Visual spatial relationship perception requires attentional shifts to individual objects

Franconeri, S. L., Scimeca, J. M., Roth, J. C., & Helseth, S. A.

Department of Psychology, Northwestern University

Address correspondence to:
Steven Franconeri
Northwestern University
2029 Sheridan Rd, Evanston, IL 60208
Phone: 847-491-1259
Fax: 847-491-7859
franconeri@northwestern.edu

RUNNING HEAD: Spatial relationship representation
Word Count: 8609

Abstract

Visual processing breaks the world into parts and objects, allowing us not only to examine the pieces individually, but also to perceive the spatial relationships among them. We argue that a flexible system may encode relations across space as a sequence of attentional shifts. Although spatial relationship judgments can give the impression of simultaneous attention to multiple objects, we demonstrate that observers do shift attention between the judged objects using an electrophysiological correlate of the locus of selection. In Experiment 1, participants judged the relative relationship between two shapes, and they reliably shifted attention toward the shape closer to fixation and then toward the farther shape. Experiment 2 demonstrates a similar effect using colors, which are less likely to require shifts of attention simply for identification. Experiment 3 verifies that the colors used do not require shifts of attention for identification by showing that they are efficiently located in a visual search task. Together, these results provide direct evidence that visual spatial relationship representation may require shifts of attention between objects.

[171 words]

Keywords: Attention, Selection, Spatial relationships, Visual routines, Binding

To understand and act on the world, our cognitive system must recognize patterns in the environment. These recognition processes often rely on matching current input to stored representations in long-term memory. We can more easily work with long strings of digits if they are chunked into numbers with existing representations, e.g., "1776 1984 2008" (Miller, 1956). Some models of word recognition specify hardwired detectors for frequent pairings of letters, or for whole words (McClelland & Rumelhart, 1981). Visual processing may take advantage of similar detectors to respond to predefined conjunctions of features, such as red and vertical (e.g. Holcombe & Cavanagh, 2001), or typical combinations of features that might occur within frequently occurring natural objects (van Rullen, 2009). These hardwired representations allow for fast and efficient processing of frequently encountered patterns. However, they have the disadvantage of being inflexible, responding to only particular stimuli.

When hardwired representations are not available for a given pattern, a more flexible system allows for recognition, though often with less efficiency and capacity. Remembering a randomized version of the same list of memorable dates (e.g., "8172 4907 6180") is possible, but much more difficult. Similarly, processing unfamiliar words may slow a reader (Rayner & Duffy, 1986), and recognition of many visual feature conjunctions might require focused processing (Treisman & Gelade, 1980).

Within the vision literature, a substantial amount of research has explored how such flexible representations underlie the recognition of simple and complex objects (e.g., Biederman, 1987). The proposed experiments explore a similar question for complex structures involving multiple objects: How do we recognize spatial relationships between objects? Relational processing for some frequently encountered objects, such as the location and appearance of facial features (Tanaka & Farah, 2006) or the location of features or structures within a scene (Henderson & Hollingworth, 1999; Oliva & Torralba, 2007; Sanocki & Sulman, 2009) might be subserved by

existing long-term representations. But for more novel combinations, a more flexible short-term system is necessary.

What are some possible representations that might support this flexible system? Consider the simple case of a left-right relation between two objects. Perhaps the most difficult computational problem for the visual system to overcome is the need to *bind* the objects with their relative locations (Hummel & Biederman, 1992; Logan & Zbrodoff, 1999; Miller & Johnson-Laird, 1976). That is, we might recognize the individual objects, and know the positions that they occupy, but perceiving the relationship correctly requires that we know which object is in which position, so that we do not confuse "X Y" with "Y X".

This problem might strike the reader as odd – after all, we know where the left object is, and we know where the right object is – so we have all of the information necessary to judge the relation. Critically, this information is only *implicitly* represented. The two locations are known, but the locations alone do not provide an explicit representation of which location is above or to the right of another. That is, you might know that the red object is at horizontal position 4, and the green object at position 6, but the relationship between them is implicit until you explicitly subtract 4 from 6 and note whether the answer is negative or positive.

A hardwired long-term representation solves this problem by having separate representations for every configuration of every pair of objects. However, this mechanism quickly encounters the same problem found by view-based recognition systems in the literature on single object recognition, where an enormous number of existing representations must store views of objects from many different perspectives (Biederman, 1987; Hayward, 2003; Hummel, 2000; Tarr & Bulthoff, 1998). While this problem might be surmounted for single objects, it is compounded for relations between multiple objects, because each template must specify the features of *both* objects as well as the angle and distance between them.

More generally, any hardwired system would have difficulty recognizing relations between novel objects, or relations between objects that are only subtly visually different or even visually identical (see the General Discussion for ways to abstract these hardwired representations into more computationally feasible solutions). At minimum, another more flexible mechanism is needed for such cases.

How do other flexible systems solve this binding problem? The object recognition literature faces a similar dilemma - if an object recognition model represents an object as a set of parts, then how are features from one part assembly represented as separate from others? One solution is to separate part structures with time, by oscillating which features are represented as more active at any moment and thus co-activating features of the same part structure simultaneously (Hummel & Biederman, 1992; Singer, 1996). A similar problem also arises in the visual search literature, with a similar solution. When searching for a horizontal green object among horizontal red and vertical green objects, finding the target requires binding of each object's color and shape. One solution is to serially shift through the set of objects, inspecting only one object at a time (Treisman, 1996; Treisman & Gelade, 1980). Though slow, this solution eliminates the confusion of which features came from which object by simply excluding information from all but one object at a time.

---------- Insert Figure 1 About Here ----------

When judging the relationship between two objects (say, a computer mouse and a paper clip), the challenge of binding each object with its location could be solved by selecting one object at a time. The present experiments test whether such dynamic selection over time also allows binding in spatial relationship judgments. This proposal provides a mechanism to explain the role of 'attention' in many past studies of spatial relationship judgments. Figure 1 depicts an example of an attentionally demanding spatial relationship search representative of those used in previous

studies. When observers are asked to find a pair of objects in a given spatial relationship within a search display, adding more distractor pairs severely impairs response time (Logan, 1994, 1995; Wolfe, 2001). Spatial relationship searches may even hold a unique position in their robustness as a difficult serial search (Wolfe, 2001).

The difficulty of these searches is not due to the need to identify the objects within the relation, but instead to processing the relation itself. When the search task is slightly altered so that observers seek a pair of objects with different identities compared to the other objects, the task becomes far more efficient (Logan, 1994). The difficulty of the search task is not tempered by practice (Logan, 1994), or by using pictures of the target pairs instead of instructional descriptions (Logan, 1994), which often improves search performance (Vickery, King, & Jiang, 2005). Instead, the difficulty added by additional objects seems to be in locating which pair of objects to process. When attention is cued to the target pair in the search task by giving the objects in that pair a unique color, response times become much more efficient (Logan, 1994). This need to first select the target pair before computing its relation can even be seen in a simpler display. When asked to quickly judge a relation between two objects, observers are significantly slowed by the presence of just one additional object (Carlson & Logan, 2001). Other demonstrations use change detection tasks to show that processing of relative spatial relationships is slow and capacity-limited compared to processing information about individual objects (Rosielle, Crabb, & Cooper, 2002; Tatler, 2002). Because change detection performance is heavily limited by attention (Rensink, O'Reagan, & Clark, 1997; Scholl, 2000), it is likely that spatial relationship judgment performance is limited by attention.

