties: (a) Zero slope in efficient search. (b) Nonzero slope in inefficient search. (c) Efficient feature search for a horizontal bar. (d) Inefficient conjunction search for a red horizontal bar. (e) Efficient conjunction search for a red horizontal bar. (f) Inefficient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

The dichotomy of efficient versus inefficient search based on slopes was later shown to be inadequate (see Thornton & Gilden, 2007 and Townsend, 1972, for discussions of serial vs. parallel search). A continuum of flat to steep slopes can be obtained by varying saliency factors (Wolfe, 1998; Wolfe, Cave, & Franzel, 1989). In particular, search efficiency increases with decreased similarity of targets to distractors and increased similarity between distractors (Duncan & Humphreys, 1989). In other words, a conjunction search can be efficient (Figure 2e) and a feature search can be inefficient (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 one class of models that consider top-down priming, parallel enhancement of target features (Guided Search: Wolfe, 1994; Wolfe et al., 1989) or suppression of non-target features (Revised Feature Integration Theory: Treisman & Sato, 1990) were introduced to the feature integration 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 (e.g., Rao, Zelinsky, Hayhoe, & Ballard, 2002; Zelinsky, 2008) or endogenously acquired from scene memory (Contextual Cueing: Chun & Jiang, 1998; Chun & Jiang, 1999).

In another class of models that are more stimulus-driven (e.g., Attentional Engagement Theory: Duncan & Humphreys, 1989; SEearch via Recursive Rejection: Humphreys & Müller, 1993; Spatial and Object Search: Grossberg, Mingolla, & Ross, 1994; COntour DEtector Theory of Visual Attention: 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 through horizontal connections within each feature map as part of parallel processing in Feature Integration Theory. In the Spatial and Object Search model of Grossberg et al. (1994), bottom-up grouping and surface color factors set the stage for the serial allocation of top-down spatial and object attention.

A parallel line of development of the feature integration framework concerns how spatial selection and attention shifts can be carried out with plausible brain mechanisms. Koch and 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 (see also Bundesen, Habekost, & Kyllingsbaek, 2005). 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 whereby attention can disengage from the winner location and continue a new selection cycle (Posner & Cohen, 1984; see also Grossberg, 1978 and Grossberg & Kuperstein, 1986 for examples of saliency choice and inhibition of return, and reviews in Klein, 2000). As a result, a saliency map is scanned in order of decreasing saliency by the focus of attention.

Computational refinements of the saliency model (e.g., Itti & Koch, 2000; Niebur & Koch, 1996) were applied to simple laboratory stimuli as well as to 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 reaction time asymmetries (Li, 2002; Treisman & Gormican, 1988) 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 et al. (1994) proposed how spatial attention and object attention interact with visual boundary and surface representations to direct visual search. In a related approach, Navalpakkam and Itti (2005) proposed that the learned feature memory of a target can be used for attentional biasing and object recognition. Torralba et al. (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, and Humphreys (2005) used an associative memory theory/model to explain basic spatial cueing effects reported by Chun and Jiang (1998). Brady and Chun (2007) simulated the locality of spatial cueing effects (Olson & Chun, 2002) by exponentially weighting input patterns surrounding a target.

None of these models provides a unified framework for memory-based evidence accumulation that incrementally integrates all available spatial and object constraints to limit the search space. Specifically, eye fixations in most search models merely function for target checking. As a result, eye movements are a series of false-target rejections (e.g., Itti & Koch, 2000; Zelinsky, 2008). In contrast, eye fixations in ARTSCENE Search function for target checking *plus* information gathering about the target: Eye movements permit the process of evidence accumulation about the target location and identity. Moreover, unlike purely psychological 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 and spatial-to-object 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 recognized identities of context objects after each eye fixation. In ARTSCENE Search, a contextual cue 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 properties 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 a stove and microwave oven. In this scenario, object contextual cueing (Chun & Jiang, 1999), or the knowledge about the correlations among object identities in a scene, is illustrated by the expectation that a refrigerator, more than a toilet, will be seen after viewing a stove and microwave oven.

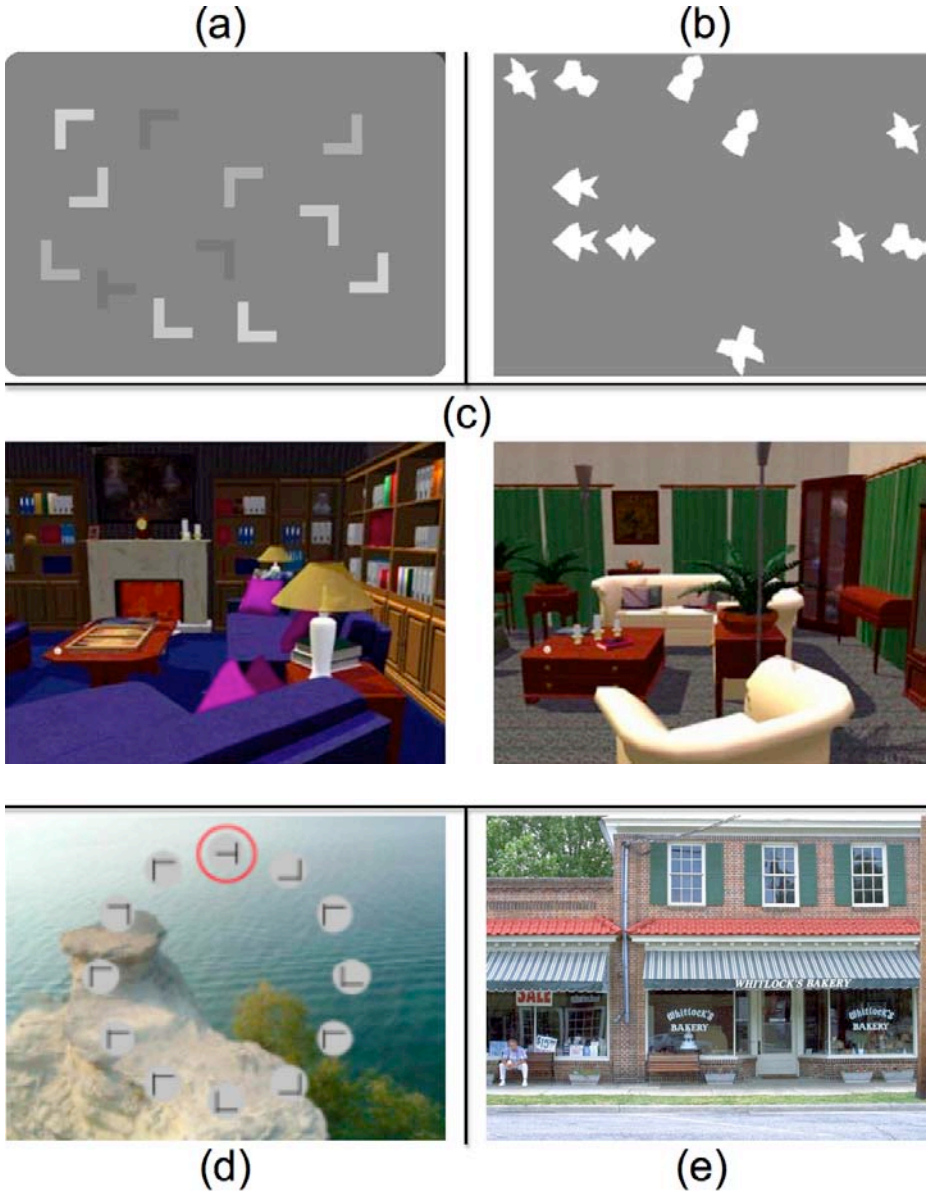


Figure 3. (a) Stimuli used by Chun and Jiang (1998), Olson and Chun (2002), Jiang and Wagner (2004), Lleras and von Mühlenen (2004), and Brady and Chun (2007). Observers searched for 'T' among 'L's. (b) Stimuli used by Chun and Jiang (1999). Observers searched for a shape symmetric around the vertical axis. (c) Stimuli used by Brockmole et al. (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).

Psychophysically, contextual cueing effects are defined as the reaction time difference in visual search between familiar and novel scenes. Compared with semantic object cueing, spatial cueing has been 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). Recently, spatial cueing effects have also been shown with photographic scenes (e.g., Brockmole & Henderson, 2006; Torralba et al., 2006) or naturalistic scenes (e.g., Brockmole, Castelhano, & Henderson, 2006). Although learning of repeated contextual cues is often reported to be implicit, without subjects' awareness of the target-cue co-variation 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 contexts persists for at least one week once acquired (Chun & Jiang, 2003; Jiang, Song, & Rigas, 2005).

After a spatial context is rapidly apprehended as the spatial gist of a scene, it guides further allocation of focal attention. Such spatial cueing can occur in 200ms (Chun & Jiang, 1998, Experiment 5), which is only long enough to accommodate one fixation in visual search (180-275ms) or scene perception (260-330ms) (see Rayner, 2009 and Nuthmann, Smith, Engbert, & Henderson, 2010, for discussions on fixation durations). In other words, a single glance at 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 (Tseng & Li, 2004), which may reflect facilitation by a learned context in spatial selection for target search in a repeated display (see also Greene, 2008, for discussions of ineffective and effective search phases).

With regard to attention deployment during search, data from Kunar, Flusberg, Horowitz, and 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 reaction time 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 contextual cues.

In terms of memory representation, a spatial context, acting as the spatial gist of a scene, is not necessarily encoded and retrieved as a whole. 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 the same 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-cue pairs, which sometimes differ from the correlation between a target and its co-varied contextual cues as a whole. In fact, training with crowded search displays can diminish cueing effects (Hodsoll & Humphreys, 2005) because a decreasing target-cue correlation may occur with an increasing probability of a distractor location being recycled in various spatial contexts where it is paired with different target locations. Similarly, small cueing effects that result when a spatial context primes more than one target location (Chun & Jiang, 1998, Experiment 3) can be intensified by

more training to strengthen correlations between fixed target-cue locations (Chun & Jiang, 1998, Experiment 6). Finally, a target-cue 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 and Watanabe (2007) 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 a spatial context is learned before a target is fixated, because the short-term memory 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.

Are spatial contexts in long-term memory encoded in egocentric (viewer-centered), spatiotopic (world-centered), or retinotopic (eye-centered) coordinates? If spatial contexts were registered retinotopically with respect to a specific gaze position, such as the target location (Jiang & Wagner, 2004; Olson & Chun, 2002; van Asselen & Castelo-Branco, 2009), then memory retrieval of encoded spatial contexts is likely to be impaired when gaze position frequently moves from one to another location. However, spatial cueing does successfully occur even when accompanied by eye movements during target search (Tseng & Li, 2004). Moreover, spatial cueing effects disappear when spatial contexts are spatially invariant in retinotopic coordinates (with respect to a foveated target location) but variant in egocentric/spatiotopic space. For example, the spatial cueing effects induced by a fixed local configuration surrounding a target in a quadrant of a search display were eliminated when the target quadrant was randomly moved to a different quadrant in the same display from repetition to repetition (Brady & Chun, 2007). This evidence is more in line with egocentric/spatiotopic than retinotopic encoding of spatial contexts, but does not rule out the possibility of a mix of reference frames underlying spatial cueing.

Aside from statistical regularities in external search displays, internal factors such as attention also regulate memory encoding and retrieval of spatial contexts. Jiang and 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 and von Mühlenen (2004) used the spatial cueing paradigm with two kinds of task instructions to bias search strategies. In their between-subjects design, one group of participants was asked to be receptive and intuitive during search, and 80% of the subjects showed context-induced *decreases* in reaction time. In contrast, the other group was asked to be active and deliberate during search, and 65% of the participants showed context-induced *increases* in reaction time. It is curious why negative cueing effects, or increased search reaction time, arose from an informative spatial context. In any case, attention modulates the efficacies of contributions from equally target-predictive context locations, and may play a role in making a local context more effective than a global/distant context for predicting target locations (Brady & Chun, 2007; Olson & Chun, 2002) or vice versa (Brockmole et al., 2006). In summary, target-cue associations and attentional modulations 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 from a visual scene directly attract saliency-based attention; 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.

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 (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). 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 parvocelluar pathway (Bar et al., 2006; Kveraga, Boshyan, & Bar, 2007). The magnocellular pathway is thus likely to extract aspects of scene gist and trigger top-down priming for object recognition or visual search, whereas the parvocellular pathway is better suited for processing detailed featural information in an object or a scene. It should be noted, however, that parvo- and magno-cellular innervations are predominant but not exclusive pathways in the ventral and dorsal streams, respectively. Each of the ventral and dorsal streams receives a mixture of inputs from both the parvo and magno pathways (for a review, see Goodale & Milner, 1992).

Top-down priming occurs in both cortical streams, and enhances effective contrast of an attended stimulus (Carrasco, Penpeci-Talgar, & Eckstein, 2000; Grossberg, 1980, 1999; Reynolds & Chelazzi, 2004). For the ventral What stream, studies of color-based attention found attentional modulations not only in inferotemporal (IT) cortex and V4, but also in early visual areas including lateral geniculate nucleus (LGN), V1, V2 and V3 (Grossberg & Mingolla, 1985; Müller et al., 2006; Saenz, Buracas, & Boynton, 2002; Sillito, Jones, Gerstein, & West, 1994). The IT cortex, among other areas, can drive such top-down priming 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 the frontoparietal attention network (Egner et al., 2008). Lateral intraparietal area (LIP) is believed to represent a feature-sensitive saliency map and to guide selection of spatial targets for saccadic eye movements (Buschman & Miller, 2007; Gottlieb, 2007; Saalmann, Pigarev, & Vidyasagar, 2007; Vidyasagar, 1999). Patients with bilateral parietal lesions lose the ability to spatially localize objects, and they perceive illusory conjunctions even with prolonged displays of only two objects (Treisman, 2006).

Prefrontal cortex (PFC) displays persistent activities in the delay period of working memory tasks (Chafee & Goldman-Rakic, 1998; Curtis & D'Esposito, 2003; Funahashi, Chafee, & Goldman-Rakic, 1993; Fuster, 1973; Fuster & Alexander, 1971; Miller, Erickson, & Desimone, 1996; Sakai, Rowe, & Passingham, 2002), and provides top-down priming in many experimental tasks (see reviews by Miller & D'Esposito, 2005). Dorsolateral PFC (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 (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 the dorsolateral PFC 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 dorsolateral PFC in goal-directed behavior (Fuster, 2008; Miller & Cohen, 2001) and frontal eye fields in visual target selection (Buschman & Miller, 2007; Schall & Thompson, 1999).

Connecting with the frontal lobe, the medial temporal lobe plays an essential role in processing contexts in a scene by which a target can be quickly defined and located. In the medial temporal lobe, the parahippocampal regions, including parahippocampal cortex and perirhinal cortex, play complementary roles (Grossberg, 2000a) in spatial and object contextual processing. Chun and Phelps (1999) reported that amnesic patients with damage in medial temporal lobe did not exhibit spatial cueing effects seen in the control group. Manns and Squire (2001) further showed that spatial cueing effects could be impaired by extensive damage to the medial temporal lobe along with variable damage to lateral temporal cortex, but not by damage confined to the hippocampal formation. Using functional MRI (fMRI), Aminoff, Gronau, and Bar (2007) examined context learning in a passive viewing paradigm, and demonstrated that the spatial-context condition elicited more activation in the posterior parahippocampal cortex (i.e., the parahippocampal place area, a.k.a. PPA) than the no-context condition, whereas the object-context condition elicited more activation in the anterior parahippocampal cortex and its adjacent perirhinal cortex than the no-context condition. Again, using fMRI, Jiang, King, Shim, and 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 perirhinal cortex is implicated in coding stimulus-stimulus associations (Murray & Richmond, 2001; Naya, Yoshida, & Miyashita, 2003; Naya, Yoshida, Takeda, Fujimichi, & Miyashita, 2003) and high-order feature conjunctions (Bussey & Saksida, 2002; Murray & Bussey, 1999).

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.

5. The ARTSCENE Search Model

5.1 Model Overview

Figure 4 provides a macrocircuit of the ARTSCENE Search model in terms of key brain areas involved in scene perception and scene memory. Structurally, the model consists of four groups of regions each of which is labeled with a different color in Figure 4. The group of feature processing regions (i.e., V1/V2 and V4/ITp) is the front-end of ARTSCENE Search in which a simulated object in a search display is represented by a set of visual features, such as colors and orientations. The group of Where stream regions (i.e., PPC, PHC, and DLPFC) regulates spatial contextual cueing, whereas the group of What stream regions (i.e., ITa, PRC, and VPFC) regulates object contextual cueing. The fourth group contains an oculomotor area (i.e., SC) responsible for generating saccadic eye movements during visual search.

The macrocircuit of ARTSCENE Search is constrained by neuroanatomical data. In particular, the early What-Where segregation is further extended into the medial temporal lobe (Eichenbaum, Yonelinas, & Ranganath, 2007) and the frontal lobe (Levy & Goldman-Rakic,

2000). In the medial temporal lobe, parahippocampal cortex is innervated by posterior parietal cortex, and perirhinal cortex is innervated by anterior inferotemporal cortex (Suzuki & Amaral, 1994). In the frontal lobe, dorsolateral PFC (area 46) is innervated by posterior parietal cortex and parahippocampal cortex, whereas ventral PFC (area 47/12) is innervated by anterior inferotemporal cortex and perirhinal cortex (Petrides, 2005). Although reciprocal connections occur between these pairs of regions in vivo, in the model reciprocal connections occur only with posterior parietal cortex and anterior inferotemporal cortex. The anatomical segregation of two visual streams is reflected by the physiological differences of brain areas involved in contextual processing. Cortical areas in the What and Where streams process, respectively, more object and spatial aspects of a scene, as in the neural data presented in Section 4 and in the model treatments presented below.

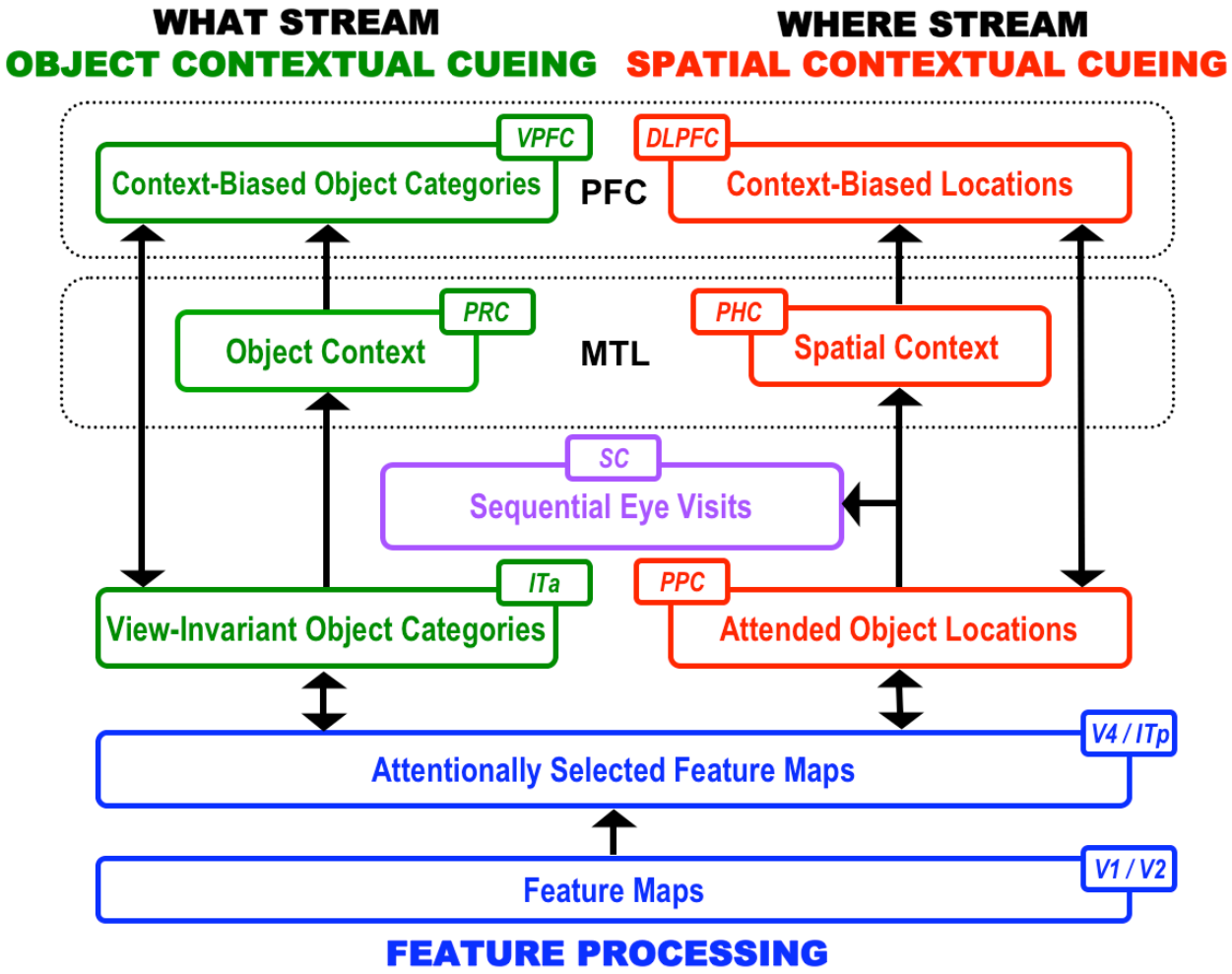


Figure 4. Macrocircuit of the ARTSCENE Search neural model for visual context processing. **V1**=First visual area or primary visual cortex; **V2**=Second visual area; **V4**=Fourth visual area; **PPC**=Posterior parietal cortex; **ITp**=Posterior inferotemporal cortex; **ITa**=Anterior inferotemporal cortex; **MTL**=Medial temporal lobe; **PHC**=Parahippocampal cortex; **PRC**=Perirhinal cortex; **PFC**=Prefrontal cortex; **DLPFC**=Dorsolateral PFC; **VPFC**=Ventral PFC; **SC**=Superior colliculus.

