Flexible visual processing of spatial relationships

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: Flexible spatial relationships
Word Count: 7,996

Abstract

Visual processing breaks the world into parts and objects, allowing us not only to examine the pieces individually, but also to perceive the relationships among them. There is work exploring how we perceive relationships within structures with long-term representations, such as faces, common objects, or prototypical scenes. But strikingly, there is almost no work on the mechanisms that allow us to *flexibly* represent spatial relationships, e.g. between objects in a novel room, or the elements within a map, graph or diagram. We describe two classes of mechanism that might allow judgments of relations between objects. In the *simultaneous* class, both objects receive attentional selection concurrently. In contrast, we propose a *sequential* class, where objects are selected individually over time. We argue that this latter mechanism is more plausible even though it violates our intuitions. To demonstrate that shifts of selection do occur during spatial relationship judgments that feel simultaneous, we used an electrophysiological correlate of the locus of selection to demonstrate that observers do shift attention between the judged objects. Static structure across space may be encoded as a dynamic sequence across time. Flexible visual spatial relationship processing may serve as a case study of more general visual relation processing beyond space, to other dimensions such as size and numerosity.

[210 words]

Keywords: Attention, Selection, Spatial Relationships, Spatial Language, Binding, Comparison

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 1980 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 (VanRullen, 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 0907 6180") is possible, but much more difficult. Similarly, processing unfamiliar words may slow a reader (Rayner & Duffy, 1986), and recognition of visual feature conjunctions often requires focused processing (Treisman & Gelade, 1980).

We explore the flexible system that allows us to judge relative spatial relationships among objects in the visual world. Relational processing for some frequently encountered objects, such as the location and appearance of facial features (Tanaka & Farah, 2006) or the location of features, patterns, 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. Some have argued for the need for flexible systems that represent part structure *within* individual objects (Biederman, 1987), and there is even one proposal for how this structural description might work (Hummel & Biederman, 1992). But there is almost no existing work on

the mechanisms underlying flexible relational representation among separate objects. There is important related work in the spatial cognition literature on similar themes in relational processing, such as how the positions of objects are encoded in coordinate frames (e.g., Mou & McNamara, 2002; Rieser, 1989; Shelton & McNamara, 2001), or when positional information can be updated across viewpoint and reference frame changes (e.g., Wang, 2003). But while this work characterizes the representations of the *positions* of objects, it does not explore the mechanism that allows the visual system to extract the *relative positions* among objects.

The difficulty of extracting relative position might strike the reader as an odd problem – 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, and is no more available from position representations as it is on the retina. 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 your computer's keyboard is at horizontal position 4, and your mouse 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 higher level of representation is needed that compares the relative positions of the objects.

Explicitly representing these relations now seems to be a daunting problem. In a given scene, there are dozens or hundreds of objects, yet we feel that we have visual access to all of them simultaneously. This is extremely unlikely, as the number of spatial relations among a set of objects expands at a geometric rate given the number of objects. Two objects have one spatial relation, but a row of four objects has six relations, five objects ten relations – to skip ahead, ten objects have forty-five relations. An important constraint that we will place on the flexible relation processing mechanism is that our intuitions about its scope are unreliable. This sense of detail *must* be an illusion, and that actual relational representations could be extremely impoverished. Our impression of broad access to the details of the visual world is frequently

wrong in many other cases, such as the resolution and color content of the visual periphery. Our illusion of detail might rest on processes that seamlessly retrieve needed details 'on demand' (Noe & O'Reagan, 2000; Rensink, 2000).

A number of studies have used visual search tasks to reveal extreme limits on our ability to judge spatial relationships (see Figure 1). 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). Objects in a given spatial relationships may even hold a unique position as the most robustly difficult target for a visual search (Huang & Pashler, 2005; Palmer, 1994; Reddy & VanRullen, 2007; Wolfe, 1998). This difficulty 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 in other visual search tasks (Vickery, King, & Jiang, 2005). Other demonstrations use change detection tasks to show that processing of relative spatial relationships is slow and capacity-limited (Rosielle, Crabb, & Cooper, 2002). The capacity limit within these visual search tasks is not related to identification of objects within the relation, but instead to processing the spatial relationship among those object identities. 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 trivially easy (Logan, 1994; Logan, 1995; Logan & Sadler, 1996). Similarly, cueing the position of the pair that contains the target restores fast response times (Logan, 1994). This need to isolate a single pair of judged objects can even be seen in a far 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). These results are all consistent with the idea that in order to judge most types of spatial