Saying that a task requires attention means that as the task is applied concurrently to more instances (e.g. pairs of objects), some limited processing resource is increasingly taxed, and speed or accuracy suffers (Pashler, 1998). But critically, although these search and change detection results point out that performance is limited by some attentional resource, they cannot specify the nature of the resource. The resource could be a mechanism that processes multiple relations at

once but with degrading performance as it is stretched more thinly (see Logan, 1995 for a discussion of parallel vs. serial mechanisms), a serial mechanism that sequentially applies a hardwired representation to each pair of objects, or a serial mechanism that requires a sequential process within the two objects (Logan & Sadler, 1996). Our results will suggest that the 'attentional' factor that limits performance is the need to shift the *locus of spatial selection* from one object in a relation to the other. As this can only happen for one pair of objects at a time, requiring an observer to process more than one instance will lead to slower responses or lower accuracy.

Shifts of attention could provide explicit relation information

Even if attentional shifts can provide a way to bind object features with their respective locations, this process still cannot explain how spatial relationships could be explicitly recovered from the implicit difference between the two locations. How does the visual system know that the first object with horizontal location X1 is to the left of the second object with horizontal location X2? We present a potential mechanism as an existence proof that explicit relation information could be recovered from the pattern of shifts, with a trivial amount of additional computational overhead.

The solution would be to require a novel type of 'feature' not present in other binding models, the *direction of the attentional shift* from one object to the other. This shift could be recorded and briefly held in heightened activation (see Figure 2), either by a circuit similar to a detector for low-level motion (Reichardt, 1969) or as an efference copy of the shift direction from the shift command itself. In one simple example (recognizing a paper clip to the right of your computer mouse), the locus of selection could be moved to the center point between the two objects. Selection could then shift toward the right object, leading to a heightened activation of that object's representation. The heightened activation of this object on the right (the paper clip) over the object on the left (the mouse), combined with the high activation of the representation of a

rightward shift, provides all of the information necessary to conclude that the paper clip is to the right of the mouse. The starting point for the shift might not be a place between the objects, but instead on one of the objects. The shift might even have to occur multiple times, perhaps from one object to the other and back again, to gain redundancy in the coding of the relation. Note that this is an unusual role for selection, which often is thought to relatively amplify relevant information at the expense of irrelevant information (Hillyard, Vogel, & Luck, 1998). Instead, both elements of the relation are highly relevant, giving attention a more active role in constructing a representation over time, similar to a visual 'routine' (Cavanagh, 2004; Logan & Zbrodoff, 1999; Ullman, 1984).

---------- Insert Figure 2 About Here ----------

Detecting shifts of spatial attention with an electrophysiological correlate

The present studies seek evidence that spatial relationship judgments entail shifting the locus of selection from object to object. There has been long interest in determining the trajectory and speed of the attentional spotlight (Eriksen & Schultz, 1977; Pinker, 1980; Yantis, 1988). Tracking attentional shifts has been made easier by the recent discovery of an electrophysiological correlate. A large body of work in the last 15 years demonstrates that a shift of attention to one side of the visual field is accompanied by greater negativity in the electrode sites on the contralateral side. This *n2pc* component, first demonstrated as negativity at 200-300ms (N2) (though sometimes as early as 175ms), is located at posterior areas of the brain (P), contralateral to the attended field (C). This posterior negativity appears when a target item must be isolated from distractor items (Luck & Hillyard, 1994), especially when the distractor items are closer to the target (Luck, Girelli, McDermott, & Ford, 1997), or when the search is more difficult (Luck & Ford, 1998). The n2pc signal is not present when the distractors are removed, releasing the requirement to attentionally filter (Luck & Hillyard, 1994). There is debate over the

degree to which the n2pc reflects distractor suppression versus target enhancement (Eimer, 1996; Hickey, McDonald, & Theeuwes, 2006) or even a combination of the two (Hickey, Di Lollo, & McDonald, in press). The signal is likely to originate in lateral extrastriate and inferotemporal cortex (Hopf, et. al., 2000).

The n2pc allows an experimenter to track the relative allocation of spatial attention between visual hemifields at a high temporal resolution. One set of studies exploited the association between shifts of attention and contralateral negativity to determine whether search could serially proceed on an item-to-item basis (Woodman & Luck, 1999; 2003). By motivating participants to search through arrays in a predictable order, through manipulations of item saliency and target probability, and arranging search displays with critical objects in the left or right visual hemifields, the authors were able to show the timecourse of item to item shifts in a serial search in the ERP waveform. For example, at approximately 250 and 350ms, there were shifts to the locations of the 1st and 2nd most likely (or otherwise most attractive) target locations, as demonstrated by increased negativity at posterior electrode sites contralateral to that item's side of the search display.

Experiments

In Experiments 1 and 2, we use n2pc to track shifts of attention during spatial relationship judgments. Note that in these experiments we always measure the n2pc for these shifts with an object-relative analysis. That is, we always ask whether the shifts were toward or away from a given object or instructional designation. We never examined the strategy of simply shifting attention to the left or right, regardless of a trial's condition. While this type of analysis might be possible in a behavioral or eyetracking paradigm, the n2pc technique cannot detect these types of shifts, because a comparison of activity at the left or right hemisphere electrodes would be confounded with any other lateralized activity across the cerebral hemispheres. In contrast, object-relative analyses average across such lateralized differences by presenting object types

equally often on either side of the display, and collapsing results across electrodes contralateral and ipsilateral to a given object type.

Experiment 1

Experiment 1 tests whether participants make sequential shifts of attention during spatial relationship judgments. Some studies show that simple identification or localization of an object, a prerequisite for spatial relationship perception, can cause a shift of attention (Hyun, Woodman, & Luck, 2009; Luck & Ford, 1998). However, in those experiments only one object was relevant for the task, while in spatial relationship judgments both objects are relevant for the task. In particular, during a spatial relationship judgment the reader may feel that, intuitively, attention envelops both objects. The present experiments will show that this impression would be an illusion, by showing systematic shifts between the two objects.

Because the n2pc technique requires averaging results across many trials, if participants do not adopt a consistent strategy the average may reflect a mixture of shift patterns. As an extreme example, if a participant first shifted toward the computer mouse on odd trials, and the paper clip on even trials, the average across trials would show no evidence for shifts. We therefore use a manipulation that biases the shift direction toward one object.

One natural shifting strategy is to start with the object that happens to be closer to fixation, and then shift toward the more distant object. Distance from fixation has been shown to reliably affect attentional priority in a visual search experiment using posterior contralateral negativity signals to track shifts of attention (Woodman & Luck, 2003). However, placing one object closer to or farther from fixation might cause a stronger signal at posterior contralateral areas of the scalp, regardless of shifts of attention. To distinguish shifts of attention from such stimulus-based effects on the ERP, we follow a solution used by Woodman & Luck (2003). By including *two* sets of objects of different colors (see Figure 3), each set with one near and one far object, one set can be task-relevant and the other irrelevant. The analysis can then be collapsed across the two

color sets, resulting in electrodes contralateral to either the task-relevant near or far objects. The retinal stimulation is identical across these conditions - only the task requirements change. Note also that to equate visibility, the farther object is slightly larger, scaled according to the cortical magnification factor (see Woodman & Luck, 2003). We predict that during the spatial relationship judgment, the locus of spatial attention will shift first to the near object, and then to the far object.

---------- Insert Figure 3 About Here ----------

Methods

Participants

15 Northwestern University undergraduates participated in a 2-hour session in exchange for payment or course credit.

Stimuli:

The experiment was controlled by a Dell Precision M65 laptop computer running SR-Research Experiment Builder. Although head position was not restrained, the display subtended 32.6° x 24.4° at an approximate viewing distance of 56cm, with a 1024x768 pixel resolution. All measurements are reported hereafter in pixels; the conversion factor for the present experimental setup was 33.6 pixels per degree.