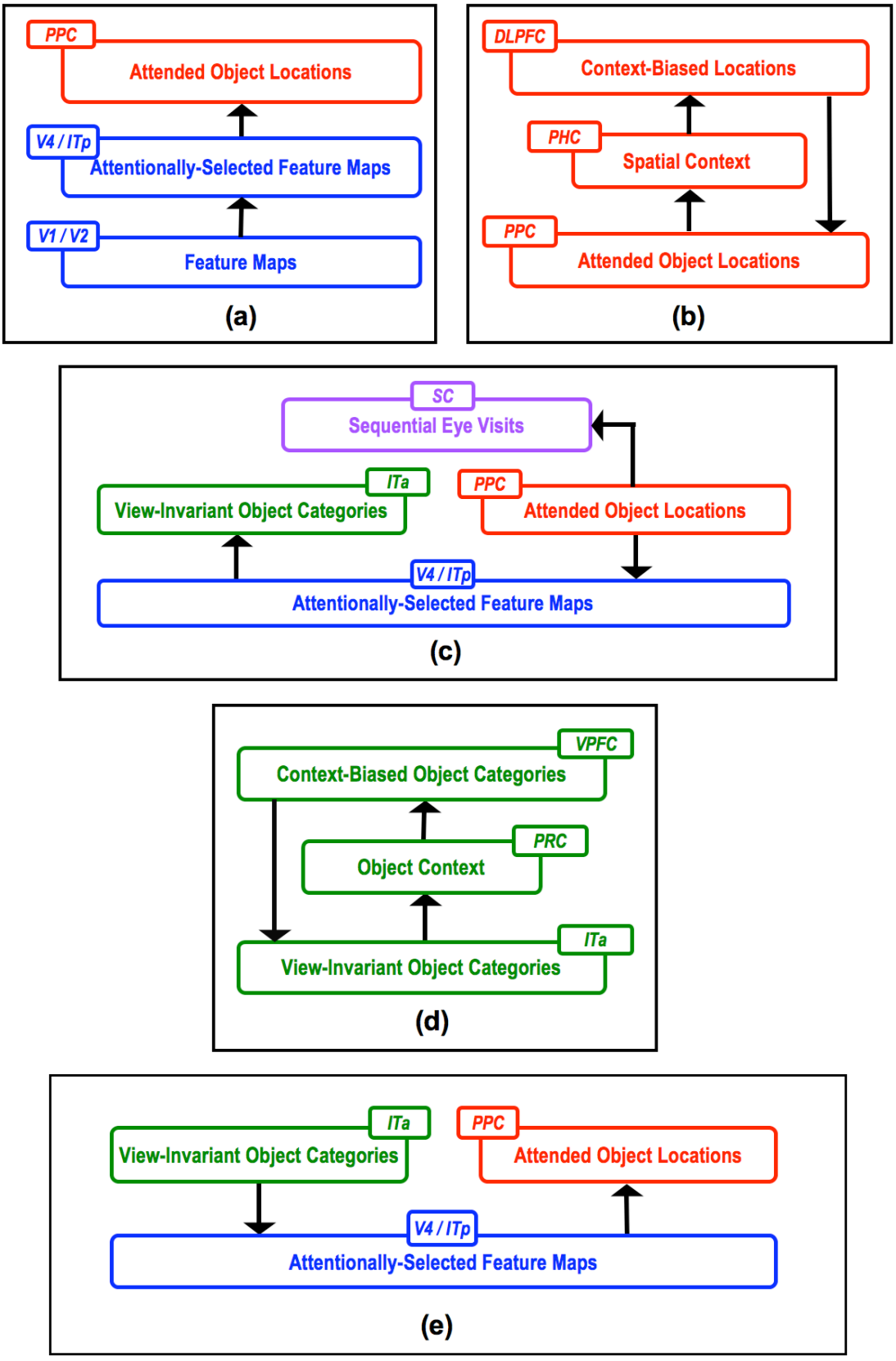

PPC
Attended Object Locations
V4 / ITp
Attentionally-Selected Feature Maps
V1 / V2
Feature Maps
(a)
DLPFC
Context-Biased Locations
PHC
Spatial Context
PPC
Attended Object Locations
(b)
SC
Sequential Eye Visits
ITa
View-Invariant Object Categories
PPC
Attended Object Locations
V4 / ITp
Attentionally-Selected Feature Maps
(c)
VPFC
Context-Biased Object Categories
PRC
Object Context
ITa
View-Invariant Object Categories
(d)
ITa
View-Invariant Object Categories
PPC
Attended Object Locations
V4 / ITp
Attentionally-Selected Feature Maps
(e)

Figure 5. A model search cycle. (a) Step 1: Feature maps in V1/V2 code different features in the search display. V4/ITp undergoes a local competition within each feature map to produce feature-level saliency for each location. Feature-level saliencies are averaged over features into location-level saliencies in PPC. (b) Step 2: The spatial priority map in PPC is cached as a spatial context in PHC, which induces a context-sensitive expectation of target locations in DLPFC. The PPC-PHC-DLPFC-PPC loop then primes possible target locations in PPC with spatial attention. (c) Step 3 (Where to What): PPC drives SC to direct an eye movement, namely overt attention, to the most active location in the spatial priority map. All features at the fixated location are then selected for object recognition in ITa. (d) Step 4: The history of viewed objects is stored in PRC, which builds up a context-induced expectation of target identities in VPFC. The ITa-PRC-VPFC-ITa loop primes possible target categories in ITa with object attention. (e) Step 5 (What to Where): Object-attended ITa further primes feature maps in ITp/V4 with feature-based attention, and thereby boosts PPC to highlight spatial locations with target-like objects. If the currently fixated and recognized object is not a target, the model goes back to Step 2 for maintaining spatial cueing, and then to Step 3 for inspecting other objects based on the updated priority map in PPC.

To focus on the learning and memory mechanisms of contextually guided visual search, ARTSCENE Search simplifies some brain processes underlying search dynamics. First, the model does not factor in all perceptual and oculomotor components that contribute to search time costs, such as stimulus registration and saccade execution. Consequently, the search reaction times in the behavioral data cannot be directly simulated, but are instead compared to the number of gaze locations inspected to discover a target (see also Brady & Chun, 2007; Tseng & Li, 2004). Second, some similarly tuned and anatomically adjacent and/or overlapping brain areas are lumped together in ARTSCENE Search. This includes the model region V1/V2 that simplifies feature computations, and the model region V4/ITp that simplifies processing of figure-ground separated surfaces and recognition categories (Fazl et al., 2009; Zeki, 1996). Third, the simplified connections among model regions are sufficient to simulate all behavioral data treated in this article. Fourth, spatial locations of objects in a search display are all represented by spatiotopic/egocentric coordinates throughout ARTSCENE Search.

In this section, the operation and functional role of each model area will be described qualitatively. The discussion is organized in the order of a model search cycle (Figure 5) from the stages of feature processing (Figure 5a; Section 5.2), to spatial contextual cueing (Figure 5b; Section 5.3), to attention shifts and eye movements (Figure 5c; Section 5.4), and finally to object contextual cueing (Figure 5d and 5e; Section 5.5). References to the Appendix equations in the following text are provided for readers who are interested in the mathematical specification of ARTSCENE Search. Figure 6 locates the mathematical variables that are discussed below, and detailed in the Appendix, to help bridge the gap between heuristic and technical descriptions of the model.

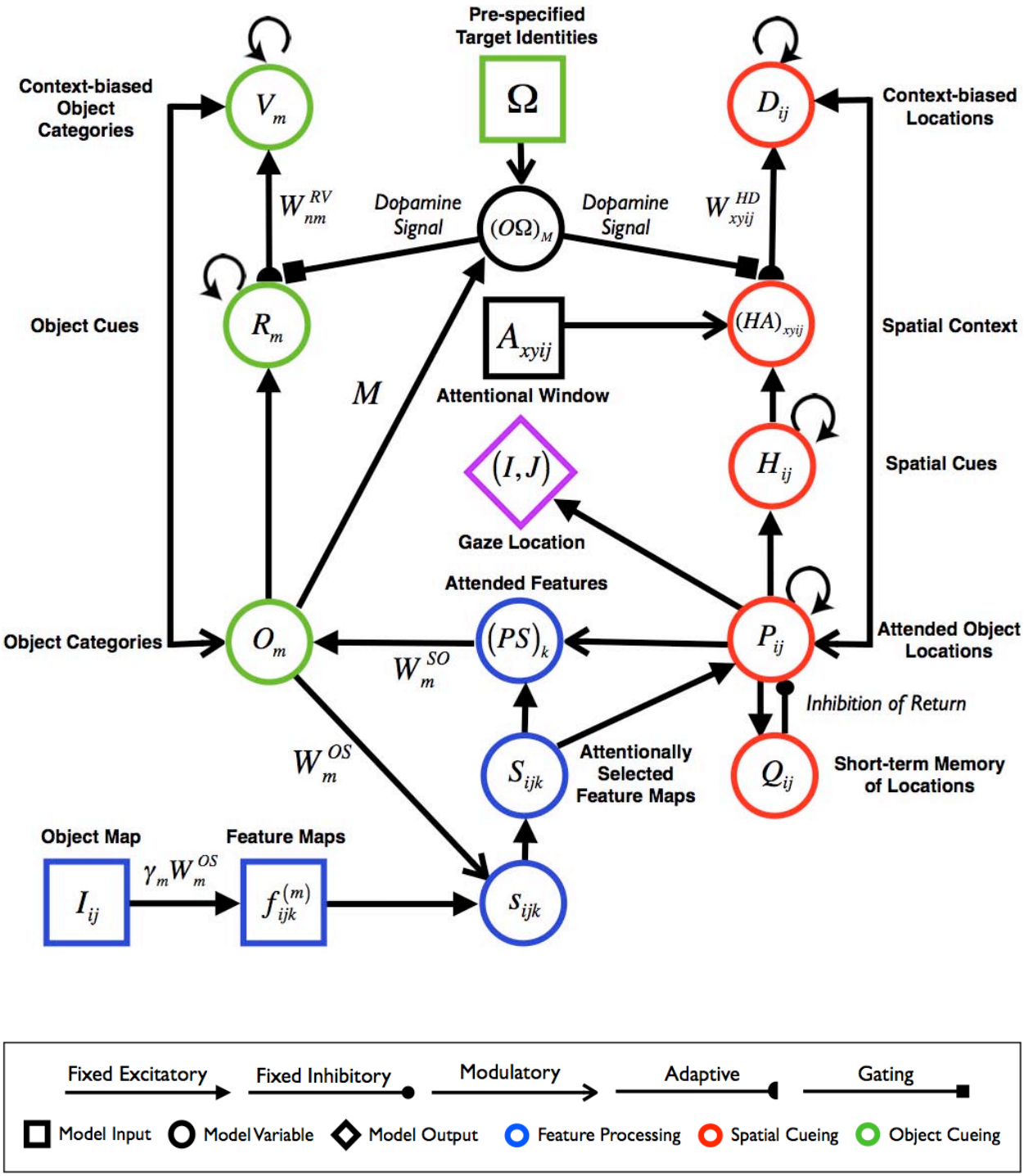


Figure 6. Model variables and their computational relations. All variables are qualitatively defined in Section 5 and mathematically specified in the Appendix. These variables are arranged in a layout similar to the one in Figure 4, although some computational details such as inhibition of return are provided here but not in Figure 4.

5.2 Feature Processing

5.2.1 V1/V2: Feature Maps. Visual areas V1 and V2 in the model represent the neuronal feature responses ($f_{ijk}^{(m)}$ in Equation 4 and Figure 6) to a search display (I_{ij} in Equation 4 and Figure 6). A search cycle begins from this model stage (see Figure 5a). In the brain, V1 and V2 compute boundary and surface properties (i.e., features) 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. Computationally, since an object on a two-dimensional search display is represented by a set, or vector, of visual features ($f_{ij}^{(m)} = \left(f_{ij1}^{(m)}, \cdots, f_{ijk}^{(m)}, \cdots \right)$ in Equation 4), the whole search display (I_{ij}), can be decomposed to a set of two-dimensional feature maps ($f_{ij}^{(m)}$), each of which encodes the strength (γ_m in Equation 4) of a feature (e.g., brightness) in each location of the search display. Such a model front-end is common in feature integration and saliency map models.

5.2.2 V4/ITp: Attentionally-Selected Feature Maps. In ARTSCENE search, model V1/V2 projects only to model V4/ITp. Visual areas V4 and posterior inferotemporal cortex (ITp) in the model represent attention-modulated feature maps (S_{ijk} in Equation 5 and Figure 6). In the early phase of a search cycle (Figure 5a), model V4/ITp receives only bottom-up inputs from model V1/V2 ($f_{ijk}^{(m)}$ in Equation 6 and Figure 6). For each input feature map, the feature-level saliency (S_{ijk}) is then computed for each location through lateral inhibition (Φ_{ijpq} in Equations 5 and 7), which contrast-enhances distinctive features and gives rise to saliency-based attention in the initial phase of attention deployment. In the middle phase of a search cycle (Figure 5c), model V4/ITp also receives a top-down spatial attention signal from posterior parietal cortex (P_{ij} in Equation 19 and Figure 6), which selects all the features at the most attended location ($\left(PS\right)_k$ in Equation 19). This feature pattern, via a bottom-up filter (W_{mk}^{SO} in Equation 18 and Figure 6), activates position-invariant object categories in anterior inferotemporal cortex (ITa: O_m in Equation 17 and Figure 6) thereby leading to selection of the best-matched object (Equation 20). In the late phase of a search cycle (Figure 5e), the active position-invariant category in ITa sends object attentional signals to V4/ITp (s_{ijk} in Figure 6) to enhance feature maps from model V1/V2 that correspond to features of this object category. In the model, V4/ITp thus mediates between the What and Where processing streams. It coordinates Where-to-What interactions (Figure 5c) as well as What-to-Where interactions (Figure 5e).

5.3 Where Stream: Spatial Contextual Cueing

5.3.1 PPC: Attended Object Locations. Posterior parietal cortex (PPC) in the model represents a spatial priority map (P_{ij} in Equation 8 and Figure 6), which indicates how much attention is allocated to each location of a search display (Figure 5c). Model PPC plays a central role in determining gaze locations ($\left(I, J\right)$ in Equation 11 and Figure 6) based on all input signals available in various stages of a search cycle: First bottom-up location saliency (Figure 5a), then top-down spatial modulation (Figure 5b), and finally top-down featural modulation (Figure 5e). Specifically, model PPC averages salient features from V4/ITp (S_{ijk} in Equation 8 and its input to P_{ij} in Figure 6) to influence spatial saliency (P_{ij}), which is gain-modulated by context-sensitive top-down priming by spatial attention from dorsolateral prefrontal cortex (D_{ij} in Equations 8 and 15, and its input to P_{ij} in Figure 6). In addition, feature-based attention (O_m-modulated s_{ijk} in Equation 6 and Figures 5e and 6) highlights target-like objects in PPC by

enhancing target-like features inputted from model V4/ITp to PPC ($s_{ijk} \rightarrow S_{ijk} \rightarrow P_{ij}$ in Figure 6). Thus, the spatial priority map in model PPC provides a feature-sensitive spatial gist of a search display.

During a search trial, model PPC carries out spatial attentional selection for the next gaze location via competition among all locations in the spatial priority map (Equation 8). This competition mechanism normalizes cell activities in the spatial priority map (cf. 'filtering' in Bundesen et al., 2005 and Grossberg & Raizada, 2000) and contrast-enhances activities across the map until the most active cell crosses a threshold (p in Equation 9). The corresponding location then becomes the winning representation in PPC and keeps other location representations under the threshold. In ARTSCENE Search, this winning attended location in PPC drives model superior colliculus (SC in Figure 5c) to trigger a saccade to that location. To disengage attention from a winning location, short-term memory of fixated locations (Q_{ij} in Equation 10 and Figure 6) builds up inhibition-of-return to a selected location (Q_{ij} in Equation 8 and its inhibitory input to P_{ij} in Figure 6), allowing a new cycle of location selection to begin. More complete analyses of how inhibition-of-return may be regulated are modeled in Brown, Bullock, and Grossberg (2004) and Grossberg, Roberts, Aguilar, and Bullock (1997).

5.3.2 PHC: Spatial Context. Parahippocampal cortex (PHC) in the model codes the spatial context, or gist, of a search display (P_{ij}) in short-term memory (H_{ij} in Equation 12 and Figures 5b and 6). Such a spatial map of cue locations is formed early in a search cycle (Figure 5b), even before any eye movements (Figure 5c), and maintained in model PHC throughout a search trial. Computationally, since inhibition-of-return in model PPC (Q_{ij} in Equation 8 and its inhibitory input to P_{ij} in Figure 6) inhibits all previously selected locations in the spatial priority map (P_{ij}), a separate representation of the spatial configuration of a search display is stored in short-term memory by model PHC (H_{ij}) to ensure learning and retrieval of spatial contexts. Due to a rapid integration rate of PHC (τ_H in Equation 12), occupied locations in the PPC spatial priority map (P_{ij} in Equation 12 and its input to H_{ij} in Figure 6) are simultaneously registered in PHC early in a search trial, and this short-term memory of spatial scene gist is only slightly perturbed by its PPC inputs (see small λ in Equation 12) later during spatial selection.

In terms of outputs (see Figure 5b), spatial cues in model PHC (H_{ij} in Equation 12) are first modulated by an attentional window (A_{xyij} in Equation 14) to form an effective spatial context ($(HA)_{xyij}$ in Equation 13 and 16 and Figure 6). Such a context in PHC is associated via the adaptive weights (W_{xyij}^{HD} in Equation 15 and Figure 6) with currently active target locations in model dorsolateral prefrontal cortex (D_{ij} in Equation 15 and Figure 6).