In the stimulus display, a fixation point was flanked by two red or green shapes on each side. Each shape was either an “x” or a “+” circumscribed by a circular border. The far shapes were 120 pixels from the fixation point, 38 pixels in diameter, and had 5 pixel thick segments, and the near shapes were 40 pixels from the fixation point, 20 pixels in diameter, and had 2 pixel thick segments. Within each color pair, one shape was always an x and the other was a + (see Figure 3). Similarly, one near shape was always green and the other was red, arranged such that the

colors of the shapes alternated from left to right.

One of the far shapes was always green (24 cd/m^2) and the other was always red (14 cd/m^2). The color values were intended to be perceptually equiluminant, as determined by a separate experiment where 8 observers were asked to minimize perceived flicker as a red and green square alternated at 15Hz. Participants performed 20 adjustments of the luminance of a red patch (alternately starting at low or high values) while the luminance of the green patch remained fixed at 24 cd/m^2. Equiluminant values of red were designated as the average of each subject's median value.

Procedure:

At the beginning of each trial a fixation point was displayed for 1800-2200ms, followed by the stimulus display for 1500ms. Each participant was tested on a total of 512 trials in 16 blocks of 32 trials. Trials were randomized within blocks, and each block included an equal number of each of the 8 possible display types (2 red shape orderings x 2 green shape orderings x 2 color orderings). At the beginning of each block, participants were instructed to report the pattern of either the red or green shapes using the M (for + x) or K (for x +) keys on a keyboard. Participants received feedback for incorrect responses and were given brief breaks in between blocks.

EEG Recording

ERP was recorded using a Biosemi Active 2 EEG/ERP system. All sites were re-referenced to the post-recording average of the left and right mastoids and low-pass filtered at 80Hz. We recorded from the following sites according to the 64-channel modification of the international 10/20 system: F3/4, C3/4, PO3/4, P5/6, P7/8, PO7/8, O1/2, POz, Oz, Horizontal and Vertical EOG. The HEOG and VEOG channels were used to reject eye movement artifacts and blinks, using a combination of automated rejection thresholds and hand-inspection. Both types of EOG

rejection used thresholds for both absolute and slope changes, defined individually for each subject, for 200ms before to 800ms after stimulus presentation. Participation in the experiment took 1.5 hours, including ERP cap preparation, breaks, and task practice. Inter-trial delays include randomized timing with at least 400ms of jitter (rectangular distribution) to minimize the impact of previous trials on the EEG signal.

Results & Discussion

Of the 15 total participants, the results from 3 were not analyzed due to an inability to maintain fixation. Two participants were removed from the analysis for excessive HEOG, and one was removed due to excessive artifact rejection overall (58%). For the remaining 9 observers, an average of 20.8% of trials were rejected due to eye movement artifacts, blink artifacts, or electrode noise (Min=6%, Max=33%). Every participant showed 2uv or less of a difference between HEOG signals for near-shape left and near-shape right trials, confirming that participants did not systematically move their eyes toward either the near or far shapes. Trials with incorrect responses or responses of over 1500ms were also removed from the analysis. Accuracy was high (M=96.6%, SD=3.4%). Response time was 741ms on average (SD=82ms)

---------- Insert Figure 4 About Here ----------

Woodman & Luck (2003) used a similar manipulation to induce shifts of attention toward objects in a visual search in a known order. Based on their results, we predicted a priori that activity would be more negative contralateral to the near shape between 200-300ms post-stimulus, and more negative contralateral to the far shape after 300-400ms post-stimulus.

The results confirm the prediction. Figure 4a depicts waveforms for electrodes contralateral to the near and far shapes, and Figure 4b depicts the difference between these two expressed as signals consistent with attentional shifts toward either shape. At earlier times 200-300ms post-

stimulus, PO7/8 amplitudes were more negative contralateral to the near target compared to the far target (Difference M=0.78uv), t(8)=4.2, p=0.003. At later times 300-400ms post-stimulus, the reverse trend appeared where amplitudes were more negative contralateral to the far target (Difference M=0.82uv), t(8)=4.4, p=0.002.

Participants in Experiment 1 shifted attention between two shapes during a spatial relationship judgment, suggesting that such shifts may typically accompany relational judgments. However, it is possible that these shifts may not have been required for the relationship judgment itself. Instead, it is possible that serially inspecting each shape was necessary simply to reliably identify them. If so, this effect still presents a challenge to any model of spatial relationship processing that relies on templates or abstracted templates for relational judgments, by providing an existence proof of a relational judgment where the objects were serially inspected. If participants had been able to use, e.g., an abstracted template for the relationship between diagonal and non-diagonal lines, these shifts should not be necessary. More broadly, real-world judgments should require the same types of shifts for object identification. Therefore, any potential underlying mechanism for spatial relationship processing must to be able to reconstruct relations that are initially represented as a sequence over time. This critique also highlights an important advantage of representing spatial relationships as a sequence of attentional shifts. If we already constantly shift attention among objects of primary interest, then the shifts themselves can serve as a carrier for spatial information about those objects, with a minimum of additional computational overhead.

Yet, it is also important to demonstrate that this serial selection of objects occurs even when those shifts are not required for identification of the objects. In Experiment 2, we replace the shape relation task with a color relation task, which should not require serial selection of objects to identify the colors (Treisman & Gelade, 1980).

Experiment 2

In Experiment 2 participants judged the color relationship between two objects. As in Experiment 1, displays were balanced such that there were always two objects on each side of fixation. Within the relevant subset of 2 objects (either diamonds or circles) one object was on each side of fixation, and one was closer to fixation. If judging the spatial relationship between the two colors requires serially selecting one or more of the objects, the shift should be toward the near object at an early time window, and toward the far object at a later time window.

Experiment 2 added two additional elements to further demonstrate that any attentional shifts would be due to the relational judgment. First, one might argue that attentional shifts seen in Experiment 1 could be related to the need to translate a purely visual representation to another representation needed to map the relation to a response. For example it could be necessary to match the stimulus to a verbally encoded response mapping, e.g. "Press X if magenta is to the right of green". We therefore eliminated the need to memorize response mappings, by providing a response screen 800ms post-stimulus depicting the two possible arrangements within the relevant objects. Participants pressed one gamepad button for the upper arrangement, and another for the lower arrangement (the buttons were congruently vertically arranged on the gamepad). Note that adding this manipulation removed the ability to measure response times, because the participants were required to wait 800ms for the response selection screen.

Second, one might also argue that in the design used in Experiment 1, shifts of attention were not necessary, but they were also not discouraged. Past work shows that such 'exploratory' shifts can indeed happen. When asked to detect simple features in a search display, a study using a similar n2pc design revealed shifts of attention to unique objects (Luck & Ford, 1998), even when other work suggested that such shifts should not be necessary. However, when participants were given a noisy letter discrimination as a dual-task, the shifts disappeared, suggesting that the shifts could be discouraged. Experiment 2 employs two conditions: a single task where participants judge only the spatial relationship between the relevant colors, and a dual task where they must

additionally perform a difficult letter discrimination similar to the one used by Luck & Ford (1998). We predict that in the dual-task condition, the shift of attention related to the spatial relationship judgment will occur later in time. Pilot results using different secondary tasks suggest that this shift, which usually occurs initially at 200-300ms, will occur even later post-stimulus (approximately 400ms).