5.3.3 DLPFC: Context-Biased Locations. Dorsolateral prefrontal cortex (DLPFC) receives direct input from the currently active attended location in PPC (P_{ij} in Equation 15 and Figure 6). DLPFC also matches the currently stored spatial context ($(HA)_{xyij}$ in Equations 13 and 15) with the long-term memory of spatial contexts (W_{xyij}^{HD} in Equations 15 and 16) to estimate the likelihood of target presence at each location (D_{ij} in Equation 15 and Figure 6). Model DLPFC hereby provides an estimate of target location that is modulated by the spatial context in which the target is found. DLPFC uses this estimate to provide top-down spatial priming to PPC (D_{ij} in Equation 8 and its input to P_{ij} in Figure 6) that biases selection toward possible target locations.

Starting in the early phase of a search cycle, the PPC-PHC-DLPFC-PPC loop (Figure 5b) carries out spatial contextual cueing, which gain-modulates the spatial priority map in PPC based on the rapidly perceived spatial context of a search display. At the end of a search trial, when a target is identified, the fixated target location becomes the most dominant representation in model DLPFC due to the corresponding PPC input (P_{ij} in Equation 15 and its input to D_{ij} in Figure 6), which is then encoded into long-term memory by being associated with the spatial context that is currently maintained in model PHC ($(HA)_{xyij}$ in Equations 13, 15 and 16). Indeed, across all search trials, spatial context learning (Equation 16) incrementally updates the connection weights between model PHC and DLPFC (W_{xyij}^{HD} in Equation 16 and Figure 6) to encode the experienced context-target correlations in long-term memory. Importantly, a volitionally-regulated attentional window surrounding a location (A_{xyij} in Equations 13 and 14), which is computationally implemented as a Gaussian function, spatially gates the context inputs from model PHC to DLPFC. This gate diminishes the long-term memory *encoding* of cue locations peripheral to a target. Note that it is impossible to retrieve non-encoded peripheral cues from long-term context memory, regardless of the attentional window size during *retrieval*. Therefore, the size of the attentional window (σ_A in Equations 13 and 14, and Table 2) in a simulated trial effectively approximates the dynamics where the attentional window opens broadly for gist processing and spatial context *retrieval* in the early phase of search, and becomes narrowly focused on fixated objects, especially during long-term memory *encoding* of a target in association with its surrounding contextual cues.

The size of the attentional window may be regulated by the frontal eye field through its involvement in covert attention (Moore & Fallah, 2004; Thompson, Biscoe, & Sato, 2005), and/or the basal ganglia via its role in gating, orienting, cognitive, and manipulative behaviors (Alexander, DeLong, & Strick, 1986; Gitelman et al., 1999; Hikosaka & Wurtz, 1989; Passingham, 1993; Strick, Dum, & Picard, 1995).

5.4. SC: Sequential Eye Visits

In the middle phase of a search cycle (Figure 5c), model superior colliculus (SC) receives attended spatial location signals from model PPC (P_{ij} in Equation 11), and generates saccades that direct overt attention and eye fixation onto the attended location ((I, J) in Equation 11 and Figure 6) in the parietal spatial priority map. In other words, any location inspection for featural analysis is always carried out by an overt eye fixation in the model. Although it has been shown that covert attention can be allocated without eye movements (e.g., Posner, Snyder, & Davidson, 1980), covert attention to a location is often followed by a saccade to that location, after which covert and overt attention coincide (see review in Findlay & Gilchrist, 2003). Therefore, overt gaze fixations are informative samples of the covert search process (Zelinsky, 2008). In ARTSCENE Search, the covert attentional window always leads to overt attention and a corresponding eye movement.

In this regard, target inspection may or may not require foveal vision, depending upon the difficulty of a search task. In the extreme case of pop-out search, peripheral vision is often sufficient for detecting or locating a distinctive target. Conversely, when targets and distractors are highly similar, foveal vision or eye fixations may be needed for observers to confirm the identity of a target candidate. For example, in the difficult 'Where's Waldo' task reported by Otero-Millan, Troncoso, Macknik, Serrano-Pedraza, and Martinez-Conde (2008), the average fixation duration was doubled from 283ms to 600ms when fixations were near identified targets. In ARTSCENE Search, fixation of a target terminates a search trial, and one gaze is required for

a pop-out search. This model treatment is a reasonable approximation to reality wherein most search tasks are difficult. Computationally, model saccades are implemented algorithmically and instantaneously by changing the simulated gaze position from one to another (Equation 11).

5.5. What Stream: Object Contextual Cueing

5.5.1 ITa: Position-Invariant Object Categories. Anterior inferotemporal cortex (ITa) in the model represents position-invariant object categories (O_m in Equation 17 and Figure 6). In the middle phase of a search cycle (Figure 5c), model ITa carries out position-invariant object recognition for fixated features ($(PS)_k$ in Equations 18-19 and its input to O_m in Figure 6). Computationally, object recognition in the model is achieved by filtering of attended features ($(PS)_k$) by the prototype features of each object category (W_{mk}^{SO} in Equation 18 and Figure 6). If the attended features match the category prototype well enough, the object category becomes activated above a recognition threshold ($\psi_{0.5}(O_m)$ in Equation 22 and Figure 6) and thereupon drives the corresponding object representation in model perirhinal cortex (PRC: R_m in Equation 22 and Figure 6) to be stored in a temporally developing object context. Moreover, if the attended object category (M in Equation 20 and Figure 6) is a search target (Ω in Equation 21 and Figure 6), then nonspecific dopamine signals in the model ($(O\Omega)_M$ in Equations 16, 21, 24, and Figure 6) are broadcast to both the What and Where streams where they trigger associative target-cue learning between the frontal and medial temporal lobes (Equations 16 and 24), after which the ongoing search trial terminates.

For top-down modulations, model ITa projects back to ITp/V4 and converts object-based priming from ventral prefrontal cortex (V_m in Equation 17 and its input to O_m in Figure 6) to feature-based priming in the feature maps (O_m in Equation 6 and its input to s_{ijk} in Figure 6). Specifically, the activation level of an object category (O_m) in model ITa proportionally gain-modulates the entire feature map in model ITp/V4 (s_{ijk}) that represents that object. 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). As a consequence of such feature-based modulation across the whole search display, attention will be allocated more to the locations that accommodate known search features, which is a major What-to-Where interaction in the model (see Figure 5e).

5.5.2 PRC: Object Context. Perirhinal cortex (PRC) in the model represents the same set of object categories as in ITa, and codes in short-term memory a temporally evolving context of recognized objects (R_m in Equation 22 and Figure 6). Such a history of overtly attended objects is incrementally developed in model PRC after a sequence of eye visits during the middle phase of a search trial (see Figures 5c and 5d). Computationally, since object recognition is carried out for overtly attended objects (M in Equation 20 and Figure 6) one at a time due to the winner-take-all spatial selection in the model ($(PS)_k$ in Equations 18-19 and its input to O_m via o_m in Equation 17 and Figure 6), the identities of all viewed objects in a search display are not simultaneously available without a short-term memory mechanism. To ensure learning and retrieval of a remembered object context, all recognized object categories in model PRC (R_m in Equation 22), unlike their upstream counterparts in ITa (O_m in Equation 17), remain active in short-term memory. In terms of outputs (see Figure 5d), the projection from model PRC to ventral prefrontal cortex (VPFC: V_m in Equations 23 and 24, and Figure 6) encodes the learned correlations between cue and target identities (W_{nm}^{RV} in Equation 23 and Figure 6). As a result, previously viewed object identities that are stored in model PRC can build up an expectation of

the current target identity in VPFC using associative votes ($\sum_n R_n W_{nm}^{RV}$ in Equation 23).

5.5.3 VPFC: Context-Biased Object Categories. Ventral prefrontal cortex (VPFC) in the model comprises the same set of object categories as the ones in model ITa and PRC. Output signals from the short-term memory of the current object context in perirhinal cortex (PRC: R_m in Equation 22) are filtered by learned long-term memory traces of object contexts (W_{nm}^{RV} in Equations 23 and 24) to predict the likelihood of a target in that context during a search trial (V_m in Equation 23). Thus, in the late phase of a search trial, when an object context is already stored in model PRC, model VPFC becomes the source of top-down context-based object attention (Figure 5d) and feature-based attention (Figure 5e), which biases attention to target-like objects in space. In particular, model VPFC directly feeds identity-based priming back to ITa (V_m in Equation 17 and its input to O_m in Figure 6), and indirectly causes feature-based priming from the gain-modulated ITa to ITp/V4 (O_m in Equation 6 and its input to s_{ijk} in Figure 6), which in turn may shift spatial attention in the PPC spatial priority map (Figure 5e; $s_{ijk} \rightarrow S_{ijk} \rightarrow P_{ij}$ in Figure 6). In summary, the ITa-PRC-VPFC-ITa loop (Figure 5d) carries out object contextual cueing during search, which gain-modulates the PPC spatial priority map based on the gradually developing object context of a search display. After a target is indentified, the category of a fixated target becomes the most dominant representation in model VPFC due to the corresponding ITa input (O_m in Equation 23 and its input to V_m in Figure 6), and is encoded into long-term memory as part of the object context maintained in model PRC (R_m in Equation 22). Across search trials, object context learning (Equation 24) incrementally updates the connection weights between model PRC and VPFC (W_{nm}^{RV} in Equation 24 and Figure 6) to encode the experienced context-target correlations.

6. Simulation Results

6.1. Simulation Overview

ARTSCENE Search provides a unified explanation and simulation of data about spatial/object, distant/local, and positive/negative contextual cueing effects. Several factors go into this explanation: First, spatial and object cueing effects are manifestations of the learned associations between a target and its context objects in location or featural appearance, respectively. In a similar manner, spatial and object cueing facilitate target search by predictively biasing attention toward target locations and features that are most likely to co-occur with the stored contexts. As a result, overt attention is less drawn to distractors during search, which is reflected by a reduced search reaction time (RT).

Second, distant versus local cueing effects can be reconciled by attention-dependent context learning. Compared with ordinary cues that are visually similar to the background, an attentionally salient cue, regardless of its spatial distance from the target, can be encoded more strongly into context memory, and thus induce spatial cueing more effectively.

Third, positive and negative spatial cueing effects result from different sizes of the attentional window, which down-regulates the efficacies of peripheral locations as spatial cues. A widely spread attentional window centered on a target allows associative learning between the target and its surrounding objects within the window, which leads to positive cueing effects. In contrast, an attentional window that focally bounds a target alone prevents that target from being contextually associated with its surrounding objects, which eliminates contextual facilitation. In this case, a repeated spatial context is no different from a novel context in distracting target

search, and sometimes even slows down search more than a novel context just by chance, as shown by some individual observers as negative cueing effects.

A complete mathematical specification of ARTSCENE Search is provided in the Appendix. Simulation codes and documents of the model can be found at http://cns.bu.edu/~tren/artscene_search. Simulations of the model equations 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 the Appendix as Table 2. In addition to experimental parameters, the three free model parameters listed in Table 2 are the attentional window size and the learning rates for spatial or object contexts. The attentional window size determines the efficacies of local, distant, and overall contexts in spatial cueing. It qualitatively changes the results between experimental conditions (e.g., separate vs. overlapping curves), whereas the two learning rates only quantitatively change how fast learning curves converge to equilibrium.

There are also fixed parameters embedded in the model equations (see Appendix). Unlike the three free parameters, these fixed parameters regulate model dynamics more at the neural than the behavioral level. Overall, these parameters determine the relative strengths of passive decay, excitatory/inhibitory signals, bottom-up/top-down inputs, and featural/spatial priming in the system. Ideally, these values should be consistent with neurophysiological data, if available. Empirically, they were chosen to ensure a functioning system, notably a proper balance among all the functionally-important processes within a model region, as well as a balance among all the model regions. For all ten simulations presented later in this section, the same set of fixed parameters were used. The model produces qualitatively similar properties when these parameters are perturbed within a reasonable range.

Although some behavioral properties of search reaction time may partially be represented within a simpler model with a few global descriptive parameters just for the sake of data-fitting, a neural model with more parameters leads to a greatly expanded explanatory and predictive range, in part by providing more insight into the underlying brain mechanisms and the interactive dynamics that give rise to behavioral data as emergent properties. For example, in the two-layer network model of Brady and Chun (2007), it is not clear how the configuration of a spatial context can be learned at the end of a search trial after a series of inhibition of returns during search, which suppress the representations or saliencies of individual locations in that spatial context. In ARTSCENE Search, such neural dynamics are ensured by the model parahippocampal cortex, which stores the spatial context of a scene (i.e., spatial gist) in short-term memory until the end of a search trial. It should also be noted that, although the model incorporates properties of multiple brain regions, all the model processes are based on variations of the same basic equations for cell activation, habituation, and learning.

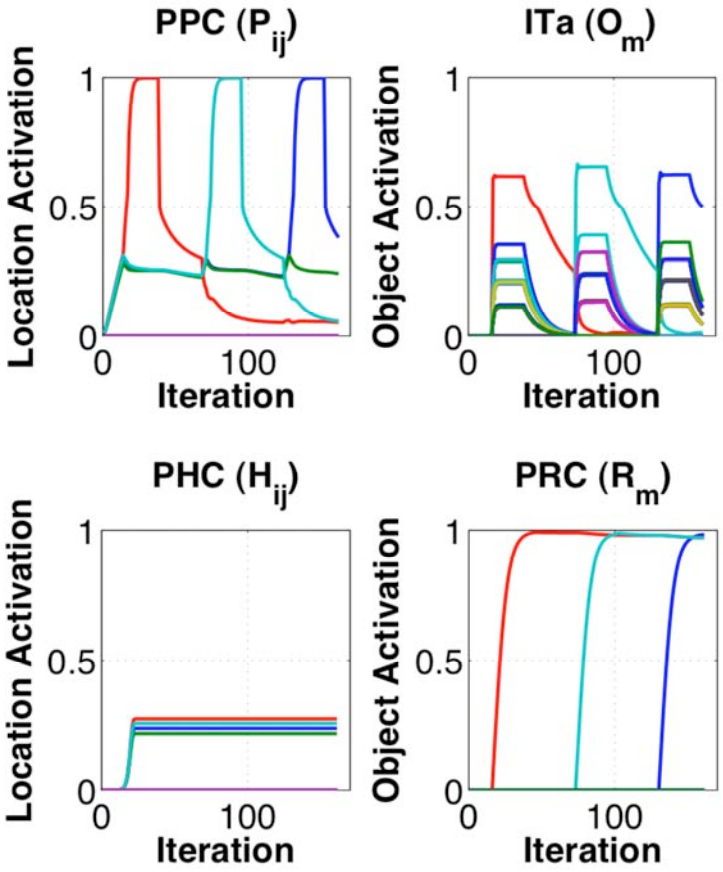
Model simulations are both quantitative and qualitative. Quantitatively speaking, all model simulations can be evaluated using standard statistical tests across conditions and factors. However, ARTSCENE Search does not factor in all possible biological processes, notably perceptual preprocessing and oculomotor execution, to simulate the exact value of the search reaction times (RT). Instead, the number of eye fixations in a trial is measured as the model output. In this sense, the simulation results are qualitative and ordinal across conditions. Due to the use of the number of eye fixations to estimate RT, we cannot utilize curve-fitting techniques such as regression to determine parameter values based on the goodness of fit. Indeed, fixation duration increases as task difficulty increases in reading, visual search, and scene perception

(Rayner, 2009). Therefore, model parameters were chosen by subjective evaluation of the match between data and simulation.

Some conventions implicit in the article should be noted here. First, a search trial always contains a single target in all experiments and simulations that are presented in Section 6. Second, experiment and simulation results are listed, respectively, on the left and right panels of Figures 10-18 for side-by-side comparisons. Third, data points and error bars in Figures 10-19 are the means and standard errors calculated across subjects. A simulated subject refers to a new simulation session, which independently generates search displays for all experimental conditions. Fourth, for easy comparison across simulation results, the number of subjects was fixed as 15 in all the simulations. Fifth, although the saliency of a location is modulated by top-down factors in the model, 'saliency' in the text often refers to only the bottom-up component of attentional valence. Sixth, while a 'spatial cue' refers to a target-predictive location, a 'spatial context' refers to a spatial configuration or a set of locations that are presented together with a target but not necessarily target-predictive. Similarly, while an 'object cue' refers to a target-predictive object, an 'object context' refers to a set of objects that are presented together with a target but not necessarily target-predictive. Finally, a 'contextual cue' refers to both the identity and location of an object in space.

6.2. Model dynamics during a Search Trial

To illustrate the model dynamics of ARTSCENE Search during a search trial, the activities of model cells in both the Where (PPC-PHC-DLPFC) stream and What (ITa-PRC-VPFC) stream are provided in Figures 7-9 for a search example. In the simulations, the overall saliency of each object/location (γ_m in Equation 4) was pre-specified in such a way that the target ranked third in saliency.



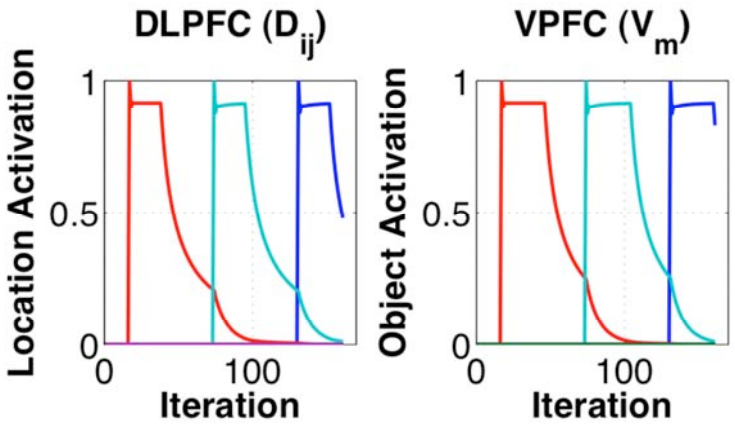
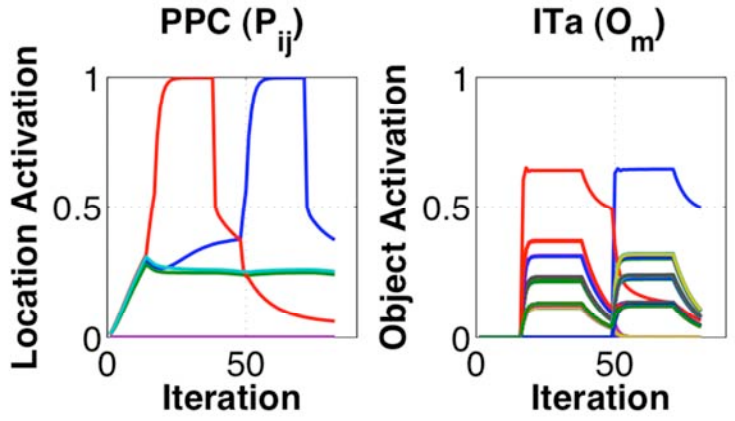


Figure 7. Model dynamics during a search trial with bottom-up saliency-based attention but no top-down priming. Each figure panel shows the cell responses in each model region. The x-axis represents iteration number during numerical integration (Equation 2), and the y-axis represents activation level of model neurons (Equation 1).

Figure 7 shows the model dynamics during a search trial before context learning. In the simulation, four objects were presented in a search display and the target was found during the third fixation. An object *m* and its corresponding location (*i, j*) are color-coded for all panels in Figure 7. Since top-down priming is not induced yet, the deployment of attention followed the order of decreasing saliency (i.e., red → cyan → blue → green curve). Notice that model PHC codes the relative strength of object/location saliencies *in parallel* since the very beginning of search due to its rapid integration rate (τ_H in Equation 12). Later during spatial selection, this short-term memory of spatial scene gist is slightly perturbed by its PPC inputs at a much finer scale (see small λ in Equation 12) that is invisible in the PHC figure panel. In model ITa, object categories that shared similar features all responded to a specific featural input, and competed with each other. Thus, more than four traces were elicited in the ITa figure panel.