Methods

Participants

12 Northwestern University undergraduates participated in a 2-hour session in exchange for payment or course credit.

Stimuli:

Stimuli were similar to those used in Experiment 1, with the following exceptions (see Figure 5). Fixation displays had a grey background (13 cd/m^2) with a small light grey fixation cross (30 cd/m^2) 12 pixels wide. The stimulus displays added four objects (in the same locations as Experiment 1) and a noise-degraded letter. The objects were colored shapes, either diamonds (24 pixels wide) or circles (20 pixels wide), in green (36.6 cd/m^2), cyan (32.6 cd/m^2), magenta (28.0 cd/m^2), or orange (36.4 cd/m^2). These colors were intended to be perceptually equiluminant, as measured for an independent group of 6 observers using a procedure similar to the one used in Experiment 1. There was also always a white square 50 pixels wide and centered 50 pixels below fixation, containing a black letter (A, E, O, U; H, S, X, N) in Helvetica font approximately 28 pixels high. The white box was noise degraded by the replacement of 65% of its pixels with randomly chosen white or black values.

The choice and positions of the colored shapes were constrained in several ways. There were always two diamonds and two circles interleaved, such that both shapes of one type never appeared on the same side of fixation, and one example of each shape was always closer to

fixation. Each shape type always contained one of two color pairings, either green and magenta, or blue and orange. Displays were fully counterbalanced such that each shape and color appeared at each of the four screen locations equally often.

---------- Insert Figure 5 About Here ----------

Procedure:

In each block, the participant was asked to judge the relationship among the colors for either the diamonds or the circles, ignoring the irrelevant shape set. Participants were also told which of the two color sets would be relevant, cyan/orange or green/magenta. At the start of each block, an instruction screen first appeared for 1500ms, depicting examples of the currently relevant objects in both possible arrangements (e.g., a cyan circle to the left of an orange circle, and the opposite pairing underneath). The instruction screen also specified whether the center letter was relevant, by asking the participant to either ignore it the letter, or press one key if it were a vowel and another if it were a consonant. Participants were also told to prioritize the spatial relationship judgment, and to report the letter only if possible. The instruction screen also reminded the participant that they should complete the relation task by attending to *both* circles (or diamonds) simultaneously, and try their best not to use a strategy of basing their response on only one object, even if it impaired their performance. The experimenter also emphasized this point repeatedly in verbal instructions.

Trials began with a fixation screen lasting 800-1200ms (rectangular distribution), and were followed by the stimulus display for 120ms, another fixation display for 680ms, and the answer display (containing the two possible arrangements of the relevant objects) until response. Eye movements were monitored by a table-mounted SR-Research Eyelink 1000 Remote eyetracker. If participants moved their eyes outside of a 30 pixel radius around the fixation point, from the time window starting from 800-1200ms preceding stimulus presentation (depending on the

randomly chosen inter-trial jitter value) to 800ms after stimulus presentation, the trial was rejected. Given the small amount of noise present in the eyetracker's position signal (approximately 15 pixels), the effective size of the allowed window was actually smaller than the permitted 30 pixels radius. On rejection, the participant was presented with a screen depicting the allowed fixation region and a dot showing real-time eye position. There was also an indicator of whether the participant had looked left, looked right, or blinked. The experimenter could then choose to recalibrate the eyetracker at his or her discretion. The trial was then repeated at a randomly chosen point within the block.

Participants repeated a 320-trial sequence twice. Of the 320 total trials, half required only the single spatial relationship task, and half required the dual task adding the noisy letter identification. For the 160 trials in each condition, there were 40 trials for each of the four combinations of relevant shape set (diamond or circle) and associated color pairs for that set (green/magenta, blue/orange). Within each of these 40 trials blocks, there were 5 trials of each of the 8 combinations of position of the relevant shape set (e.g., diamonds shifted to the left), the color relationship within the relevant shapes, and the color relationship within the irrelevant shapes. The order of these 40 trial blocks was randomized, but single/dual task was blocked such that one task was entirely completed before the other, in random order. Self timed breaks were given after each of these 40 trial blocks, followed by the instruction screen depicting the relevant shape and color sets for the next block.

Results & Discussion

Accuracy in the spatial relationship judgment task was high (M=97%) in the single task condition, and only slightly lower (M=93%) in the dual task condition. In the dual task condition, accuracy for the letter identification task, which participants understood had a lower priority than the relational task, was lower (M=62%). Two subjects were removed from the analysis due to an excessive number of trials rejected due to eye movements.

Figure 6a depicts activity at PO7/8 for electrodes contralateral to the near and far targets in the single task condition. We predicted a similar pattern in Experiment 1, where activity would be more negative to the near object between 200-300ms, and more negative for the far object between 300-400ms, suggesting a shift of attention from the near object to the far object over time. The earlier trend emerged (Difference M=0.33uv), $t(9)=3.0$, $p=0.014$, but there was no difference at the 300-400ms time window. Although not predicted a priori, there was a trend from 500-600ms for a return to more negativity for electrodes contralateral to the near object (Difference M=0.35uv), $t(9)=2.0$, $p=0.08$. This pattern is consistent with a shift of attention toward the relevant object closer to fixation, and perhaps later a second confirmatory shift back to that object. Figure 6b depicts this pattern as a difference wave, expressed as more negativity for electrodes contralateral to the near vs. far object.

---------- Insert Figure 6 About Here ----------

Figure 7a depicts activity at PO7/8 for electrodes contralateral to the near and far targets in the dual task condition. We predicted a similar pattern in Experiment 1, where activity would be first be more negative to the near object. Pilot experiments using a different dual-task manipulation suggested that this shift would occur later in time (approximately 400ms), and would not occur for both the near and far objects. The present results were similar to the pilot results, except that instead of more negativity toward the near object at a late time window, there was more negativity in electrodes contralateral to the far object between 400-500ms (Difference M=0.5uv), $t(9)=2.6$, $p=.026$. Although not predicted a priori, this negativity continued into the 500-600ms time window (M=0.31uv), $t(9)=2.3$, $p=.047$. This pattern is consistent with a shift of attention toward the relevant object *farther* from fixation at a later time window, perhaps after discrimination of the letter at fixation.

---------- Insert Figure 7 About Here ----------

These results demonstrate that participants shifted attention in systematic ways during spatial relationship judgments, even under difficult dual task conditions where such shifts should be discouraged. The shifts should not be necessary to discriminate the identity of the object colors (see Experiment 3 for additional evidence that the color identities were available without selection of each object). In particular, when using a similar noisy-letter judgment task, a past study using a similar n2pc design showed no shifts of attention toward an object when participants were asked to simply identify it (Luck & Ford, 1998). In contrast, in the present study this same dual-task manipulation did not prevent participants from systematically shifting attention to one of the colored objects.

The results also present two differences compared to Experiment 1. First, each condition only shows one shift toward one object, instead of shifts toward both the near and far objects. Note that this pattern of results is equally consistent with our claim that such shifts are required for spatial relationship processing, and that shifting to both objects should not be necessary. To know that magenta is to the left of green, it is sufficient to know that magenta is on the left. It is likely that participants in Experiment 2 were more inclined to inspect only one object because the display was only presented for 120ms, instead of being presented until after the response in Experiment 1. Second, at the late time window in the dual-task experiment, participants shifted toward the farther object instead of the closer object. One possibility is that while inspecting the degraded letter, the near objects were enveloped in a penumbra of inhibition that would accompany the 'spotlight' of selection focused on the degraded letter (Bahcall & Kowler, 1999; Hopf et. al., 2006), making the far object a more attractive target. A more speculative possibility is that, if the direction of the relation were carried by the direction of the vector produced by the shift of attention, then the vector from the degraded letter to the near object would be a diagonal, a poor exemplar of a horizontal relationship (Logan & Sadler, 1996). In contrast, the vector from

the letter to the far object would have a much stronger horizontal component and would carry a more effective horizontal directional signal.

Experiment 3

Based on past work using visual search paradigms (e.g., Treisman & Gelade, 1980), it is likely that the colors used in Experiment 2 did not require selection in order to be identified. However, because the set of colors was more heterogeneous than in past experiments, we conducted a visual search task to ensure that a singleton color could be efficiently localized in a visual search display. If adding additional distractors to the display does not substantially increase response times for target localization, then the color identification should not require focused attention.

Methods

Participants

8 Northwestern University undergraduates participated in a 30 minute session.

Stimuli:

Stimuli were similar to those used in Experiment 2, except that up to 3, 6, or 12 colored shapes were distributed across the display (see Figure 8, dotted lines for illustration only). To maintain inter-object density across these set sizes, triplets of objects were constrained to quadrants of the search display. In three object displays a random quadrant was chosen, in six object displays the two quadrants were always within the same hemifield, and twelve object displays used all four quadrants. The target was randomly chosen from the four possible colors, and distractors were chosen from the remaining colors without replacement for each quadrant. All objects were randomly either circles or diamonds, with the constraint that at least one of each

shape be present in each quadrant, and that the ratio between shapes be the same across quadrants (to maintain homogeneity of shape ratios across set sizes). The dominant shape was randomized and counterbalanced within subject. The fixation point was a small ring, and the shape sizes, colors, and eccentricities from fixation were identical to Experiment 2. Inner shapes were placed 45 degrees off the display's vertical or horizontal axes (22.5 degrees for outer objects).

---------- Insert Figure 8 About Here ----------

Procedure:

There were 288 trials, divided by target color into 72 trial blocks presented in a random order for each participant. Each sequence began with an instruction screen depicting the target color, followed by 24 trials of each set size, in random order. For the two smaller set sizes, the quadrant(s) where shapes would appear was blocked so that their locations would be as predictable as in the full displays. In each trial, there was a 1000ms fixation display, followed by the search display until response. Participants used a keypress response to report the shape of the single object in the target color. Incorrect responses were followed by an 'incorrect' message and a 5-second delay.

Results & Discussion

Accuracy was 96% at each set size. Response times were 606, 629, and 648ms for the 3, 6, and 12 shape displays, showing a positive ($t(7)=2.4$, $p=.034$) but very small slope (4.6 ms/item). Slopes were 10 ms/item or less for every participant.

Even though the displays in Experiment 3 were more populated and more crowded than those used in the spatial relationship judgment task in Experiment 2, there was virtually no cost in identifying a target color, suggesting that identifying the colors used in Experiment 2 does not require that they be serially selected.

General Discussion

When we judge visual spatial relationships among objects, we may feel as though we selectively attend to both objects in the relation simultaneously. We argue here that in many cases, simultaneous selection of both objects would present a binding problem, leading to ambiguity about which features were associated with which object. Instead, we argue spatial relationships may be judged by sequentially selecting each object in a sequence over time. We additionally speculate that the direction of the relation could be coded by the direction of the shift. Experiment 1 showed that shifts of attention occurred during spatial relationship judgments, with participants shifting first toward the object closest to fixation, and then to the more distant object. Experiment 2 showed similar shifts when judging simple colors that should not require attentional isolation in order to identify them. Experiment 3 confirmed that these colors could be identified highly efficiently in a visual search task.

This evidence is consistent with past work showing a role for attention in spatial relationship judgments (Logan, 1994; 1995; Logan & Sadler, 1996; Wolfe 2001; see Carlson & Logan, 2005 for review), while specifying attention's role more concretely. Specifically, the attentional resource needed for relational judgments may be the locus of selection itself, which must shift between the two objects being judged. We speculate that recording the direction of these shifts, either passively through 'motion' detectors or actively through efference copy of the shift command, could provide the direction of the relation. This is a computationally feasible way to flexibly represent many aspects of visual structure. The shift directions require simple circuitry to detect and store, and this signal would automatically produce relational information as a by-product of attentional exploration of a scene.

This attention shift mechanism presents alternative, or simply more specific, ways to implement some stages of Logan & Sadler's (1996) model of visual spatial relation judgments. For example, in the 'spatial indexing' stage, the objects in a relation are found and isolated from others in the display. The two objects are then fitted to a 'template' for a given relation, where

one object is specified as the reference and the other as a target, and their spatial arrangement is evaluated for how well it matches the typical examples of that relation. For example, objects 'above' other objects should ideally be directly above, without large amounts of horizontal displacement. Another stage binds the objects to their correct roles in the relation. The present account is similar, but specifies these steps at a lower level of implementation. However, some aspects might be qualitatively different. For example, evaluating how well a set of objects matches a template for a given relation might not be a separate stage. Instead, the goodness of the relation might be determined by how well the direction of the shift (the vector itself) matches the prototypical shift vector for that relation.

The present account is similar to the Attention Vector Sum (AVS) model of Regier & Carlson (2001). The model describes an algorithm for predicting evaluations of how well two objects fit a prototypical relation. The relation is similarly described as a vector, created by the sum of vectors from multiple points within the reference object to the target object. Each vector's contribution is weighted by the proximity of its starting point to a point on the reference object close to the target. The present account would alter this one only slightly, such that instead of summing vectors, only one vector is created and evaluated. The starting point of this vector is created through a process isomorphic to the one used to create the final vector in the AVS model. That is, the same processes described by the AVS model, which take into account the shape of the reference object and its arrangement relative to the target, could produce a single starting point on the reference object for a shift of spatial attention toward the target object. This account would then produce the same predictions and results as specified by Regier & Carlson (2001).

Are all spatial relationships judged by shifting attention between objects?

Almost certainly not. The introduction presented a restrictive version of hardwired LTM mechanisms that appeared incapable of supporting spatial relationship judgments without encountering a combinatorial explosion of existing representations. There are versions of a

hardwired model that could overcome this combinatorial explosion. More generalized templates might 'cheat', and respond not to complex object identity, but rather to more abstract properties like relative differences in brightness or size. Knowing that the small object is to the right of the large object is enough information to conclude that your bicycle is to the right of your garage. Hardwired representations might also be restricted to frequently encountered pairings. The spatial regions that are acceptable might also be diffusely specified (Logan & Sadler, 1996), allowing fewer templates with more categorical outputs.

As an example of such relations at a large spatial scale, one model of scene recognition relies on memory for the spatial organization of features in a scene to aid in scene categorization and priming of subsequent recognition processes (Oliva & Torralba, 2001; Oliva & Torralba, 2007). At a smaller scale, some frequently encountered spatial relationships may be processed by mechanisms that detect local clusters (or "scenelets") of commonly paired objects in a scene (Hayworth, Lescroart, & Biederman, 2007). As evidence of this possibility, there is at least one instance of a visual search for spatial relationships that does not lead to inefficient search rates. When observers were asked to find a cube with dark shading on the top among cubes with light shading on the top, the target object was easy to find (Enns & Rensink, 1990). The fact that top-shaded cubes were easy to find suggests that this relation does not require attention to process, and this relation may instead be processed using long-term memory representations of local spatially organized features. Although we do not have long-term experience with top-shaded objects (light sources usually illuminate the tops of objects), other search results suggest that long-term experience with top-illuminated distractors may have allowed participants to reject them efficiently (Wang, Cavanagh, & Green, 1994).