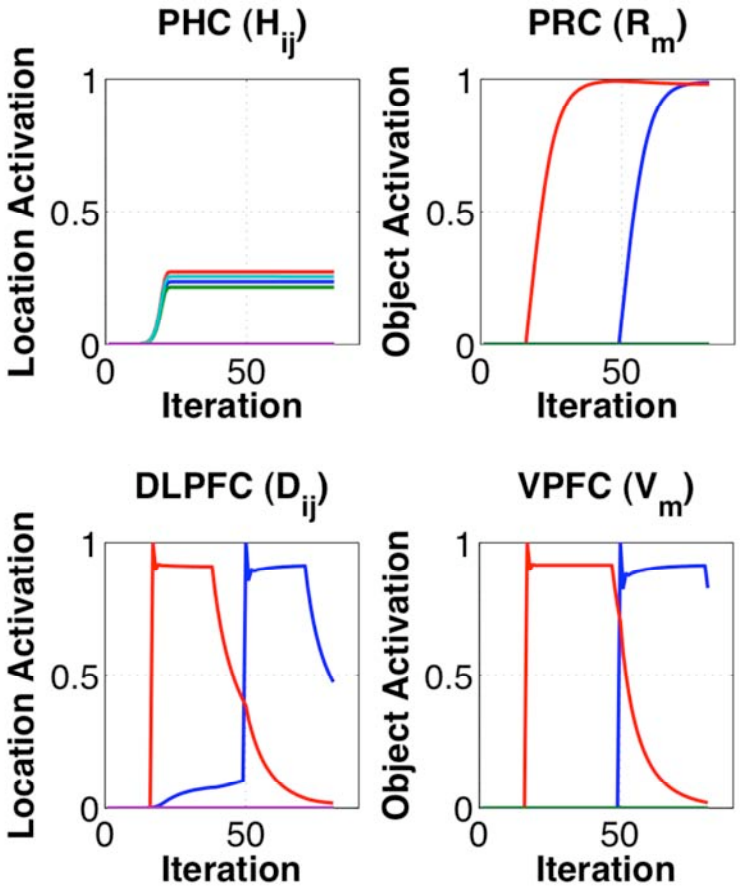


Figure 8. Model dynamics during a search trial with top-down priming of spatial attention. The x-axis represents iteration number during numerical integration (Equation 2), and the y-axis represents activation level of model neurons (Equation 1).

Figure 8 shows the model dynamics during a search trial with spatial contextual cueing. The simulation is the same as the one in Figure 7 except that the connection weights between model PHC and DLPFC are nonzero after learning of the spatial context. Due to spatial contextual cueing, the target was found during the second rather than the third fixation. Notice the gradually developed target representations (blue curves) in model DLPFC and PPC illustrate subthreshold priming for the target location.

Figure 9 shows the model dynamics during a search trial with object contextual cueing. The simulation is the same as the one in Figure 7 except that the connection weights between model PRC and VPFC are nonzero after learning of the object context. Due to object contextual cueing, the target was found during the second rather than the third fixation. Notice the gradually developed target representations (blue curves) in model VPFC and ITa illustrate subthreshold priming for the target identity.

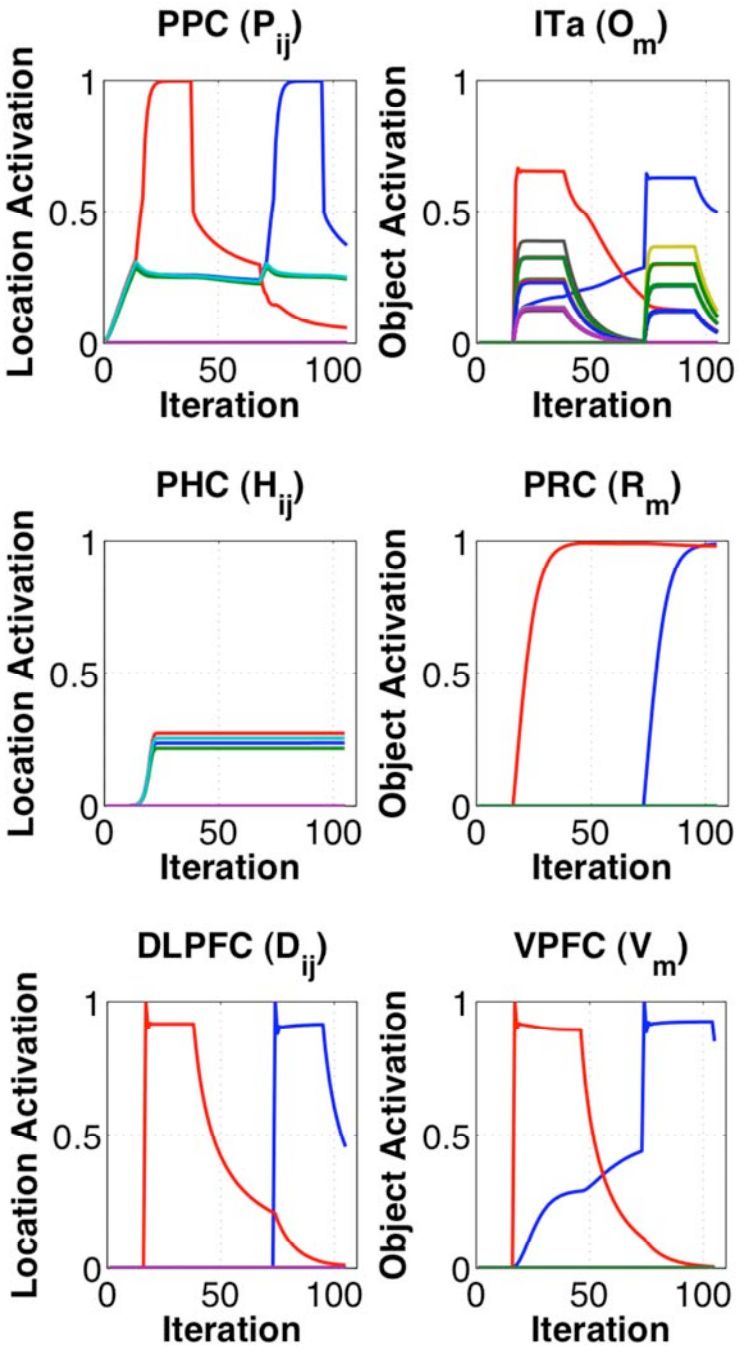


Figure 9. Model dynamics during a search trial with top-down priming of object attention. Each figure panel shows the cell responses in each model region. The x-axis represents iteration number during numerical integration (Equation 2), and the y-axis represents activation level of model neurons (Equation 1).

6.3. Positive Spatial Cueing

Positive spatial cueing effects are the reaction time (RT) reductions for search in a familiar spatial context compared to a new context. In the spatial cueing paradigm (Chun & Jiang, 1998), a fixed target location was chosen from a grid search display without replacement, and presented

in one trial per block. Across blocks of search trials, a target location was accompanied by either a repeated spatial configuration of distractors ('Old' condition) throughout the entire experiment, or by a random configuration that was newly generated in each block ('New' condition). Figure 10 shows a typical result from the spatial cueing paradigm. In the figure, the x-axis represents search epochs grouped from blocks of trials (see Table 2 in the Appendix), and the y-axis represents search RT for completing a trial. Since the upper and lower curve in each panel of Figure 10 corresponds, respectively, to the 'New' and 'Old' spatial context condition, the separation between these two curves indicates the amount of contextual facilitation in search RT from a regular spatial context. Notice that the RT in the 'New' spatial context condition dropped across epochs, and a further RT reduction in the 'Old' spatial condition also developed as the session progressed.

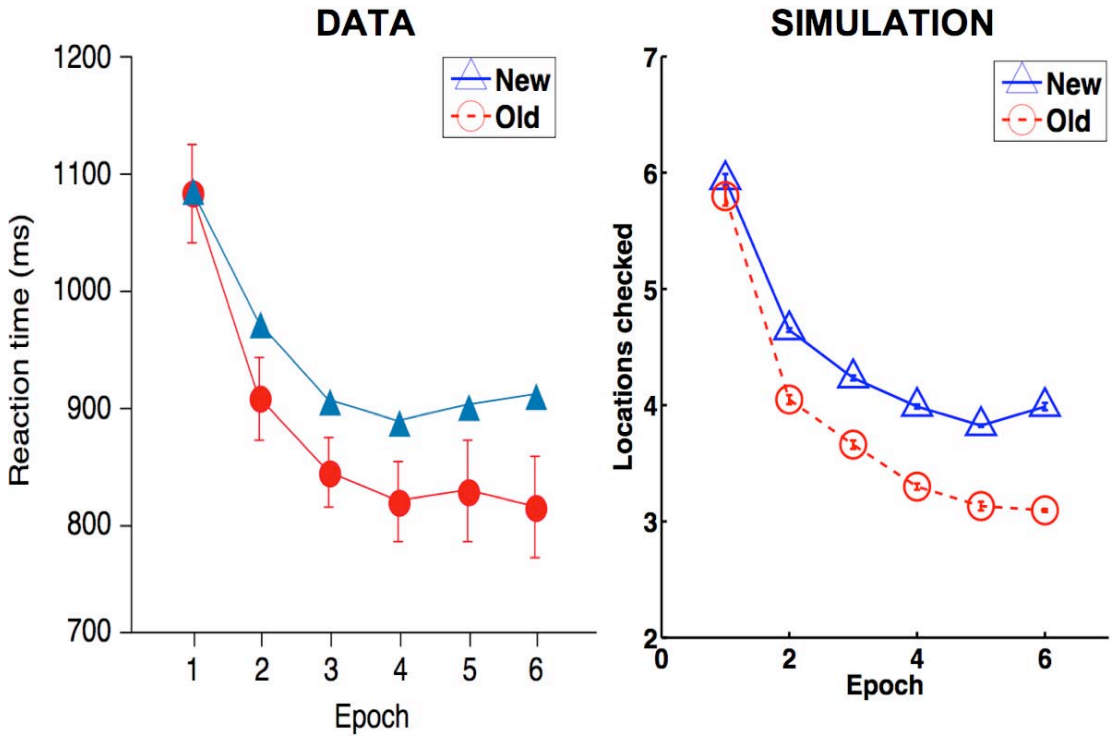


Figure 10. Positive spatial cueing effects are the RT reductions for search in a familiar spatial context (i.e., 'Old' condition) compared to a novel context (i.e., 'New' condition). The x-axis represents training epoch grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time for completing a trial. (Data reprinted with permission from Chun, 2000).

ARTSCENE Search replicates spatial cueing effects through learning of pairwise associations between a context location and a target location. Specifically, when a search display is presented, the layout of search items forms a spatial scene gist, which activates model posterior parietal cortex and its downstream parahippocampal cortex as the eyes search a scene. Each context location represented in model parahippocampal cortex then learns to vote for its correlated target locations represented in dorsolateral prefrontal cortex (PFC), collectively building up a spatial map in model dorsolateral PFC about the likelihood of seeing a target at each location. Afterwards, the spatial priority map in model posterior parietal cortex is gain-modulated by top-down feedback from dorsolateral PFC to bias attention toward possible target locations given the current scene layout (see Figure 5b). As a consequence, an eye-scan path

becomes more target-based rather than saliency-based. Accordingly, the probability of fixations on salient distractors is reduced, reflected as spatial cueing effects. Importantly, the strongest pairwise association learned by the model is typically from a target location to itself due to its perfect self-correlation. Unlike the locations that never accommodate a target, a target location itself, once it re-appears in a search trial, signifies target presence and strongly attracts overt attention. Therefore, search RT can still decrease during the course of training even if a target location is presented in combination with a new context.

6.4. Memory Representation of Spatial Contexts

Spatial contextual cueing can be obtained from a novel context configuration consisting of target-predictive individual locations, as shown by Jiang and Wagner (2004, Experiment 1). In their experimental design, two sets of spatial configurations of distractors were generated for each target location, and each set was presented together with the corresponding target location in one search trial per block throughout the entire training session. Immediately after training blocks of trials was the transfer test whose results are shown in Figure 11. In the figure, the y-axis represents the median search RT for completing a trial in each condition, and the 'New', 'Old', and 'Recombined' conditions in order refer to search in a novel, familiar, or recombined spatial configuration with respect to the trained configurations. A recombined context configuration for a target location was a half-half blend of two 'Old' configurations that were initially paired with that target location during training. Interestingly, the 'Recombined' and 'Old' conditions were equally effective for reducing search RT, compared to the 'New' context condition.

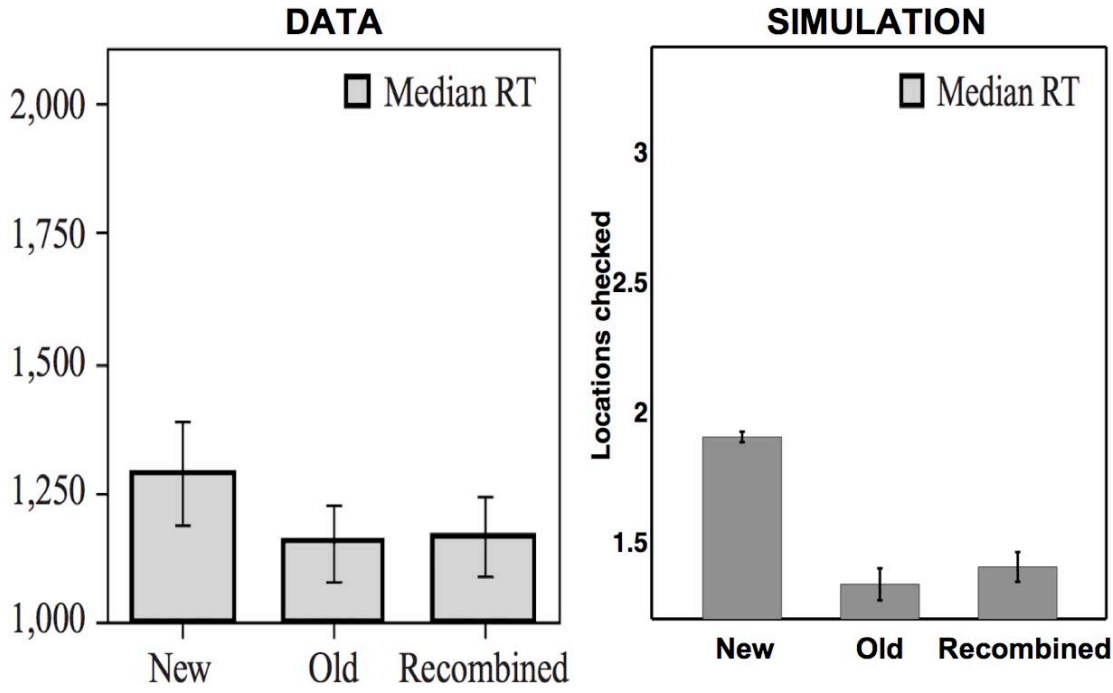


Figure 11. Spatial contextual cueing can be obtained from a novel context configuration consisting of target-predictive individual locations. In the graphs, the y-axis represents search reaction time for completing a trial, and the 'New', 'Old', and 'Recombined' conditions refer to, respectively, search in a novel, familiar or recombined configuration. The recombined configuration was a half-half blend of two 'Old' configurations that were initially paired with the

same target location during training. (Data reprinted with permission from Jiang & Wagner, 2004, Experiment 1).

ARTSCENE Search can replicate the learning transfer from recombined spatial contexts, again through learning of pairwise associations between a target location and a context location. In the model, 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. In consequence, a novel spatial scene gist in model parahippocampal cortex can still drive certain location representations in dorsolateral prefrontal cortex to generate spatial cueing effects as long as individual locations in the layout consistently predict the same target location.

6.5. The Set Size Effect in Spatial Cueing

Using the spatial cueing paradigm with a different number of search items in the display of a search trial (i.e., set size), Chun and Jiang (1998, Experiment 4) found that the magnitude of spatial cueing effects increased as the set size increased. In their data shown in Figure 12, the x-axis represents search epochs grouped from blocks of trials (see Table 2), and the y-axis is the magnitude of spatial cueing effects obtained by subtracting search RT for old contexts from that for new contexts. Hence, the positive values indicate RT benefits for search in old contexts. In each block and hence epoch, three set sizes (8, 12, 16) of trials were intermixed, and spatial cueing effects were more pronounced when the search set size was larger (see also Tseng & Li, 2004).

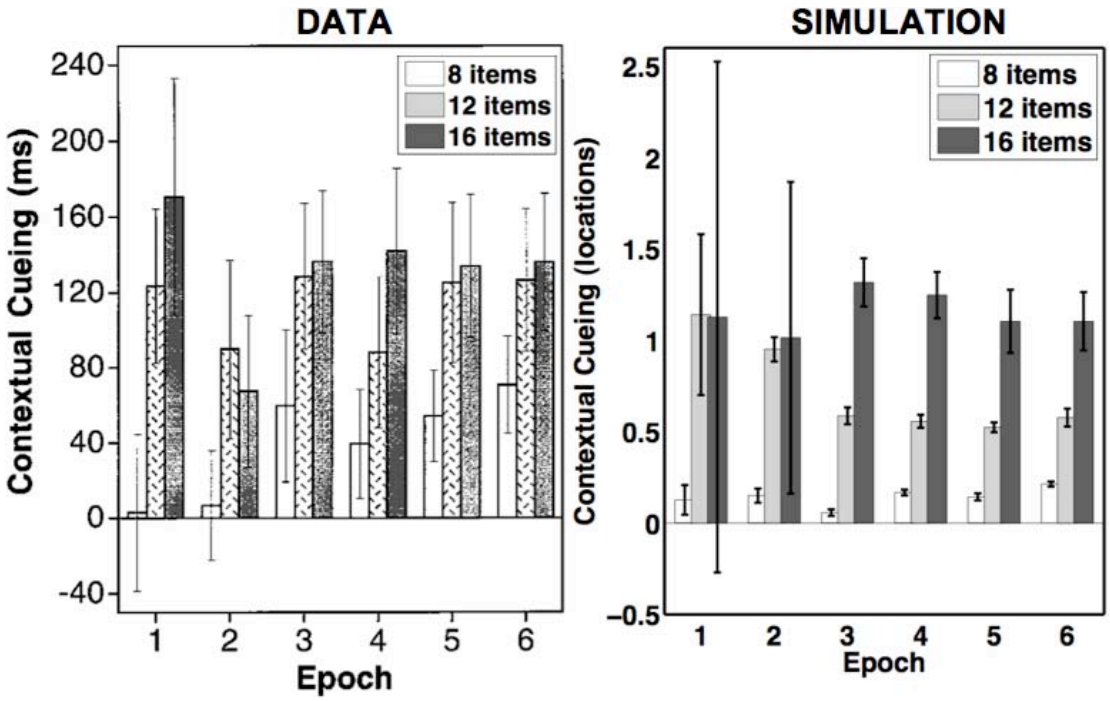


Figure 12. The set size effect in spatial contextual cueing. Context-induced RT reductions on the y-axis were more pronounced when the search set size of a trial was larger. The x-axis represents training epoch grouped from blocks of trials (see Table 2). (Data reprinted with permission from Chun & Jiang, 1998, Experiment 4).

In ARTSCENE Search, the set size effect in spatial cueing is a natural result from the learned pairwise associations between a context location and a target location, because associative votes from context locations in model paraphippocampal cortex to a target location are additively accumulated in model dorsolateral prefrontal cortex (PFC). The more associative votes or inputs to a specific location representation in model dorsolateral PFC, the stronger spatial priming/cueing for that location will be projected from model dorsolateral PFC back to posterior parietal cortex (see Figure 5b).

6.6. Locality of Spatial Cueing

Locations equally predictive of a target location may disproportionately contribute to spatial cueing effects, which cannot be explained solely by associative learning. Figure 13 shows the data reported by Olson and Chun (2002) who compared spatial cueing effects from spatial contexts in separate visual hemifields (left vs. right in Experiment 1 or upper vs. lower in Experiment 2). In the figure, the x-axis represents search epochs grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time for completing a trial. The trial conditions 'New', 'GlobalCxt', 'LRCxt', and 'SRCxt' refer, respectively, to search in novel, globally repeated, long-range, and short-range spatial contexts with respect to the target location. In Figure 13, 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 (see also Alvarez & Cavanagh, 2005 and Kingstone, Enns, Mangun, & Gazzaniga, 1995 for hemifield differences in visual search and attentional tracking). A follow-up study by Brady and Chun (2007) strengthened this locality finding by showing that a local context surrounding a target in a quadrant of a search display was as effective as a global context across quadrants.

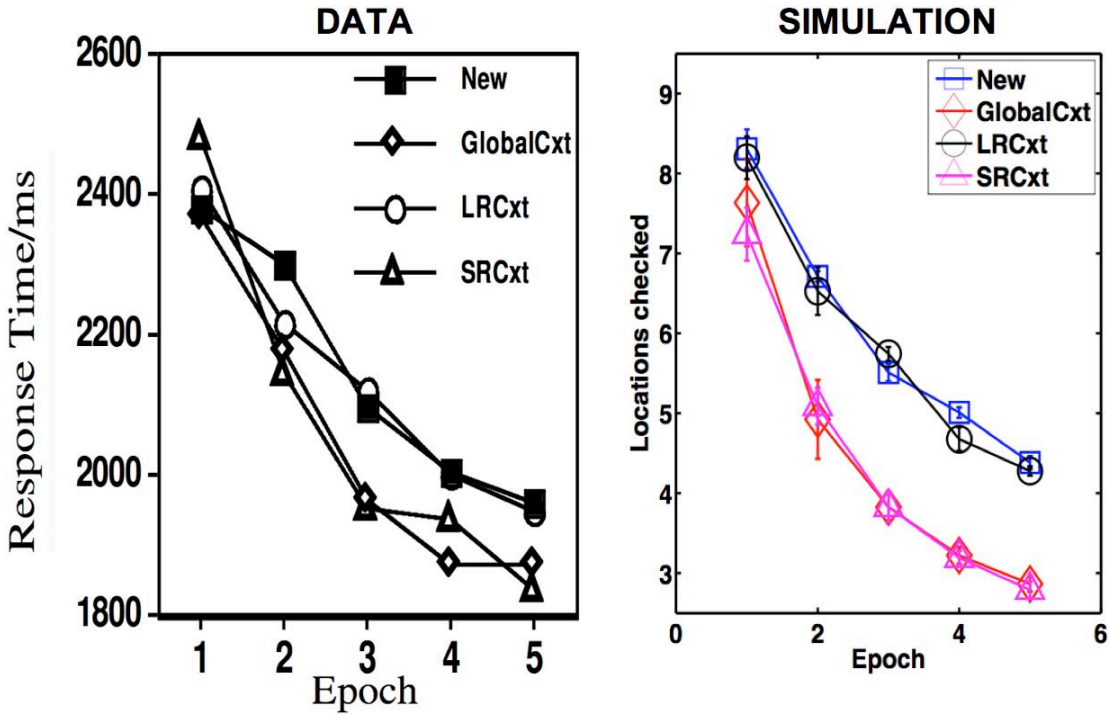


Figure 13. Spatial cueing effects can be mainly attributed to target-predictive locations closer to the target, such as those in the same visual hemifield. In the graphs, the conditions 'New',

'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 epoch grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time for completing a trial. (Data reprinted with permission from Olson & Chun, 2002, Experiment 2).

ARTSCENE Search reproduces the locality effects in spatial cueing by introducing an attentional window to the model. The attentional window is computationally implemented as a Gaussian function that spatially gates the stored contextual inputs from model parahippocampal cortex to dorsolateral prefrontal cortex, and disengages peripheral cue locations from the remaining processes of spatial cueing. Functionally, the attentional window imposes a perceptual span (for relevant data, see Bertera & Rayner, 2000; Geisler, Perry, & Najemnik, 2006; Loschky, McConkie, Yang, & Miller, 2005; McConkie & Rayner, 1975; Miellet, O'Donnell, & Sereno, 2009; Nelson & Loftus, 1980; Rayner & Bertera, 1979; Saida & Ikeda, 1979; van Diepen & d'Ydewalle, 2003; for a review, see Rayner, 2009) for incoming contexts by down-regulating the short-term memory representation of peripheral locations in model parahippocampal cortex. Therefore, when a target location is fixated at the end of a search trial, only local cue locations surrounding the target location are selected in short-term memory, and further encoded into long-term context memory in association with the target location represented in model dorsolateral prefrontal cortex.

6.7. Null or Negative Spatial Cueing

Target-predictive spatial contexts do not necessarily reduce search reaction time (RT). Using the same spatial cueing paradigm (Section 6.3) that was introduced by Chun and Jiang (1998), Lleras and von Mühlenen (2004) found that task instructions for a visual search task critically determined whether regular spatial contexts could be learned by observers to show spatial cueing effects. In their study, the 'passive strategy' group of participants was asked to be receptive and intuitive during search, whereas the 'active strategy' group of participants was instructed to be active and deliberate during search. The results of the 'passive strategy' group replicated classic spatial cueing effects. The data averaged from the 'active strategy' group are shown in Figure 14 where the x-axis represents search epochs grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time (RT) for completing a trial. Interestingly, the 'Old' and 'New' context conditions resulted in comparable search RT (i.e., a null cueing effect). In other words, spatial cueing benefits disappeared when observers searched with highly focused attention. In this case, all non-learned context locations simply distract search. Some observers, as statistical fluctuations of the group average, even searched more slowly in the 'Old' than in the 'New' context condition (i.e., negative cueing effects).

ARTSCENE Search simulates the null/negative cueing effects by constricting the attentional window in the model to minimize context learning. As described in Section 6.6, a Gaussian function is introduced as an attentional window to gate inputs from model parahippocampal to dorsolateral prefrontal cortex (PFC). When the width or variance of the Gaussian window is sufficiently small to only cover a single location in the search display, all contexts surrounding a target location are down-regulated for learning, and self-priming of a target location is the only association that can be established from model parahippocampal cortex to dorsolateral PFC. In this scenario, the search RT can still drop across epochs in the 'New' context condition due to the self-priming of a target location, but there is no further RT reduction in the 'Old' context condition because contexts surrounding a target are not well perceived and learned.

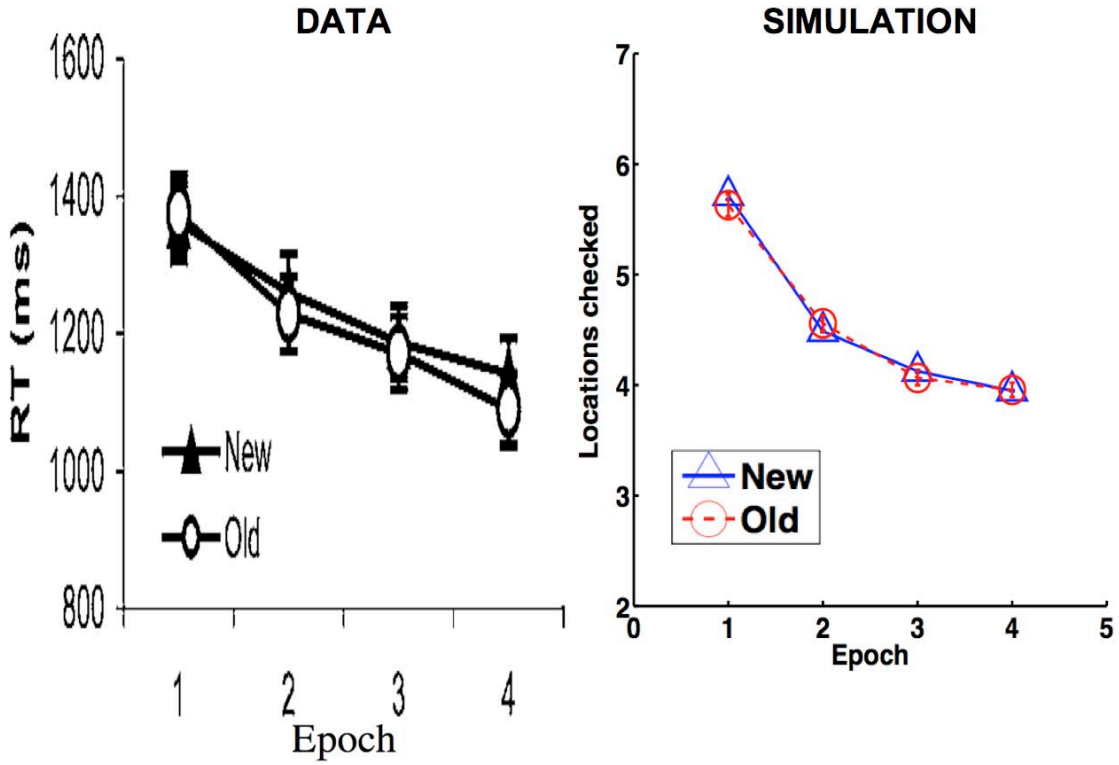


Figure 14. Negative cueing effects or context-induced search RT increases can arise at the single subject level due to focused attention. At the group level in which search RTs were averaged across subjects, there was no significant RT difference for search in a familiar spatial context ('Old' condition) or a novel one ('New' condition). The x-axis represents training epoch grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time for completing a trial. (Data reprinted with permission from Lleras & von Mühlenen, 2004, Experiment 3).

6.8. Saliency-Based Distant Cueing

Opposite to the locality of spatial cueing discussed in Section 6.6, Brockmole et al. (2006) reported that global/distant contexts yielded greater cueing effects than local contexts. In their study using naturalistic scenes, a search target was always positioned on a table in a room (see Figure 3c). With respect to the target, a table was thus a local context, whereas other room objects were global or distant contexts. During training, only one scene was repeated across blocks for the 'repeated' context condition, and the other scenes were presented exactly once in the entire experiment for the 'novel' context condition. During transfer, observers in the local change condition saw a different table in the repeated room, whereas observers in the global change condition saw the same repeated table in a different room. Their experimental results are shown in Figure 15 where the x-axis represents three training epochs grouped from blocks of trials (see Table 2) plus one transfer block, and the y-axis represents the search reaction time (RT) for completing a trial. Remarkably, RT benefits for search in a repeated scene were retained when the local context (i.e., table) was changed, but were eliminated when the global or distant context (i.e., room) was changed. In other words, the distant rather than local context was responsible for the observed RT reductions in their repeated scene.

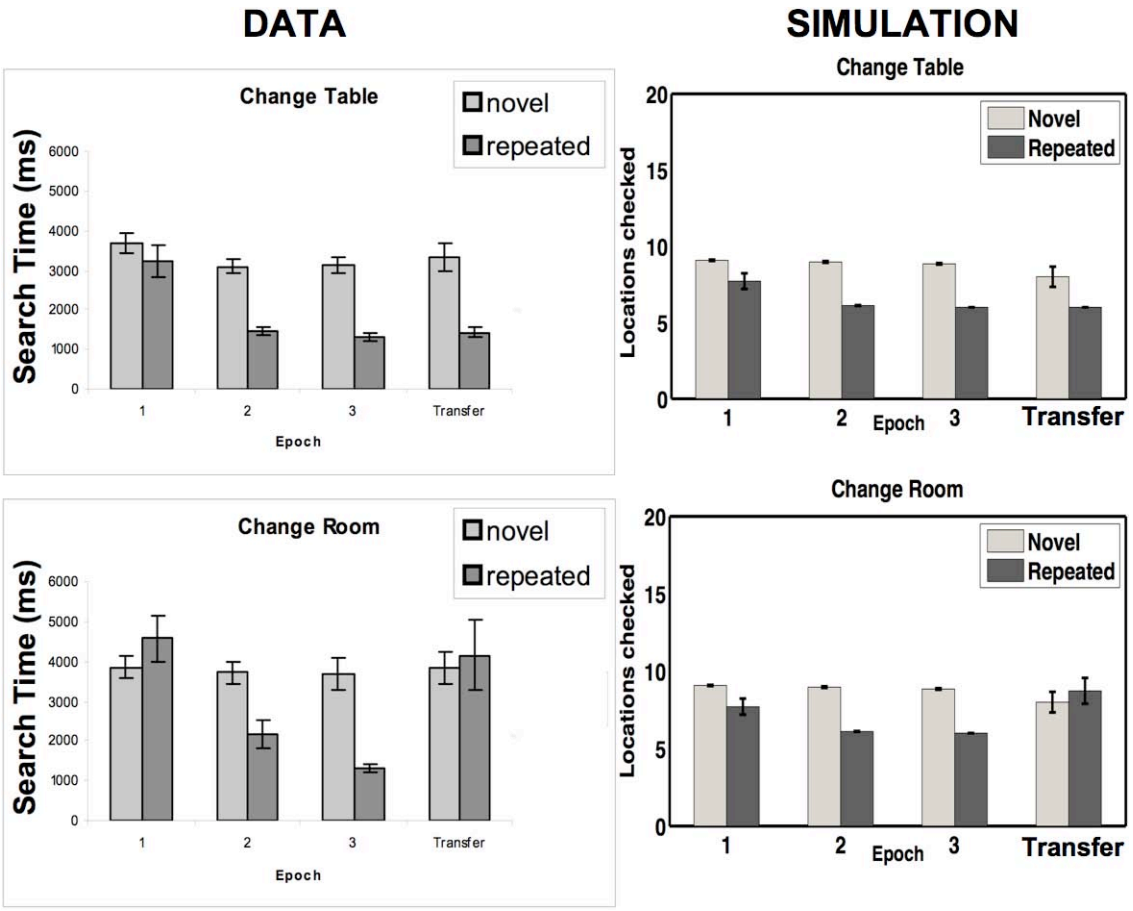


Figure 15. Spatial cueing effects due to target-predictive global/distant contexts. The graphs show RTs in the three learning epochs and the transfer block during which subjects searched for a target letter on a table (i.e., the local context) in a furnished room (i.e., the global/distant context) but with either the familiar table or room changed (upper and lower panels, respectively). The x-axis represents training epoch grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time for completing a trial. (Data reprinted with permission from Brockmole et al., 2006, Experiment 1).

There are differences between local and distant cueing experiments that may account for the seemingly conflicting observations. Unlike Brady and Chun (2007) or Olson and Chun (2002) who used discrete letter displays, Brockmole et al. (2006) employed naturalistic scenes as search displays, which are much richer inputs to the visual 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 attended visual search change in response to varying visual stimuli. On the other hand, if the same search mechanisms are assumed, the data of Brockmole et al. (2006) are then 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 salient contexts still outweigh

local contexts even under the peripheral down-regulation of spatial attention, and remain viable for long-term memory encoding.

ARTSCENE Search reconciles local and distant cueing under the same framework using saliency and attention-dependent context learning. In the model, the relative saliency of each location in the search display is maintained in the short-term memory of model parahippocampal cortex (Figures 5a and 5b). During learning, the strength of associative weights between model parahippocampal cortex and dorsolateral PFC is updated proportionally to the location saliency represented in model parahippocampal cortex. Accordingly, attentionally salient spatial cues, no matter where they are in a scene, will be encoded into context memory more strongly than less salient counterparts, and become an effective spatial context of that scene. For simplicity and ease of comparison with other data simulations in this article, the simulation of distant cueing was carried out in discrete search displays with point objects (see Appendix) rather than the naturalistic images used in the original study.

6.9 Semantic Object Cueing

Familiar object contexts can also facilitate visual search. Chun and Jiang (1999) found that target search among a set of co-occurring objects became faster even when the spatial arrangement of those objects changed from repetition to repetition. In their object cueing experiment, a target is a shape symmetric around the vertical axis in the search display. In the 'Old' object context condition, a target was consistently paired with a specific set of distractors across blocks. In the 'New' object context condition, a target was randomly paired with various sets of distractors. The locations of all objects in the search display were randomized for every trial in a block. The cueing effects from regular object contexts are shown in Figure 16 where the x-axis represents training epochs grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time for completing a trial. Interestingly, the RT curves of both conditions in Figure 16 resemble the ones seen from spatial cueing effects in Figure 10. In particular, the RT in the 'New' object context condition dropped across epochs, and a further RT reduction in the 'Old' object context condition also developed as the session progressed.

As reflected by the similarity between spatial and object cueing effects, ARTSCENE Search proposes an object cueing process in the model What stream (Figure 5d) that is a homolog of the spatial cueing mechanism in the model Where stream (Figure 5b). Specifically, in the model, a spatial context is decomposed as pairs of correlated *locations*, whereas an object context is treated as pairs of correlated object *identities*. However, unlike those parallel processes in the model Where stream such as spatial gist formation and priming, object cueing in the model What stream is inherently a sequential process requiring 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 model posterior parietal cortex. The corresponding object is then recognized in model anterior inferotemporal cortex and stored temporarily in model perirhinal cortex to initiate formation of an object context. If the foveated object is not a target, spatial attention at the selected location is then gradually disengaged due to inhibition of return and a new selection cycle resumes.

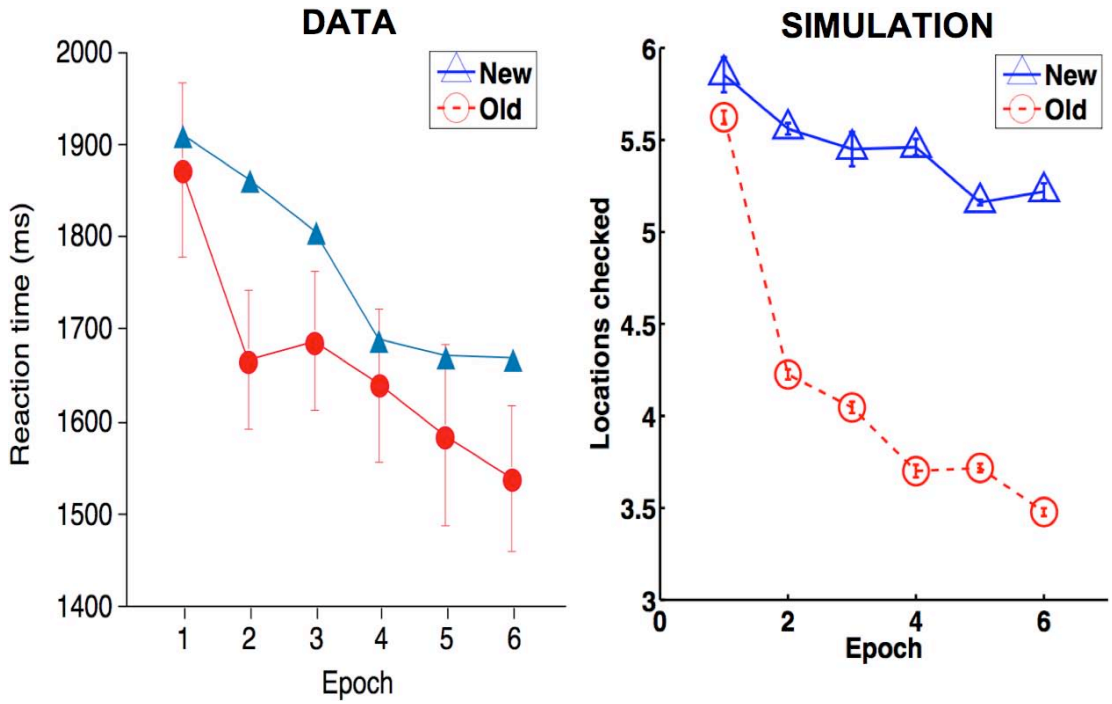


Figure 16. Object cueing effects are the RT reductions for search in a congruent or familiar object context (i.e., 'Old' condition) compared to an incongruent or a novel context (i.e., 'New' condition). The x-axis represents training epoch grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time for completing a trial. (Data reprinted with permission from Chun, 2000).