There are also two emerging views on spatial relationship processing that present exciting new potential mechanisms for encoding between-object structure. One view suggests that at late stages in the visual processing hierarchy (e.g., LOC), cells that represent relatively complex shapes can also be biased toward a preferred position *relative to the current attentional window*

(Biederman, Lescroart, & Hayworth, 2007; Hayworth, Lescroart, & Biederman, 2008). Thus, when attending simultaneously to, e.g., a computer mouse and a paper clip, one subject of cells would be more active when the paper clip were on the left, and one would be more active when the paper clip were on the right. This account is supported by evidence that when presenting a pair of objects twice over time, LOC shows a greater release from adaptation when the two objects flip their relative positions, relative to when they translate the same distance but while maintaining their relation, which is consistent with the possibility that a new set of cells represents the group when the relation is changed. It is not clear that this account can explain the present results, because its predictions are similar to a template model, in that participants judge a relation by spreading the window of attention over both objects at once, instead of each object in sequence. But it is also possible that both mechanisms exist, each specialized to process different kinds of relations.

Under a second view, the ventral visual stream is divided into two effectively separate visual systems, each capable of representing one object in a relation. These two systems would each have an independent 'spotlight' of attention, and the relation between these spotlights could provide the relation between the represented objects (Hayworth, Lescroart, Kim, & Biederman, 2009). It will be an intriguing future direction to explore how this account might explain the present results, but at present it is not clear how the activity of multiple distinct spotlights would be reflected by the n2pc component used here. If correct, this account would more generally represent a transformational change in our understanding of the visual system.

Outstanding questions

Future work should test coordinate spatial relationship judgments, instead of the categorical judgments used here. Coordinate spatial relations capture precise distances between objects in metric units, while categorical spatial relations abstract over precise position to produce generalized relationships such as 'above/below', or 'left/right of'. Behavioral, neuroimaging, and

neuropsychological evidence suggest that these are dissociable processes (Chabris & Kosslyn, 1998; Jager & Postma, 2003; Kosslyn, 1987). The present experiments are limited to categorical judgments because it is more counter-intuitive to demonstrate shifts of attention for relations that would be more likely to be simultaneously processed. In contrast, coordinate judgments are defined even more explicitly as a vector with direction, and may be more likely to be subserved by shifts of attention.

The spatial shifting account can easily explain our ability to process left-right and top-bottom relations, but others are more difficult. For example, how would this account deal with front-back relations? Searching for such relations also leads to inefficient search (Moore, Elsinger, & Lleras, 2001), perhaps because they require observers to switch attention back and forth between each pair of background and foreground objects (Logan, 1994; 1995). Some studies suggest that selection is not possible for a given depth (Ghirardelli & Folk, 1996; Theeuwes, Atchley, & Kramer, 1998), while others suggest that it is possible as long as observers have a continuously available object to select (Atchley & Kramer, 2001; Marrara & Moore, 2000), or a visual surface to select (He & Nakayama, 1995). Selection in depth might also be possible only via mechanisms that select specific objects or object features (Scholl, 2001). Thus, the mechanisms supporting selection in depth are not yet understood well enough to specify how a detector for such shifts might work. As another example, inside-outside relationships *could* be supported by this shifting mechanism. There is a large body of existing work on shifting the locus of selection from local to global scales (see Kimchi, 1992). We would only need to add a detector circuit that fired whenever the scale switched from local to global, or vice-versa. That is, if you would like to judge whether the basket were in the cup, or the cup in the basket, you would know that the latter were true if you shifted to the local scale and were now attending to the cup.

Another issue for future research is how shifts of attention relate to spatial reference frames. Relational judgments can be made relative to some reference frame, such as gravity, the retina, the head, the body, or other external objects (Carlson, 2000). For the present work, we

assume that this reference frame is retinal, but we cannot distinguish this frame from any of the others with the present design. If the direction of the attentional shifts were coded in a retinal reference frame, there would need to be a mapping between the coordinate space of this frame and the frame needed for a given task.

The sequential attention shift account makes several predictions. When using this flexible system, observers should only be able to make one spatial relationship judgment at a time, and this amount of time should be constrained by estimates of attention shift speed (Egeth & Yantis, 1997; Pylyshyn & Storm, 1988). These judgments should also not be possible when the locus of spatial attention is drawn to another task. There should be a coupling between the shift order as determined by instructional inducements (as used here) and electrophysiological and eyetracking correlates of the locus of spatial attention over time. It also should not be possible to judge relations among three or more objects without first 'chunking' the objects into two groups, or selecting only two of the objects at a time to judge. Finally, to specifically show that the direction of the shift itself helps bind the relative locations of the objects, it may be possible to selectively adapt the shifting mechanism through repetitive judgments, to create relational illusions under conditions of brief presentation.

In conclusion, while we may have an intuition that we make visual spatial relationship judgments by simultaneously selecting multiple objects at once, the present results show that relations over space may instead be represented by a sequence of attentional shifts over space and time. This flexible mechanism would complement other more fixed mechanisms for visual structure representation, and provide a way for vision to compositionally represent arbitrary arrangements of objects.

Acknowledgements

We thank the following people for helpful discussion: Irving Biederman, Laura Carlson, Heeyoung Choo, Todd Handy, Kenneth Hayworth, Jim Hoffman, Jiye Kim, Mark Lescroart, Gordon Logan, Steve Luck, Marty Woldorff, Satoru Suzuki, and Geoff Woodman. We are grateful to Derek Tam, Alison Gschwend, Trixie Lipke, Roxana Malene, and Sally Martinez for their assistance in data collection.

References

Atchley, P., & Kramer, A. F. (2001). Object and space-based attentional selection in three-dimensional space. *Visual Cognition, 8*(1), 1-32.

Bahcall, D. O. & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research 39*, 71-86.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115-147.

Biederman, I., Lescroart, M., & Hayworth, K. (2007). Sensitivity to object-centered relations in LOC [Abstract]. *Journal of Vision, 7*(9):1030, http://journalofvision.org/7/9/1030/

Carlson, L. A. (2000). Selecting a reference frame. *Spatial Cognition and Computation 1*, 365–379.

Carlson, L. A. & Logan, G. D. (2001). Using spatial terms to select an object. *Memory & Cognition, 29*, 883-892.

Carlson, L. A., & Logan, G. D. (2005). Attention and spatial language. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 330-336). San Diego, CA: Elsevier.

Cavanagh, P. (2004). Attention routines and the architecture of selection. In Michael Posner (ed.), *Cognitive Neuroscience of Attention*. New York: Guilford Press, 13-28.

Chabris, C. F., & Kosslyn, S. M. (1998). How do the cerebral hemispheres contribute to encoding spatial relations? *Current Directions in Psychological Science, 7*(1), 8-14.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48*, 269-297.

Eimer, M. (1996). The n2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology, 99*(3), 225-234.

Enns J.T., and Rensink R.A. (1990). Influence of scene-based properties on visual search. *Science, 247*, 721-723.

Eriksen, C. W., & Schultz, D. W. (1977). Retinal locus and acuity in visual information processing. *Bulletin of the Psychonomic Society, 9*(2), 81-84.

Ghirardelli, T. G., & Folk, C. L. (1996) Spatial cuing in a stereoscopic display: Evidence for a `depth-blind' attentional spotlight. *Psychonomic Bulletin & Review*, *3*, 81-86.

Hayward, W. (2003). After the viewpoint debate: Where next in object recognition? *Trends in Cognitive Sciences, 7*(10), 425-427.

Hayworth, K., Lescroart, M., & Biederman, I. (2008). Explicit relation coding in the Lateral Occipital Complex [Abstract]. Journal of Vision, 8(6):35, http://journalofvision.org/8/6/35/

He, Z. J., & Nakayama, K. (1995). Visual attention to surfaces in three-dimensional space. *Proceedings of the National Academy of Sciences of the USA, 92*, 11155-11159.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.