During each target selection cycle, a distractor is a new piece of evidence for correlated target identities, and associative contextual votes are carried out through the learned weights from model perirhinal cortex to ventral prefrontal cortex (VPFC). As the expectation of target identities is built up from the current object context, model ventral PFC feeds back to anterior IT cortex (ITa), which in turn feeds back to V4 and posterior IT cortex (ITp) to enhance target-like features in space (see Figure 5e). Consistent with data of Vickery, King, and Jiang (2005), the implemented top-down knowledge in visual search primes not only target categories but also visual features of target prototypes. Accordingly, the probability of fixations on salient distractors is reduced, reflected as object cueing effects. Importantly, while an old object context primes a specific target identity (e.g., a butterfly) in model ventral PFC, a new object context that has been randomly paired with different targets in the past non-specifically primes all paired target identities (e.g., objects symmetric around the vertical axis), which then prevents exhaustive search (e.g., objects both symmetric and asymmetric around the vertical axis) and causes search reaction time to drop as a session progresses.

6.10. Top-Down Modulations in Contextual Cueing

The efficacy of a contextual cue in spatial cueing is determined not only by bottom-up factors such as saliency (Section 6.8), but also by top-down attentional modulation, as observed by Jiang and Chun (2001). In their study, a search display contained an equal number of red and green items. Participants were instructed to search through items that shared the target color, which is fixed as green or red for each participant throughout the whole experiment. Figure 17 shows the results of this manipulation. In the figure, the x-axis represents search epochs grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time (RT) for completing

a trial. Each chosen target location was presented in one trial per block within which four conditions were intermixed. Across blocks, the spatial configuration of distractors in a trial was randomly varied in the 'Control' condition, but fully repeated in the 'Both-old' condition. The 'Ignored-old' and 'Attended-old' conditions preserved spatial locations across blocks for distractors in the ignored or attended color, respectively. In Figure 17, the 'Attended-old' condition led to faster search than the 'Ignored-old' condition. In other words, the efficacies of contextual cues in spatial cueing were strengthened by modulation from top-down feature-based attention.

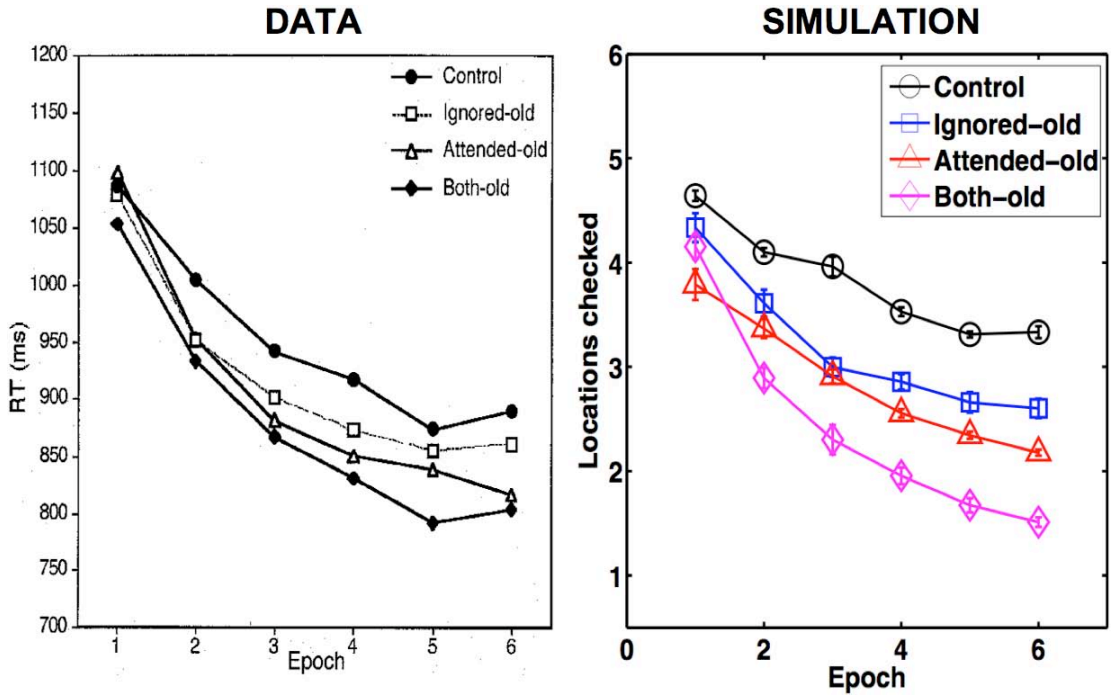


Figure 17. Selective feature-based attention modulates contextual cueing. In the experiment and simulation, a search trial consisted of red and green items including the target whose color was maintained and attended to throughout the entire session. Across blocks, the spatial configuration of distractors in a trial was randomly varied in the 'Control' condition, but fully repeated in the 'Both-old' condition. The 'Ignored-old' and 'Attended-old' conditions preserved spatial locations across blocks for distractors in the ignored or attended color, respectively. The x-axis represents training epoch grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time for completing a trial. (Data reprinted with permission from Jiang & Chun, 2001, Experiment 3).

ARTSCENE Search simulates such attentional modulation via the What-Where interaction in the model, channeled by the pathway in Figure 5e from model anterior inferotemporal (IT) cortex, through posterior IT cortex/V4, to posterior parietal cortex (PPC). Such What-Where interaction occurs when a subject holds an expectation of target features in mind. For example, expected target identities (e.g., red items) in model ventral prefrontal cortex (PFC) prime position-invariant object categories in model anterior IT cortex, which in turn prime spatial maps in model posterior IT cortex and V4 that code these target features (e.g., the redness map) by a non-specific boost of all these featural representations. In this way, items that share target features such as color at any location in a scene are primed as candidate targets. Then, when boosted by a bottom-up featural input to a particular location, such What modulation in

posterior IT cortex and V4 propagates to the Where pathway including the spatial priority map in model PPC, the control of eye movements in model superior colliculus (SC; see Figure 5c), and beyond to the short-term memory of spatial contexts in model parahippocampal cortex, and the long-term memory of such spatial contexts in association with a target stored in dorsolateral PFC (see Figure 5b). As a result, attended locations are encoded more strongly than ignored location as contextual cues in spatial cueing effects.

6.11. Target-Based Attention Allocation in Context Learning

As a consequence of attention-gated context learning (Sections 6.8 and 6.10), equally attended contexts are equally effective for spatial cueing. With a design similar to the one used by Jiang and Chun (2001), Olson and Chun (2002, Experiment 4) found that color did not modulate the effectiveness of a spatial context when color was not a predictive feature for the target. Their results are shown in Figure 18 where the x-axis represents search epochs grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time (RT) for completing a trial. Notice that the conditions 'Control', 'Ignored-old', and 'Attended-old' in Figure 17 used the same manipulations as in the conditions 'New', 'Old-Oppose', and 'Old-Match' in Figure 18, 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 and 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 and Chun (2002) had targets randomly colored in red or green on half of the trials so that color was an uninformative and ineffective cue for search guidance, and participants paid attention to both green and red items.

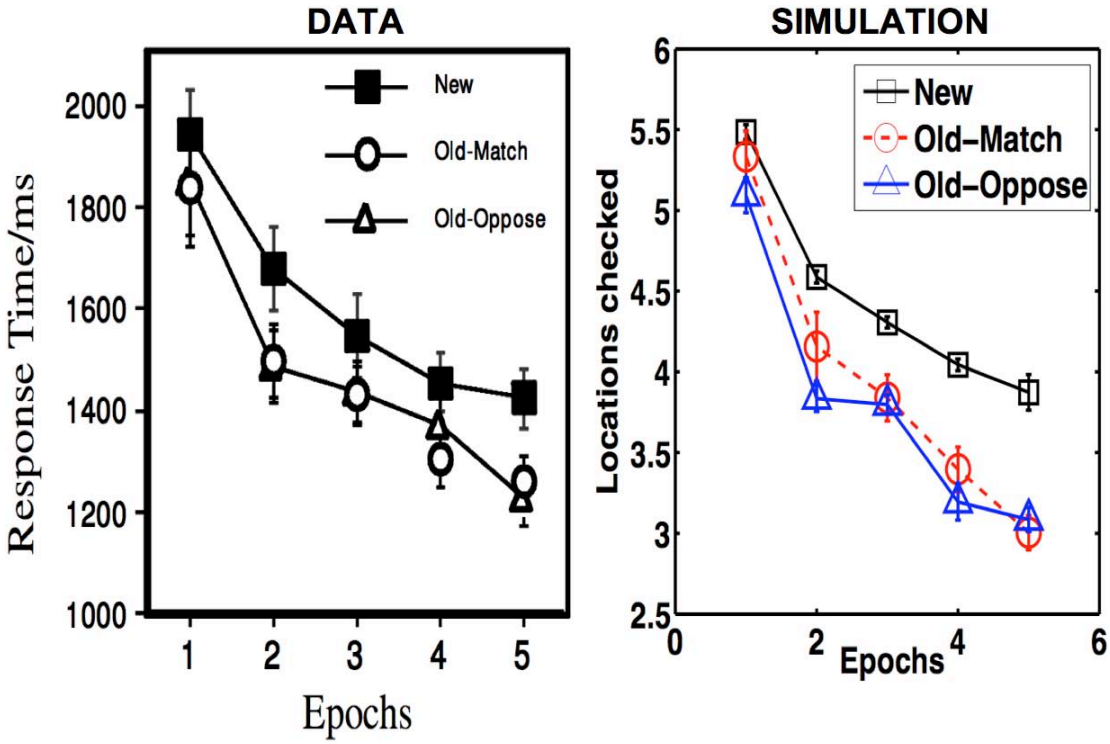


Figure 18. Task-irrelevant colors did not affect spatial cueing. In the experiment and simulation, the target color was non-predictable, either red or green. The context layouts were varied in the 'New' condition, but preserved across blocks for half of the items that shared the target color in the 'Old-Match' condition. In contrast, the 'Old-Oppose' condition preserved locations for half

items that differed in color from the target. The x-axis represents training epoch grouped from blocks of trials (see Table 2), and the y-axis represents search reaction time for completing a trial. (Data reprinted with permission from Olson & Chun, 2002, Experiment 4).

ARTSCENE Search simulates the data from Olson and Chun (2002, Experiment 4) using the same mechanism that was described in Section 6.10. Specifically, the model What system not only regulates object cueing but also interacts with the model Where system to modulate spatial cueing via feature-based attention (see Figures 5d, 5e, and 5b in order). When targets are equally likely to be red or green, all red and green object representations in model ventral prefrontal cortex are then equally active. Therefore, all red and green items in the search display are primed simultaneously, so that color plays no role in search guidance.

6.12. Integration of Spatial and Object Cueing

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 targets acquired from spatial or object cueing alone, as shown in Figure 19; cf., Figures 8 and 9. In this model simulation of What-Where interactions, targets and distractors were two disjoint sets of objects. Each target was assigned to one of the following four conditions, which were intermixed in a block. In the 'Object Cueing Alone' condition, all items in the search display were relocated on each new trial, and a target was paired with a fixed set of context objects, together presented in one trial per block. In the 'Spatial Cueing Alone' condition, a target location was paired with a fixed set of context locations, repeated in one trial per block. However, the objects in all locations of this invariant spatial configuration varied across repetitions. In the 'Spatial+Object Cueing' condition, the same search display was repeated in one trial per block. Conversely, in the 'No Regular Contexts' condition, no spatial and object regularity was paired with a particular target except the fact that the target set, as a whole, repeated across blocks. In Figure 19, the x-axis represents the search trials grouped into blocks, and the y-axis corresponds to search reaction time (RT) for completing a trial. Clearly, the 'Spatial+Object Cueing' condition led to the greatest search RT benefits. This simulation result is consistent with behavioral data from experiments using other conditions (e.g., Endo & Takeda, 2004, Experiment 3) and designs (e.g., Gronau, Neta, & Bar, 2008) showing that task reaction times were the lowest among conditions when visual stimuli were both spatially and semantically related.

In terms of search dynamics, the further RT reductions from the 'Spatial Cueing Alone' condition to the 'Spatial+Object Cueing' condition in Figure 19 suggests that the global-to-local and spatial-to-object evidence accumulation in ARTSCENE Search is an effective strategy for target search. In each simulated trial of Figure 19, ARTSCENE Search first engages the spatial cueing circuit (Figure 5b) to generate expectations of target locations based on the bottom-up inputs of spatial gist. The top-down spatial priming for targets in the Where stream is then integrated with bottom-up signals in the spatial priority map. Then, the expectations of target identities and features are also developed in the object cueing circuit (Figure 5d). The top-down object priming for targets from the What stream further gain-modulates the bottom-up signals sent to the spatial priority map in model posterior parietal cortex (Figure 5e). From the perspective of evidence accumulation, in the early search phase before any eye movements, the model first distributes spatial attention across the visual field to apprehend the spatial gist of the search display, which rapidly gives rise to a first-order estimate of where a target may be located. Later in a search when an object context is gradually formed during subsequent eye movements,

the initial hypothesis of target locations is incrementally refined after each eye fixation when the object at a fixated location can be further identified using focal attention.

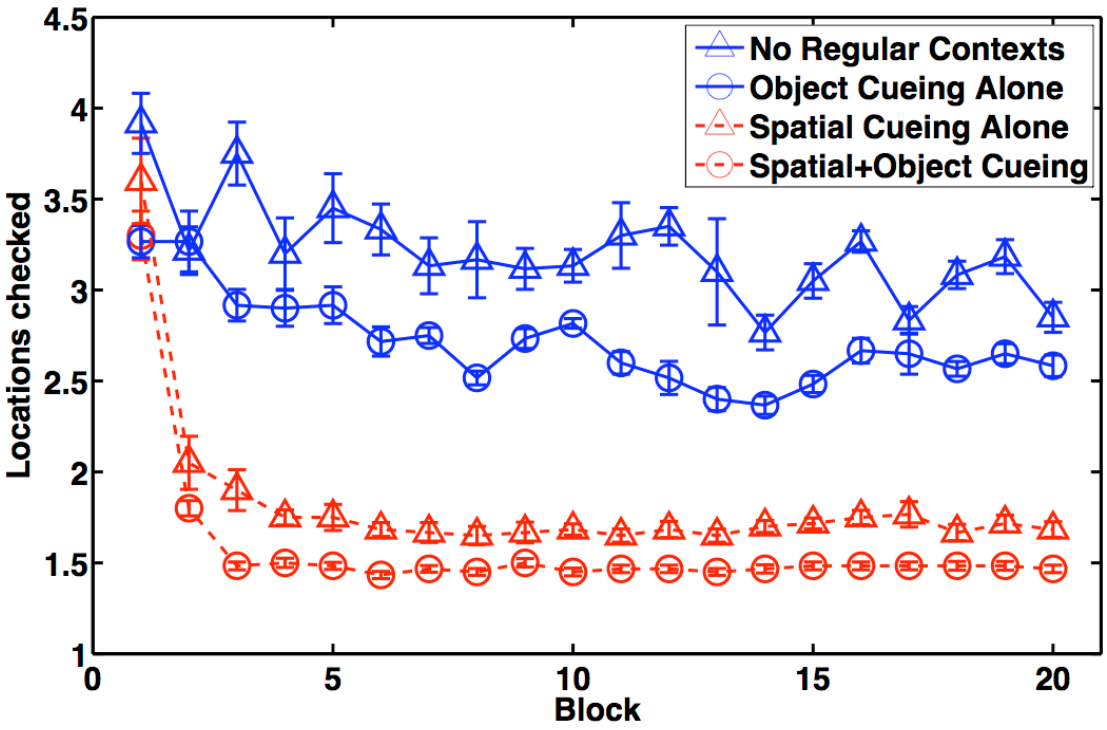


Figure 19. Integrated contextual cueing effects. The simulation used a with-in subjects design to show that spatial-plus-object regularities in the search set reduced RT more than the conditions where only spatial or object information was target-predictive. In the graph, the x-axis represents training block (see Table 2), and the y-axis represents search reaction time for completing a trial.

With regard to What-Where interactions, five properties of ARTSCENE Search are worth noting. First, where to look determines what to see. In addition to the explicitly discussed What-to-Where modulation in feature-based attention, the Where-to-What modulation in the model is simply the process of object recognition following spatial selection.

Second, 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 have an effect during object cueing because a target is always the last fixated object in a search trial. This asymmetry between the target location and identity in contextual cueing may account for the observation that search reaction times decreased when target locations were fixed and consistently paired with certain distractor identities, but not when target identities were fixed and consistently paired with certain distractor configurations (Endo & Takeda, 2004, Experiment 4).

Third, 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 (see Section 6.2 and the cell activities of model PHC in Figures 7-9), but local object cues are later accumulated in a sequence of eye fixations.

Fourth, given additional sources inputting to model ventral prefrontal cortex, top-down feature-based attention along the model What pathway can gain-modulate bottom-up scene

percepts in the Where pathway to form an effective spatial context in model posterior parietal and parahippocampal cortex before eye movements and object cueing occur.

Fifth, the strength of contextual spatial priming from model dorsolateral prefrontal cortex is commensurate with the degree of match between short-term and long-term memory representations of spatial cues from parahippocampal cortex. Since attention modulates representations of spatial cues, an unattended spatial context is registered weakly in short-term memory, and not well matched with the learned ones in long-term memory. Hence, spatial cueing effects are small when target-predictive spatial contexts are ignored. One such example is the latent learning phenomenon in which consistent yet ignored spatial cues can barely reduce search RTs during training, but suddenly become effective when attended to during testing. On the other hand, consistent and attended spatial cues can reduce search RTs during training, but suddenly become less effective when ignored during testing (Jiang & Leung, 2005).

To summarize, although V4 and posterior inferotemporal (ITp) cortex are the only regions in the model that directly interface model What and Where streams (Figures 5c and 5e), What-Where interactions are ubiquitous through all model regions and computations as emergent system properties. Notably, this link between the What and Where streams in the model is sufficient to enable the model to simulate all the behavioral data presented in Section 6.

7. Discussion

7.1. Model lesions

The functional role of each model region in ARTSCENE Search can be understood via Figure 5. Lesion of a particular model region will impair the corresponding processing in the model. However, the impaired model functions or properties may or may not be necessary to carry out a given task.

Model PPC and SC play a central role in the model. Model PPC is involved in bottom-up attention (Figure 5a), top-down spatial attention (Figure 5b), and top-down feature-based attention (Figure 5e). Model SC controls fixation shifts (Figure 5c). Lesioning model PPC or SC will remove all model outputs.

Model V1/V2 and V4/ITp are essential front-ends for visual inputs. Lesion of model V1/V2 cuts out all bottom-up inputs (Figure 5a) but does not impair willful eye movements (Figure 5c) such as the ones during visual imagery or eye closure. Removal of model V4/ITp disconnects What and Where steams in the model (Figure 4), severely impairing bottom-up attention (Figure 5a), and featural selection for recognizing (Figure 5c) or locating (Figure 5e) an object.

Model PHC and DLPFC are crucial components in top-down spatial attention. Damage of either model area impairs spatial contextual cueing (Figure 5b), but not object contextual cueing (Figure 5d). Note, however, that lesion of model PHC alone does not completely disable top-down spatial priming, which is ultimately commanded by model DLPFC.

Model PRC and VPFC are responsible for top-down object-based attention. Removal of either model area impairs object contextual cueing (Figure 5d), but not spatial contextual cueing (Figure 5b). Nonetheless, lesion of model PRC alone does not totally disable top-down object priming, which is under the direct control of model VPFC.

Finally, model ITa is engaged heavily in all processes along the ventral What stream. Loss of model ITa affects object recognition (Figure 5c), object contextual cueing (Figure 5d), and top-down feature-based attention in general (Figure 5e). However, as long as the dorsal

Where stream remains intact, the model is still able to learn and retrieve spatial contexts based on the spatial gist of a scene (Figures 5a and 5b).

7.2. Comparison with other theories of visual attention and visual search

ARTSCENE Search employs the principle of global-to-local visual processing, embodies biologically plausible neural mechanisms, and is capable of guiding attention deployment for the most efficient target search based on all the spatial and object regularities in a scene. Under the same framework, the model offers an integrated explanation of challenging behavioral data of positive/negative, spatial/object, and local/distant cueing effects during visual search. Table 1 compares properties of ARTSCENE Search with those of other theories of visual attention and search.