Hickey, C., Di Lollo, V., & McDonald, J. J. (in press). Electrophysiological indices of target and distractor processing in visual search. *Journal of Cognitive Neuroscience*.

Hickey, C., McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, *18*(4), 604-613.

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society: Biological Sciences*, *393*, 1257-1270.

Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, *4*(2), 127-128.

Hopf, J. M., Boehler, C. N. Luck, S. J., Tsotsos, J. K., Heinze, H. J., & Schoenfeld, A. M. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences*, *103*, 1053-1058.

Hopf, J. M., Luck, S. J., Girelli, M., Hagner, T., Mangun, G. R., Scheich, H., & Heinze, H. J. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex*, *10*(12), 1233-1241.

Hummel, J. E. (2000). Where view-based theories break down: The role of structure in shape perception and object recognition. In E. Dietrich & A. Markman (Eds.), *Cognitive Dynamics: Conceptual Change in Humans and Machines* (pp. 157-185). Mahwah, NJ: Erlbaum.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480-517.

Hyun, J-S., Woodman, G. F., & Luck, S. J. (2009). The role of attention in the binding of surface features to locations. *Visual Cognition, 17*, 10-24.

Jager, G., & Postma, A. (2003). On the hemispheric specialization for categorical and coordinate spatial relations: a review of the current evidence. *Neuropsychologia*, *41*(4), 504-515.

Kimchi, R. (1992). Primacy of holistic processing and global/local paradigm: A critical review. *Psychological Bulletin, 112*(1), 24-38.

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review, 94*, 148-175.

Lescroart, M. D., Hayworth, K. J., & Biederman, I. (2009). Is there an object-centered map in LOC? [Abstract] *Vision Sciences 2009*.

Logan, G. D. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance, 20*(5), 1015-1036.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology, 28*(2), 103-174.

Logan, G. D. & Sadler, D. D (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and Space* (pp. 493-529). Cambridge, MA: MIT Press.

Logan, G. D., & Zbrodoff, N. J. (1999). Selection for cognition: Cognitive constraints on visual spatial attention. *Visual Cognition*, *6*, 55-81.

Luck, S. J., & Ford, M. A. (1998). On the role of selective attention in visual perception. *Proceedings of the National Academy of Science, 95*, 825-830.

Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology, 33*, 64-87.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*, 291-308.

Marrara M. T., & Moore C. M., (2000). Role of perceptual organization while attending in depth. *Perception & Psychophysics, 62*, 786-799.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of Basic Findings. *Psychological Review*, *88*, 375-407.

Miller, G. (1956). The magical number seven, plus or minus two. *Psychological Review*, *63*, 81.

Miller, G. A. & Johnson-Laird, P.N., Ed. (1976). *Language and perception*. Cambridge, Mass: Belknap Press of Harvard University Press.

Moore, C. M., Elsinger, C. L., & Lleras, A. (2001). Visual attention and the apprehension of spatial relations: the case of depth. *Perception & Psychophysics, 63*, 595-606.

Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175.

Oliva, A. & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*(12), 520-527.

Pashler, H. (1998). The Psychology of Attention. Cambridge, MA: MIT Press.

Pinker, S. (1980). Mental imagery and the third dimension. *Journal of Experimental Psychology: General, 109*, 254-371.

Pylyshyn, Z.W.; Storm, R.W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity. *Memory & Cognition, 14*(3), 191-20l.

Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, *130*(2), 273-298.

Reichardt, W. (1969). Movement perception in insects. In W. Reichardt (Ed.), *Processing of Optical Data by Organisms & Machines* (pp. 465-493). New York, NY: Academic Press.

Rensink R. A., O'Regan J. K., & Clark J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368-373.

Rosielle, L. J., Crabb, B. T., & Cooper, E. E. (2002). Attentional coding of categorical relations in scene perception: Evidence from the flicker paradigm. *Psychonomic Bulletin & Review, 9*(2), 319-326.

Sanocki, T., & Sulman, N. (2009). Priming of simple and complex scenes: Rapid function from the intermediate level. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 735-74.

Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition, 7*(1-3), 377-396.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80*(1-2), 1-46.

Singer, W. (1996). Neuronal synchronization: A solution to the binding problem? In R. Riascos Llinás, & P. Smith Churchland (Eds.), *The Mind-Brain Continuum: Sensory Processes* (pp.101-131). Cambridge, MA: MIT Press.

Tanaka, J. W. & Farah, M. J. (2006). The holistic representation of faces. In M. Peterson & G. Rhodes (Eds.), *Analytic and Holistic Processes in the Perception of Faces, Objects, and Scenes* (pp. 53-91). New York, NY: Oxford University Press.

Tarr, M. J. & Bulthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67(1-2), 1-20.

Tatler, B. W. (2002). What information survives saccades in the real world? *Progress in Brain Research*, *140*, 149-163.

Theeuwes J., Atchley P., & Kramer A. F. (1998). Attentional control within three-dimensional space. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1476-1485.

Treisman, A. 1996. The binding problem. *Current Opinion in Neurobiology, 6*, 171-178.

Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Pscyhology, 12*, 97-136.

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

Van Rullen, R. (2009). Binding hardwired versus on-demand feature conjunctions. *Visual Cognition*, *17*(1-2), 103-119.

Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*, 81-92.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics*, *56*, 495-500.

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, *400*, 867-869.

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 121-138.

Wolfe, J. M. (1998). What can 1,000,000 trials tell us about visual search? *Psychological Science, 9*(1), 33-38.

Yantis, S. (1988). On analog movements of visual attention. *Perception & Psychophysics, 43*, 203-206.

Figure Captions

Figure 1: A difficult spatial relationship search task. Find the target pair with the grey object on the left. Now find the second one.

Figure 2: Two possible classes of mechanism for visual spatial relationship judgments.

Figure 3: Sample stimulus for Experiment 1. Participants were instructed to report the spatial relationship between shapes of the relevant color while ignoring shapes of the other color.

Figure 4: (a) Average ERPs from PO7/8 electrodes contralateral to the closer object of relevant color (dark line) or contralateral to the farther object (grey line). More negative values (plotted upward) indicate shifts of attention toward that object. (b) Difference waves between the lines in Figure 4a, indicating shifts toward the near object (leftward deviation), or far object (rightward deviation).

Figure 5: Sample display for Experiment 2. Participants reported the relative relationships between the colors of the relevant shape set (either diamonds or circles). In the dual-task condition, they additionally reported whether the noise-degraded letter was a vowel or consonant.

Figure 6: For the *single-task* condition of Experiment 2, (a) Average ERPs from PO7/8 for electrodes contralateral to the closer object of relevant color (dark line) or the electrodes contralateral to the farther object (grey line). More negative values (plotted upward) indicate shifts of attention toward that object. (b) Difference waves between the lines in Figure 6a, indicating shifts toward the near object (leftward deviation), or far object

(rightward deviation).

Figure 6: (a) Identical analysis as shown in Figure 6, but for the *dual-task* condition of Experiment 2, including (b) differences waves.

Figure 8: Sample display for Experiment 3. The dotted lines indicate visual quadrant boundaries and were not present in the actual displays. Participants reported whether the target shape (in this case, cyan) was a diamond or circle.