ARTSCENE Search reduces to simpler search models given specific conditions. Without any top-down expectations, ARTSCENE Search may be compared to feature integration or saliency map models (e.g., Itti & Koch, 2000; Koch & Ullman, 1985; Li, 2002; Treisman & Gelade, 1980) that determine eye-scan paths based purely on bottom-up location saliency. When specific target features are expected before search, ARTSCENE Search functions like appearance-guided search models (e.g., Navalpakkam & Itti, 2005; Rao et al., 2002; Treisman & Sato, 1990; Wolfe, 1994; Zelinsky, 2008). 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 contextual search models (e.g., Backhaus et al., 2005; Brady & Chun, 2007; Torralba et al, 2006). As a synthesis of these established 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, and various search asymmetries, as discussed earlier in Section 2. ARTSCENE Search, however, differs greatly from other search models in several key respects, as discussed below.

In terms of search guidance, top-down priming is influenced by learning and integrated with stimulus-driven attention in ARTSCENE Search. In some attention theories, although top-down guidance for visual search is considered, it is supplied as *a priori* knowledge rather than knowledge acquired through learning (e.g., Bundesen et al., 2005; Logan, 1996; Rao et al, 2002; Treisman & Sato, 1990; Wolfe, 1994; Zelinsky, 2008). In the extreme, regions of interest for gaze deployment in some models (e.g., Ballard & Hayhoe, 2009; Rao et al, 2002; Zelinsky, 2008) are primarily determined by top-down factors such as target appearance or task-dependent reward expectation (see also Reichle & Laurent, 2006, for discussions of reinforcement learning in reading). Consequently, this group of goal-driven models cannot allocate attention in a purely bottom-up manner based on intrinsic saliencies of items or locations. In contrast, ARTSCENE Search can direct attention with or without learned top-down priming, accounting for a broad spectrum of visual search behavior.

In contrast to purely psychological theories (e.g., Backhaus et al., 2005; Ballard & Hayhoe, 2009; Brady & Chun, 2007; Duncan & Humphreys, 1989; Humphreys & Müller, 1993; Itti & Koch, 2000; Logan, 1996; Rao et al., 2002; Thornton & Gilden, 2007; Torralba et al., 2006; Treisman & Gelade, 1980; Treisman & Sato, 1990; Wolfe, 1994; Zelinsky, 2008), ARTSCENE Search is a neuropsychological model, which makes specific predictions about what cortical areas and brain mechanisms underlie human search behavior. In particular, the neural mechanisms of 'biased competition' and spatial/featural priming (Carpenter & Grossberg, 1987, 1991; Desimone & Duncan, 1995) are implemented in model cells by a shunting recurrent on-center off-surround equation (e.g., Equations 1, 8 and 17), which allows the model to

normalize attentional weights and bias spatial selection toward an object (cf., 'filtering' in Bundesen et al, 2005). Moreover, model neuronal feature responses can be gain-modulated via top-down feature-based attention (cf., 'pigeonholing' in Bundesen et al, 2005) in the model V4 and posterior inferotemporal cortex (ITp; see Equations 5 and 6).

Authors	Uses continuous stimuli as inputs	Models functional field of view or attention	Models bottom-up saliency-based attention	Models top-down spatial attention	Models top-down object-based or feature-based attention	Learns regular spatial contexts	Learns regular object contexts	Offers neural interpretations
Treisman and Gelade (1980)	No	No	Yes	No	No	No	No	No
Koch and Ullman (1985)	No	No	Yes	No	No	No	No	Yes
Duncan and Humphreys (1989)	No	Via perceptual grouping	Via perceptual grouping	No	No	No	No	No
Treisman and Sato (1990)	No	Yes	Yes	No	Yes	No	No	No
Humphreys and Müller (1993)	No	Via perceptual grouping	Via perceptual grouping	No	No	No	No	No
Grossberg et al. (1994)	No	Via perceptual grouping	Implemented algorithmically	Implemented algorithmically	Implemented algorithmically	No	No	Yes
Wolfe (1994)	No	No	Yes	No	Yes	No	No	No
Desimone and Duncan (1995)	No	No	Proposed but not implemented	Proposed but not implemented	Proposed but not implemented	No	No	Yes
Logan (1996)	No	Via perceptual grouping	Yes	Yes	Yes	No	No	No
Findlay and Walker (1999)	No	Proposed but not implemented	Proposed but not implemented	Proposed but not implemented	Proposed but not implemented	No	No	Yes
Itti and Koch (2000)	Yes	Yes	Yes	No	No	No	No	No
Li (2002)	No	No	Yes	No	No	No	No	Yes
Rao et al. (2002)	Yes	Yes	No	No	Yes	No	No	No
Backhaus et al. (2005)	No	Yes	No	Yes	No	Yes	No	No
Bundesen et al. (2005)	No	No	Yes	Yes	Yes	No	No	Yes
Navalpakkam and Itti (2005)	Yes	Yes	Yes	No	Yes	No	Pre-specified	Yes
Torralba et al. (2006)	Yes	No	Yes	Yes	No	Yes	No	No
Brady and Chun (2007)	No	Yes	Iso-saliency map	Yes	No	Yes	No	No
Thornton and Gilden (2007)	No	No	Iso-saliency sampling	No	No	No	No	No
Zelinsky (2008)	Yes	Yes	No	No	Yes	No	No	No
Ballard and Hayhoe (2009)	Yes	No	No	Via reward-based mechanisms	Via reward-based mechanisms	Via reward-based mechanisms	Via reward-based mechanisms	No
Huang and Grossberg (Our model)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 1. Properties of several visual attention and visual search models.

Beyond functional specification, ARTSCENE Search further simulates the neural dynamics of each model region using differential equations, which enable the model to emulate aspects of brain dynamics in real time. In particular, the model simulates how various brain regions in the cortical What and Where pathways, including subregions of the medial temporal lobe and the prefrontal cortex, may 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 global-to-local evidence accumulation through eye movements, ARTSCENE Search implements eye fixations as a series of information gathering events by which the likelihood of seeing a target at every location, or the saccadic plan, can be dynamically revised in the model posterior parietal cortex 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 (e.g., Brady & Chun, 2007; Itti & Koch, 2000; Koch & Ullman, 1985; Navalpakkam & Itti, 2005; Torralba et al., 2006; Treisman & Gelade, 1980; Treisman & Sato, 1990; Wolfe, 1994), and extends more abstract search models that illustrate evidence accumulation using diffusion or race mechanisms (e.g., Bundesen et al., 2005; Logan, 1996; Thornton & Gilden, 2007).

With regard to eye movements, saccade generation in ARTSCENE Search is much more simplified than in models that are devoted to explaining the oculomotor system (e.g., Brown et al., 2004; Gancarz & Grossberg, 1999; Grossberg et al., 1997; Srihasam, Bullock, & Grossberg, 2009) or gaze deployment during reading (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Heinzle, Hepp, & Martin, 2010; McDonald, Carpenter, & Shillcock, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998). In the current implementation, the predictions of gaze locations from ARTSCENE Search do not attempt, for example, to simulate the center-of-gravity phenomenon where a saccade directed to two modestly separated targets often lands at an intermediate location between the two (e.g., Findlay & Walker, 1999; Rao et al, 2002; Zelinsky, 2008). Similarly, unlike models that operate on natural images or continuous stimuli (e.g., Ballard & Hayhoe, 2009; Itti & Koch, 2000; Navalpakkam & Itti, 2005; Rao et al., 2002; Zelinsky, 2008), ARTSCENE Search does not currently address why visual fixations sometimes land on regions of a scene background rather than on objects, although the model may be naturally extended to process figure-ground properties of extended objects that partially occlude each other in a 3-dimensional (3D) scene. Such model development will require an elaboration of model region V1/V2 and V4/ITp (Figures 4 and 5a) that incorporates recent models of 3D viewing, figure-ground separation, and invariant object category learning and recognition (e.g., Fazl et al., 2009; Grossberg, 1994; Grossberg & Yazdanbakhsh, 2005).

7.3. Concluding Remarks

The ARTSCENE Search model clarifies how the brain combines locally uncertain combinations of scenic information, through learning, into predictive decisions 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 anterior inferotemporal cortex (ITa), perirhinal cortex (PRC), and ventral prefrontal cortex (VPFC) in the What stream, and posterior parietal cortex (PPC), parahippocampal cortex (PHC), and dorsolateral prefrontal

cortex (DLPFC) in the Where stream. In a like manner, Grossberg and Pilly (2008) developed a detailed neural model of how primary visual cortex (V1), middle temporal area (MT), medial superior temporal area (MST), lateral intraparietal area (LIP), and the basal ganglia (BG) interact to make eye movements to the direction of coherent motion in noise. Their model quantitatively simulated all the key data properties of the psychophysical and neurophysiological experiments of Roitman and Shadlen (2002) and Shadlen and Newsome (2001) in response to probabilistically defined motion stimuli. Compared to Bayesian models that generally sketch probabilistic relations among observables, neural models further delineate how specific brain cells, circuits, and systems embody particular principles and mechanisms to carry out perceptual decisions and actions. Such a neural level of understanding helps uncover brain organizational properties that are not easily revealed by Bayesian or other statistical approaches (for more discussions on Bayesian inference in the brain, see Grossberg & Pilly, 2008).

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 with point objects. As noted in Section 7.2, for the model design to work on real images, future work needs to incorporate mechanisms of visual boundary and surface processing which clarify how the brain accomplishes 3D vision, figure-ground separation, and invariant object category learning and recognition. In particular, the FACADE and 3D LAMINART models clarify how 3D boundary and surface representations of object and their backgrounds form (e.g., Grossberg, 1994; Grossberg & Yazdanbakhsh, 2005). The ARTSCAN model (Fazl et al., 2009) clarifies how invariant recognition categories of extended objects can be learned as attention shifts and eye movements scan a scene. In particular, surface surface-fitting 'attentional shrouds' mold spatial attention into the shape of an attended object to regulate view-invariant and position-invariant object category learning and recognition. Such spatial-to-object attentional interactions show how we can attend to objects even before they are recognized (Walther & Koch, 2006). ARTSCAN learns to categorize spatially extended objects using Adaptive Resonance Theory, or ART, circuits (Carpenter & Grossberg, 1987, 1991; Grossberg, 1980, 1999). Likewise, the ARTSCENE model uses ART to classify different scenes as multi-scale textures, including the gist of a scene (Grossberg & Huang, 2009).

Future work needs to unify the competences embodied within the 3D LAMINART (Cao & Grossberg, 2005; Fang & Grossberg, 2009; Grossberg & Howe, 2003; Grossberg & Raizada, 2000; Grossberg & Yazdanbakhsh, 2005), ARTSCAN (Fazl et al., 2009), ARTSCENE (Grossberg & Huang, 2009) and ARTSCENE Search models. In such an extended model, the treatment of spatial coordinates needs to be enriched. All model regions in the Where stream of ARTSCENE Search encode positions in egocentric/spatiotopic coordinates. In vivo, a mix of reference frames is needed. Retinotopic organizations may be found in visual areas V1-V4 (Fize et al., 2003) and superior colliculus (Schneider & Kastner, 2005). Posterior parietal neuronal responses are gain-modulated by eye, hand, head and body positions, which are thought to mediate transformations between retinotopic and egocentric coordinates (Cohen & Andersen, 2002; Fazl et al., 2009; Pouget & Snyder, 2000). Parahippocampal cortex is more egocentric than retinotopic (Epstein, 2008). Spatial representations of targets in dorsolateral PFC neurons are more egocentric/spatiotopic, which may be converted by FEF to retinotopic coordinates for oculomotor planning or execution (Funahashi, Bruce, & Goldman-Rakic, 1989; Grossberg et al., 1997). In addition to spatial coordinates, future work needs to address why a learned but moderately rescaled spatial configuration can reduce search reaction times (Jiang & Wagner, 2004, Experiment 2), which may involve how multiple-scale filters (e.g., Cohen & Grossberg,

1987; Grossberg & Huang, 2009) interact with eye movement circuits to make the best prediction using all the regular information in a scene.

Further What and Where interactions may be considered. Jiang & Song (2005) showed that the expression of spatial cueing can either be identity-independent or identity-contingent given different training procedures (e.g., all items were white in training trials vs. all items were either white or black in half of the training trials). Endo and 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, spatiotemporal integration of cues may play a role in contextual guidance. 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 and Kawahara (2005) reported that target-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.

Finally, incentive motivational processing may be integrated with contextual cueing. Ventral prefrontal cortex (area 47/12) in ARTSCENE Search overlaps the orbitofrontal cortex (see review in Kringelbach, 2005), which is implicated in regulating motivational and emotional processing in conjunction with the amygdala (Dranias, Grossberg, & Bullock, 2008; Ghashghaei & Barbas, 2002; Grossberg, 2000b; Grossberg, Bullock, & Dranias, 2008; Schoenbaum, Setlow, Saddoris, & Gallagher, 2003). This additional circuit can drive ventral prefrontal cortex using motivational signals from the amygdala 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 occurred when repeated context blocks were presented *before* novel context blocks, but vanished when repeated context blocks were presented *after* novel context blocks (Jungé, Scholl, & Chun, 2007).

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.

References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9*, 357-381.

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.

Armony, 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.

Ballard, D. H., & Hayhoe, M. M. (2009). Modeling the role of task in the control of gaze. *Visual Cognition, 17*, 1185-1204.

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.

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of effective vision in visual search. *Perception & Psychophysics, 62*, 576-585.

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.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology, 103*, 597-600.

Bradski, G., Carpenter, G.A., & Grossberg, S. (1994). STORE working memory networks for storage and recall of arbitrary temporal sequences. *Biological Cybernetics, 71*, 469-480.

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., Castelhano, 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.

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112*, 291-328.

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.

Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision, 38*, 515-578.

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, MA: 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*, *9*(8):173, 173a, http://journalofvision.org/9/8/173/, doi:10.1167/9.8.173.

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., & Jiang, Y. (2003). Implicit, long-term spatial context memory. *Journal of Experimental Psychology: Learning, Memory, Cognition, 29,* 224-234.

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, Y. E., & Andersen, R.A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Review Neuroscience, 3*, 553-562.

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'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7*, 415-423.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222.

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.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777-813.

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.

Epstein, R. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences, 12*, 388-396.

Fang, L., & Grossberg, S. (2009). From stereogram to surface: How the brain sees the world in depth. *Spatial Vision, 22*, 45-82.

Fazl, A., Grossberg, S., & Mingolla, E. (2009). Position-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology, 58*, 1-48.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision*. New York: Oxford University Press.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences, 22*, 661-674.

Fize, D., Vanduffel, W., Nelissen, K., Denys, K., Chef d'Hotel, C., Faugeras, O., & Orban, G. (2003). The retinotopic organization of primate dorsal V4 and surrounding areas: A functional magnetic resonance imaging study in awake monkeys. *Journal of Neuroscience, 23*(19), 7395-7406.

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.

Gancarz, G., & Grossberg, G. (1999). A neural model of saccadic eye movement control explains task-specific adaptation. *Vision Research, 39*, 3123-3143.

Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision, 6*(9):1, 858-873, http://journalofvision.org/6/9/1/, doi:10.1167/6.9.1.

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.

Gitelman, D. R., Nobre, A. C., Parrish, T. B., LaBar, K. S., Kim, Y. H., Meyer, J. R., & Mesulam, M.-M. (1999). A large-scale distributed network for covert spatial attention. *Brain, 122,* 1093-1106.

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.

Greene, H. H. (2008). Distance-from-target dynamics during visual search. *Vision Research, 48*, 2476-2484.

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 & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233-374). New York, NY: Academic Press.

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. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics, 55*, 48-120.

Grossberg, S. (1999). The link between brain learning, attention, and consciousness. *Consciousness and Cognition, 8*, 1-44.

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., & Howe, P. D. L. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research, 43*, 801-829.

Grossberg, S., & Huang, T.-R. (2009). ARTSCENE: A neural system for natural scene classification. *Journal of Vision, 9*(4):6, 1-19, http://journalofvision.org/9/4/6/, doi:10.1167/9.4.6.

Grossberg, S., & Kuperstein, M. (1986). *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements*. Amsterdam, North-Holland: Elsevier.

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.

Grossberg, S., Roberts, K., Aguilar, M., & Bullock, D. (1997). A neural model of multimodal adaptive saccadic eye movement control by superior colliculus. *Journal of Neuroscience, 17*, 9706-9725.

Grossberg, S., & Yazdanbakhsh, A. (2005). Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vision Research, 45*, 1725-1743.

Hayhoe M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*(4), 188-193.

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience, 9*, 467-479.

Heinzle, J., Hepp, K., & Martin, K. A. C. (in press). A biologically realistic cortical model of eye movement control in reading. *Psychological Review*.

Hikosaka, O., & Wurtz, R. H. (1989). The basal ganglia. In R. Wurtz, & M. Goldberg (Eds.), *The neurobiology of saccadic eye movements* (pp. 257 – 281). Amsterdam: Elsevier.

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. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489-1506.

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. (2005). Hyper-specificity in visual implicit learning: Learning of spatial layout is contingent on item identity. *Journal of Experimental Psychology: Human Perception & Performance, 31*(6), 1439-1448.

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, 9*(8):1180, 1180a, http://journalofvision.org/9/8/1180/, doi:10.1167/9.8.1180.

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.

Loschky, L. C., McConkie, G. W., Yang, J., & Miller, M. E. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition, 12*, 1057-1092.

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.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*, 578-586.

McDonald, S. A., Carpenter, R. H. S., & Shillcock, R. C. (2005). An anatomically constrained, stochastic model of eye movement control in reading. *Psychological Review, 112*, 814-840.

Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2009) Parafoveal magnification: Visual acuity does not modulate the perceptual span in reading. *Psychological Science, 20*(6), 721-728.

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.

Moore, T., & Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *Journal of Neurophysiology, 91*(1), 152-162.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782-784.

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.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353–383.

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.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during search. *Vision Research, 46*, 614-621.

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 391-399.

Niebur, E., & Koch, C. (1996). Control of selective visual attention: Modeling the 'Where' pathway. *Neural Information Processing Systems, 8*, 802-808.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (in press). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*.

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, 17*, 1244-1258.

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.

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision, 8*(14):21, 1-18, http://journalofvision.org/8/14/21/, doi:10.1167/8.14.21.

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.

Passingham, R. (1993). *The frontal lobes and voluntary action*. Oxford: Oxford University Press.

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 and performance X: Control of language processes* (pp. 531-556). Hillsdale, NJ: Erlbaum.

Posner, M. I., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160-174.

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.

Pouget, A., & Snyder, L. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience, 3*, 1192-1198.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). Cambridge, UK: Cambridge University Press.

Rao, R., Zelinsky, G., Hayhoe, M., & Ballard, D. (2002). Eye movements in iconic visual search. *Vision Research, 42*(11), 1447-1463.

Rayner, K. (2009). Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457-1506.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science, 206*, 468-469.

Reichle, E. D., & Laurent, P. (2006). Using reinforcement learning to understand the emergence of "intelligent" eye-movement behavior during reading. *Psychological Review, 113*, 390-408.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125-157.

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.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*(14):16, 1-20, http://journalofvision.org/7/14/16/, doi:10.1167/7.14.16.

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.

Saalmann, 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.

Saida, S., & Ikeda, M. (1979). Useful field size for pattern perception. *Perception & Psychophysics, 25*, 119-125.

Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience, 5*, 479-484.

Sanocki, T. (2003). Representation and perception of spatial layout. *Cognitive Psychology, 47*, 43-86.

Schall, J. D., & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience, 22*, 241-259.

Schneider, K. A., & Kastner, S. (2005). Visual responses of the human superior colliculus: A high-resolution fMRI study. *Journal of Neurophysiology, 94*, 2491-2503.

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.

Srihasam, K., Bullock, D., & Grossberg, S. (2009). Target selection by frontal cortex during coordinated saccadic and smooth pursuit eye movements. *Journal of Cognitive Neuroscience, 21*, 1611-1627.

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.

Strick, P. L., Dum, R. P., & Picard, N. (1995). Macro-organization of the circuits connecting the basal ganglia with the cortical motor areas. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 117 – 130). Cambridge, MA:

MIT Press.
Thompson, K. G., Biscoe, K. L., & Sato, T. R. (2005). Neuronal basis of covert spatial attention in the frontal eye field. *Journal of Neuroscience, 25*(41), 9479-9487.
Thornton, T. L., & Gilden, D.L. (2007). Parallel and Serial Processes in Visual Search. *Psychological Review, 114*(1), 71-103
Torralba, A., Oliva, A., Castelhano, M., & Henderson, J. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*, 766-786.
Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology, 25*, 168-199.
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. (2004). Oculomotor correlates of context-guided learning in visual search. *Perception & Psychophysics, 66*(8), 1363-1378.
Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology, 15*(1), 121-149.
Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In M. A. Ingle, M. A. Goodale, & R. Mansfield (Eds.), *Analysis of Visual Behaviour* (pp. 549-586). Cambridge, MA: MIT Press.
van Asselen, M., & Castelo-Branco, M. (2009). The role of peripheral vision in implicit contextual cuing. *Attention, Perception, & Psychophysics, 71*(1), 76-81.
van Diepen, P. M. J., & d'Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition, 10*, 79-100.
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):8, 81-92, http://journalofvision.org/5/1/8/, doi:10.1167/5.1.8.
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, NY: Plenum Press.
Zeki, S. (1996). Are areas TEO and PIT of monkey visual cortex wholly distinct from the fourth visual complex (V4 complex)? *Philosophical transactions of the Royal Society of London Series B Biological Sciences, 263*, 1539-1544.
Zelinsky, G.J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115*(4), 787-835.

Acknowledgments

This work was supported in part by CELEST, a National Science Foundation Science of Learning Center (NSF SBE-0354378) and HRL Laboratories LLC (subcontract #801881-BS under DARPA prime contract HR0011-09-C-0001). The authors wish to thank Chien-Hua Wang for valuable assistance in manuscript preparation, and Keith Rayner, Claus Bundesen, Kyle Cave, Erik Reichle, and two anonymous reviewers for their constructive comments on an earlier version of the article.

Appendix

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

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

In Equation 1, $X(t)$ is the neuron voltage; $\frac{dX(t)}{dt}$ is the rate at which $X(t)$ changes; parameter τ is the membrane capacitance and characterizes cell response time; parameter A_X is 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:

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

where the initial value, $X(0)$, is set or reset to zero for each search trial except for the long-term memory of contexts: W_{xyij}^{HD} in Equation 16 and W_{nm}^{RV} in Equation 24. The integration time step, ΔT, is 0.1 for all model equations. A smaller time step, ΔT, is computationally feasible for single-trial simulations, but not for the large-scale model simulations that are presented in Section 6, each of which searched through thousands of trials (see the number of subjects, blocks and trials per block in Table 2). However, the model results should hold for the simulations using a smaller ΔT since search reaction time is simulated by the number of inspected locations, which is not affected by the size of ΔT, or equivalently the number of iteration cycles for numerical integration.

The computational relations among the model variables are summarized in Figure 6. This circuit delineates model operations on a finer scale than its macroscopic version in Figure 4. Each variable in Figure 6 is discussed in detail below.

Stimuli

Each search trial is specified by an object map I_{ij} in which $I_{ij} = 0$ represents object-absent locations (i, j), and a positive integer $I_{ij} = m$ indexes the location of a point object m and its corresponding 100-dimensional feature vector at the viewer-centered location (i, j). The integer and vector representations of a simulated object code, respectively, the identity (e.g., baseball) and features (e.g., white, red, round) of an object. 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.

Targets of a search task are pre-specified before a simulated search session by the object indices m in the target set Ω (Figure 6). In real search experiments, the knowledge of Ω 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. Note that such conceptual understanding of a target differs from the perceptually driven object categories in model VPFC whereby a target template can be perceptually primed for search guidance (Vickery et al., 2005). The abstract target identities, Ω, may be maintained throughout search in more anterior or medial parts of prefrontal cortex,

whose detailed cognitive operations and neural mechanisms are beyond the scope of ARTSCENE Search.

Also pre-specified before simulations, the prototype of each object *m* serves as a bottom-up filter matching V4/ITp inputs to ITa for object recognition, and also a top-down prime from ITa back to ITp/V4 when model neurons in ITa are primed by the corresponding object representations in VPFC. Since S_{ijk} (Equation 5) and O_m (Equation 17) represent, respectively, the activities of V4/ITp and ITa neurons, the interconnection weights between V4/ITp and ITa, or object prototypes, are denoted by

$$\vec{W}_m^{OS} = \left(W_{m1}^{OS}, W_{m2}^{OS}, W_{m3}^{OS}, \cdots, W_{mk}^{OS}, \cdots\right) = \vec{W}_m^{SO}, \tag{3}$$

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 remaining 90 components are set equal to 0. A target in this set-up possesses some unique features while sharing other features with distractors. To simulate the cases where color is a major attribute in an experimental manipulation (i.e., Figures 17 and 18), 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*.

The number of nonzero components in each object feature vector determines the effectiveness of top-down feature-based attention. In the extreme case where only one dimension of the feature vector is nonzero, the chance of any overlapping features between two objects is very small. As a result, feature-based attention can precisely prime a target object without simultaneously enhancing the saliency of a target-like distractor. In contrast, if 99 out of 100 dimensions of the object feature vector are nonzero, the effect of feature-based attention will be small due to almost non-specific featural priming across objects. Therefore, if 10 or 20 components are set to 1, the effect of feature-based attention will be moderate, as in real-world situations where targets and distractors share some common features. The dimensionalities of color and shape representations were chosen based on this rationale.

Primary and secondary visual areas (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 context 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*:

$$f_{ij}^{(m)} = \left(f_{ij1}^{(m)}, f_{ij2}^{(m)}, f_{ij3}^{(m)}, \cdots, f_{ijk}^{(m)}, \cdots\right) \equiv \gamma_m \vec{W}_{I_{ij}}^{OS} = \gamma_m \vec{W}_m^{OS}, \tag{4}$$

where γ_m controls overall saliency of the object *m*, 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 egocentric coordinate (*i*, *j*) in response to the object *m* at location (*i*, *j*). In all simulations, γ_m was 1 plus a random white noise between 0 and 10^{-8} for all objects in a search display (i.e., a symmetry-broken iso-saliency map) with the following exception: To simulate distant cueing effects (Figure 15), γ_m was lowered to 0.1 plus a random white noise between 0 and 10^{-8} for a

target and half of its distractors that are nearest to that target in a search display. 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.

Fourth visual area and posterior inferotemporal cortex (V4/ITp, S_{ijk})

Model area V4/ITp receives bottom-up inputs $f_{ijk}^{(m)}$ from V1/V2 and top-down primes from anterior inferotemporal cortex (ITa). Specifically, the V4/ITp cell activity S_{ijk} is driven bottom-up by the k^{th} feature in its receptive field centered at the egocentric position (i, j) and is gain-modulated by the activities of ITa neurons O_m (cf. 'pigeonholing' in Bundesen et al., 2005):

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

where the top-down-modulated input obeys

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

In Equation 6, W_{mk}^{OS} is the template of synaptic connection strengths from the m^{th} object primed from ITa to V4/ITp, and Φ_{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:

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

The featural priming in Equation 6 across all locations (p, q) from a position-invariant object representation O_m simplifies the ITa-ITp-V4 feedback pathway whereby a position-invariant object category in ITa primes position-variant object categories in ITp, which in turn primes the corresponding features in V4/ITp within a specific receptive field (Chang, Cao, & Grossberg, 2009). The competition in Equation 5 normalizes the output of each feature map into the range of zero to one, and enhances the contrasts of visual inputs in each feature map.

Where Stream (PPC-PHC-DLPFC):

Posterior parietal cortex (PPC, P_{ij})

The model PPC forms an egocentric/spatiotopic priority map whose activities P_{ij} pool feedforward inputs S_{ijk} from V4/ITp (Equation 5), and are gain-modulated by top-down attentive feedback projections D_{ij} (Equation 15) from DLPFC corresponding to the egocentric location (i, j):

$$\frac{d}{dt}P_{ij} = -.01P_{ij} + (1 - P_{ij})\left[.05\sum_{k} S_{ijk}(1 + 10D_{ij}) + \psi_{0.3}(P_{ij})\right] - P_{ij}\left[\sum_{(x,y)\neq(i,j)}\psi_{0.3}(P_{xy}) + 5\psi_{0.9}(Q_{ij})\right], \tag{8}$$

where the signal function $\psi_p(x)$ determines *when* to switch on a signal x based on the threshold p (e.g., winner-take-all competition begins when $P_{ij} = 0.3$ in Figures 7-9):

$$\psi_p(x) = \begin{cases} 1, & x \geq p \\ 0, & x < p \end{cases}. \tag{9}$$

In Equation 8, the activities P_{ij} are contrast-enhanced by a recurrent shunting on-center off-surround network (Grossberg, 1973) 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 the short-term memory of visited locations Q_{ij}, which obeys:

$$\frac{d}{dt}Q_{ij} = (1 - Q_{ij})\psi_{0.5}(P_{ij}). \tag{10}$$

Computationally, Q_{ij} is switched on when the corresponding PPC location representation, P_{ij}, exceeds $p = 0.5$, the threshold that is chosen in signal function ψ_p (see Equation 9). It then builds up over time to break the positive feedback loop of the maximum P_{ij} in Equation 8, initiating a new cycle of location selection. This inhibition of return by recurrent negative feedback prevents the network from perseverating on the same choice of locations in later selection cycles (Grossberg, 1978; Koch & Ullman, 1985).

Functionally, P_{ij} guides overt attention and gaze location. In particular, before the model triggers a saccade to the next object, fixation is maintained at the location (*I, J*) where the model PPC cell is most active on the attention priority map P_{ij}:

$$(I, J) = \arg\max_{i,j} P_{ij}, \tag{11}$$

and (*I, J*) is the location that is chosen by the recurrent competitive dynamics in Equation 8. In other words, model PPC selects a location in a winner-take-all manner.

Parahippocampal Cortex (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 shunting competitive network with linear feedback signals (Grossberg, 1973, 1980):

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

where τ_H is the characteristic response time of H_{ij}, and λ 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 saliencies in model PPC (namely P_{ij}), which is dynamically updated in the course of *sequential* spatial selections and inhibition of return (Equations 8 and 10). This property of the short-term memory in model parahippocampal cortex, in tandem with the saliency-dependent long-term memory of spatial contexts (Equation 16), allows ARTSCENE Search to explain why attended locations are more effective contexts than non-attended locations (see Figures 13, 15, and 17). To obtain this model property, fast $\tau_H = 0.1$ ensures that H_{ij} rapidly converges to equilibrium for all locations (*i, j*) based on the initial values of P_{ij}, and small $\lambda = 10^{-5}$ makes later P_{ij} inputs more weakly perturb the stored values in H_{ij}. In all, PHC stores a *primacy gradient* of location saliencies that represents spatial scene gist; cf., a small input term generates a primacy gradient in the STORE working memory model of Bradski, Carpenter, & Grossberg (1994).

Dorsolateral prefrontal cortex (DLPFC, D_{ij})

The DLPFC cell activity D_{ij} represents the likelihood of target presence at an egocentric location (*i, j*), which can be predicted from the context of spatial cues stored in PHC as H_{xy} in the egocentric space (*x, y*). The efficacy of H_{xy} for driving D_{ij} is defined as

$$(HA)_{xyij} \equiv H_{xy}A_{xyij}(\sigma_A), \tag{13}$$

where $A_{xyij}(\sigma_A)$ is a volitionally-modifiable attentional window of width σ_A surrounding the location (i, j):

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

which may be controlled by gating signals from the basal ganglia.

In model DLPFC, the cell activities D_{ij} are driven by bottom-up inputs P_{ij} from PPC as well as by inputs $(HA)_{xyij}$ from PHC, and stored by a shunting recurrent competitive network with linear feedback signals (Grossberg, 1973):

$$\frac{d}{dt}D_{ij} = -D_{ij} + (1 - D_{ij})\left(10\psi_{0.5}(P_{ij}) + \sum_{xy}(HA)_{xyij}W_{xyij}^{HD} + D_{ij}\right) - D_{ij}\sum_{(x,y)\neq(i,j)}D_{xy}. \quad (15)$$

In the excitatory term of Equation 15, the winning PPC representation $\psi_{0.5}(P_{ij})$ is the strongest input due to the multiplicative factor 10, and drives the corresponding D_{ij} to be the most active location representation in model DLPFC during an eye fixation. This mechanism ensures a target location, once found, to be the winning representation for spatial context learning in Equation 16. Term $\sum_{xy}(HA)_{xyij}W_{xyij}^{HD}$ is an inner product for matching $(HA)_{xyij}$ with W_{xyij}^{HD}, the long-term memory of a spatial context associated with a target location (i, j). The learned weights W_{xyij}^{HD} between location (x, y) in model parahippocampal cortex and location (i, j) in dorsolateral PFC 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 $(O\Omega)_M$ (Equation 21) in model dorsolateral PFC (Dranias, Grossberg, & Bullock, 2008):

$$\frac{1}{\mu_D}\frac{d}{dt}W_{xyij}^{HD} = (O\Omega)_M\,\psi_{0.8}(D_{ij})\left[(HA)_{xyij} - W_{xyij}^{HD}\right], \quad (16)$$

where μ_D is the learning rate for spatial contexts in the dorsal Where stream. Both σ_A in Equation 13 and μ_D in Equation 16 varied in different simulations (see Table 2). Note that 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 σ_A	Learning Rate μ_D / μ_V
Fig.10: Positive spatial cueing	15	12	8x6	24x30	0.8	10^{-3} / 0
Fig.11: Recombined context	15	11	12x8	36x20 (36x3)	1.2	10^{-3} / 0
Fig.12: Set size effect	15	8/12/16	12x8	24x30	1.2	10^{-3} / 0
Fig.13: Local cueing	15	16	12x8	32x20	1.2	10^{-3} / 0
Fig.14: Negative spatial cueing	15	12	8x6	24x24	0.1	10^{-3} / 0
Fig.15: Global/Distant cueing	15	11	8x6	6x9 (6x1)	5.0	10^{-3} / 0
Fig.16: Object cueing	15	11	8x6	16x24	5.0	0 / 10^{-3}
Fig.17: Attentional learning	15	16	12x8	24x30	2.0	10^{-3} / 10^{-3}
Fig.18: Non-predictive features	15	16	8x12	24x20	2.0	10^{-3} / 10^{-3}
Fig.19: What-Where Integration	15	6	8x6	16x20	0.5	10^{-3} / 10^{-3}

Table 2. Simulation parameters. The three free parameters in the model are the attentional window size σ_A (Equations 13 and 14), the learning rate for spatial contexts μ_D (Equation 16), and the learning rate for object contexts μ_V (Equation 24). Except the last three simulations, either μ_D or μ_V was set to zero to examine how much object or spatial cueing alone can account for the observed contextual cueing effects (see also Figure 8 and 9). When both parameters are non-zero, nonspecific target priming due to irregular spatial or object contexts yields no search facilitation, as if no context learning occurs (i.e., μ_D and/or μ_V is zero).

What Stream (ITa-PRC-VPFC**):**

Anterior inferotemporal cortex (ITa, O_m)

The ITa cell activities O_m are driven by bottom-up object recognition signals o_m from V4/ITp (Equation 18) and top-down object primes V_m from VPFC (Equation 23) in a shunting on-center off-surround competitive network (Grossberg, 1973):

$$\frac{d}{dt}O_m = -O_m + (1 - O_m)(o_m + V_m) - .05O_m\left(\sum_{n \neq m} o_n + \sum_{n \neq m} V_n\right). \tag{17}$$

In Equation 17,

$$o_m = \sum_k W_{mk}^{SO}(PS)_k \tag{18}$$

is computed by matching the foveated features S_{ijk} with the m^{th} position-invariant object prototype W_{mk}^{SO} after PPC drives an eye movement to its focus of spatial attention:

$$(PS)_k \equiv \sum_{ij} \psi_{0.5}(P_{ij}) S_{ijk}. \tag{19}$$

In other words, $\psi_{0.5}(P_{ij})$ selects the V4/ITp input S_{ijk} (Equation 5) from the most salient location in model PPC (Equation 8), and kernel W_{mk}^{SO} is the V4/ITp-to-ITa bottom-up object filter. Thus, the identity of the most active O_m corresponds to the object *M* situated at the currently attended location, where:

$$M = \arg\max_m O_m. \tag{20}$$

When the winning object category, *M*, matches a target, a dopamine burst $(O\Omega)_M$ is triggered to initiate top-down learning from VPFC to PRC, and from DLPFC to PHC (Figure 6 and Equations 16 and 24):

$$(O\Omega)_M = \begin{cases} 1, & M \in \Omega \\ 0, & M \notin \Omega \end{cases}, \tag{21}$$

where Ω 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 $(O\Omega)_M = 1$), 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 spatial inhibition of return (Equations 8 and 10).

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 17, 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 that are not necessarily foveated. In contrast, the explicit multiplication of bottom-up by top-down inputs (cf., Bhatt, Carpenter, & Grossberg, 2007) in Equation 8 avoids spatial selection of no-object locations. Hence, scene context does not strictly confine search to likely target locations when targets are not actually there (Neider & Zelinsky, 2006).

Perirhinal cortex (PRC, R_m)

Model PRC object category neurons receive one-to-one inputs from the object category neurons in model ITa, and store the sequentially activated object representations O_m from ITa in short-term memory via a recurrent shunting competitive network with linear feedback signals:

$$\frac{d}{dt}R_m = -.01R_m + (1 - R_m)\left(\psi_{0.5}(O_m) + R_m\right) - .01R_m \sum_{n \neq m} R_n . \tag{22}$$

In Equation 22, the ITa input $\psi_{0.5}(O_m)$ in the excitatory term has 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. Although the exact form of object short-term memory in model perirhinal cortex is not critical to simulate the behavioral data presented in Section 6, a *recency gradient* of stored visual cues in model perirhinal cortex 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). Correspondingly, because of the form of Equation 22 (cf., Bradski et al., 1994), PRC cell activities R_m exhibit a recency gradient over time whereby recently viewed object cues can be associated with the reward-like target more strongly than earlier ones in ARTSCENE Search.

Ventral prefrontal cortex (VPFC, V_m)

Model ventral PFC activity V_m for the m^{th} object is driven by bottom-up object category inputs O_m from ITa (Equation 17) and object context inputs R_n formed in model perirhinal cortex (Equation 22) that are stored by a recurrent shunting competitive network with linear feedback signals:

$$\frac{d}{dt}V_m = -V_m + (1 - V_m)\left(10\psi_{0.5}(O_m) + \sum_n R_n W_{nm}^{RV} + V_m\right) - V_m \sum_{n \neq m} V_n . \tag{23}$$

In the excitatory term of Equation 23, the winning ITa representation $\psi_{0.5}(O_m)$ is the strongest input due to the multiplicative factor 10, and drives the corresponding VPFC activity V_m to be the most active object category during an eye fixation. This mechanism ensures a target, once found, to be the winning representation for object context learning in Equation 24. Term $\sum_n R_n W_{nm}^{RV}$ matches the stored viewed object context R_n with W_{nm}^{RV}, the long-term memory of the target-biased object context. The learned weights W_{nm}^{RV} between the n^{th} object category in PRC and the m^{th} object category in VPFC obey an instar learning rule that is doubly gated by the dominant VPFC object category $\psi_{0.8}(V_m)$ and target-triggered dopamine bursts $(O\Omega)_M$ (Equation 21):

$$\frac{1}{\mu_V}\frac{d}{dt}W_{nm}^{RV} = (O\Omega)_M \psi_{0.8}(V_m)\left[R_n - W_{nm}^{RV}\right]. \tag{24}$$

In Equation 24, μ_V is the learning rate for object contexts in the ventral What stream and varied in different simulations (see Table 2). Note that the instar learning here simultaneously updates all pairs of stored object-target associations.