Figure 1

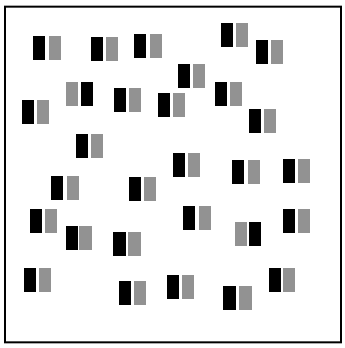


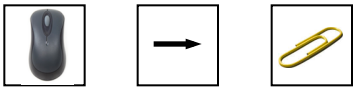
Figure 2

Is the paperclip to the right of the mouse?

'Hardwired' long-term representations

Problem: Need existing unit for each object combination, relation, & distance

Serial shift of selection



Existing recognition system with selection of one object at a time

Figure 3

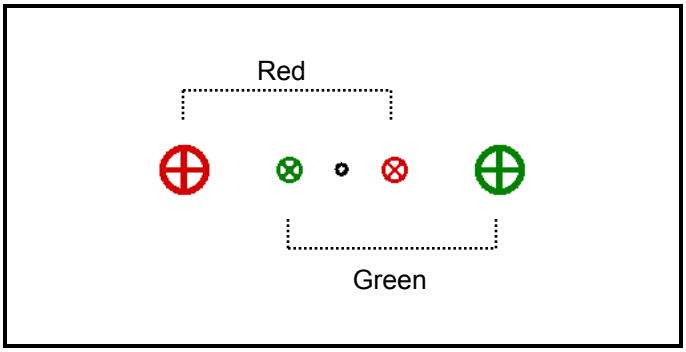


Figure 4a

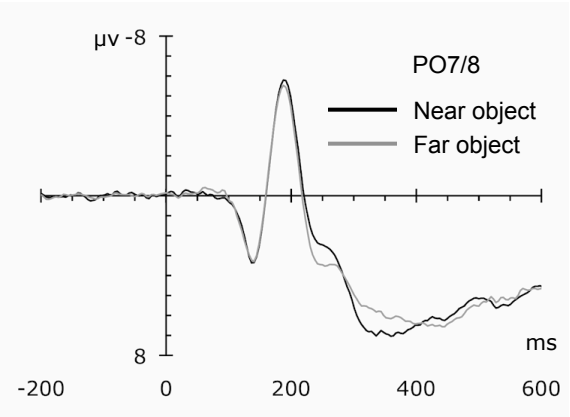


Figure 4b

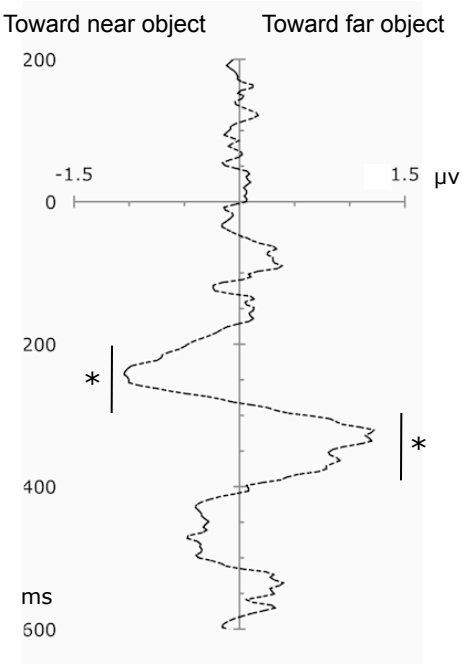


Figure 5

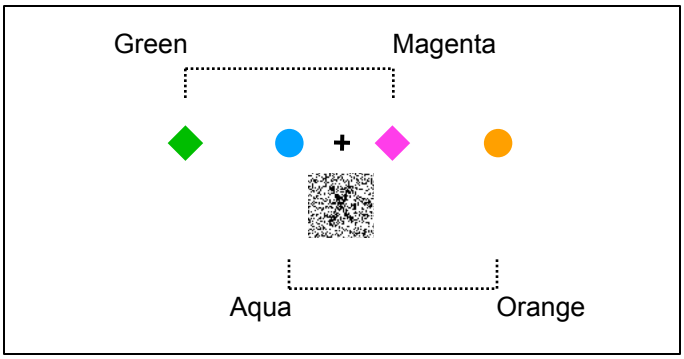


Figure 6a

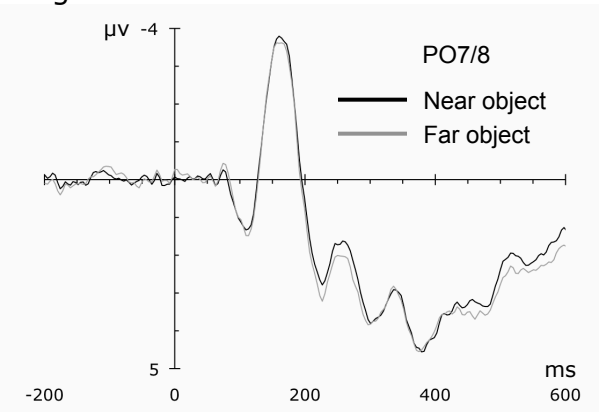


Figure 7a

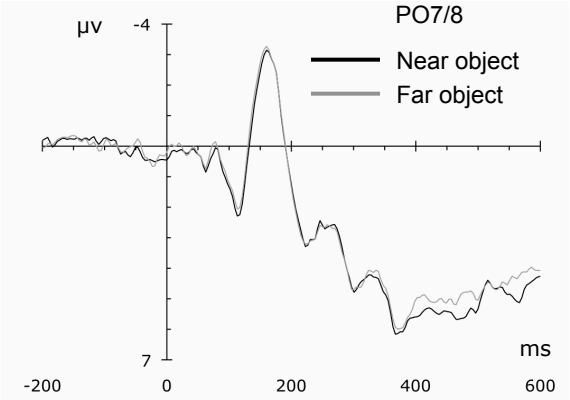


Figure 6b

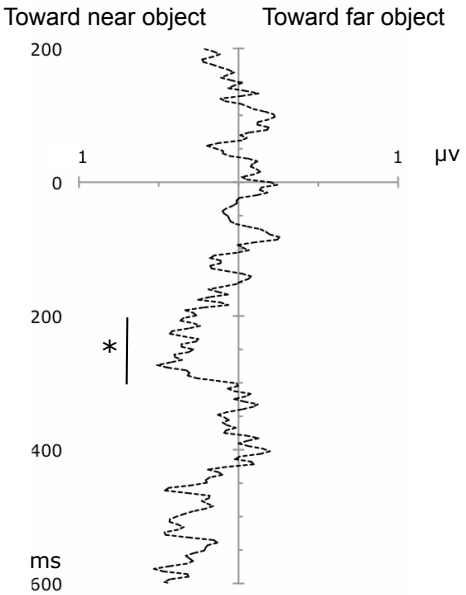


Figure 7b

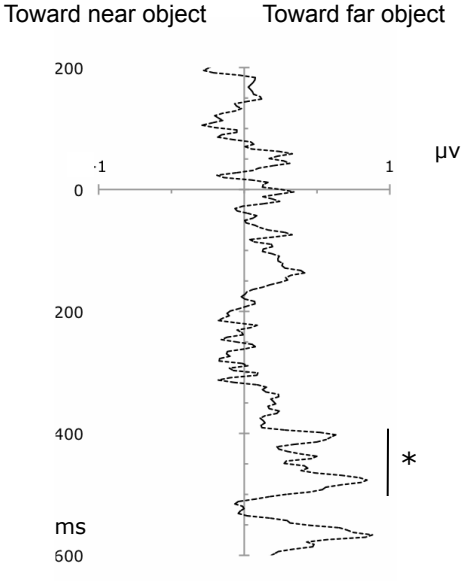


Figure 8