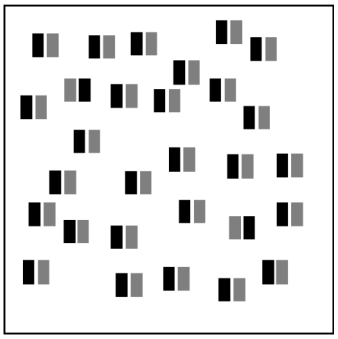

Figure 1: A difficult spatial relationship search task. Find the target pair with the gray object on the left. Now find the second one.

relationships, the visual system must *select* the relevant subset of objects for further processing, and relatively inhibit other aspects of a scene.

But what happens under this selection? Thus far, this process represents a 'black box'. We outline two classes of potential mechanisms that might allow the visual system to compare the relative spatial relationships between objects, based on emerging work from multiple laboratories. For simplicity, we will consider a single left-right judgment between just two objects. In the *simultaneous* class, the object pair receives attentional selection concurrently. We then propose a novel *sequential* class, where each object within the pair is selected individually over time.

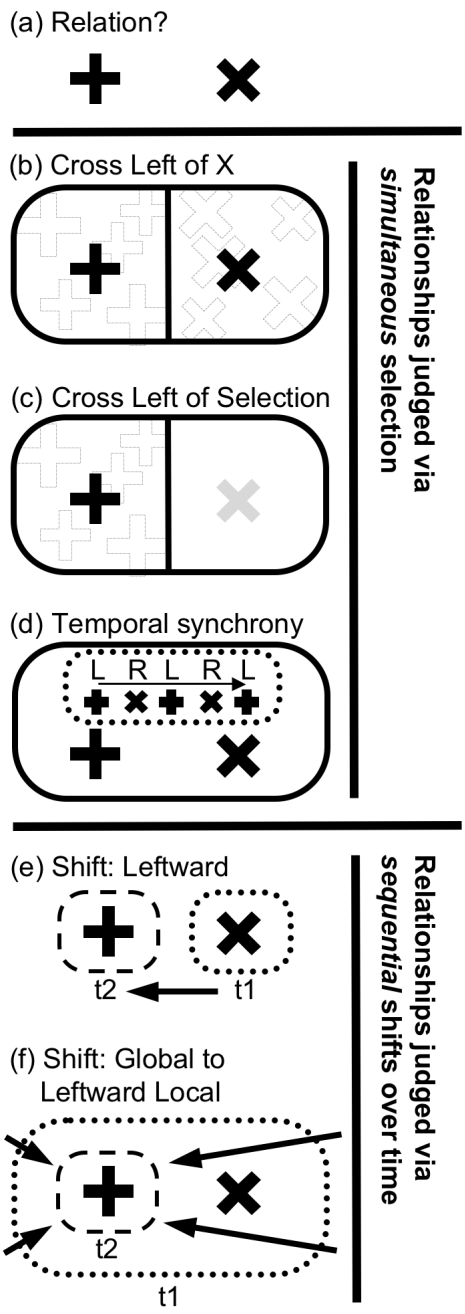

Figure 2: Two possible classes of mechanism for visual spatial relationship judgments.

Simultaneous selection

This class of mechanism compares the relative positions of two objects (see Figure 2a) by treating them as a single object, according to how the objects are selected by the 'spotlight' of attention. A pair of objects could be selected simultaneously, allowing higher-level areas of the ventral stream a strong representation of the object pair, and a relatively suppressed representation of the rest of the scene (Moran & Desimone, 1985; Reynolds & Desimone, 1999). Given this 'clipping' of the visual scene consisting of the selected pair of objects, one simple mechanism might recognize relative

relations with a long-term representation (e.g. a 'grandmother cell') for that relation among those objects. Figure 2b presents an example of such a long-term relation detector that fires upon encountering (+ x). This mechanism solves the problem of relation detection, but does not meet our requirement of flexibility, because it requires an existing representation for every possible configuration of every pair of objects. There is debate over whether such systems would cause an unrealistic combinatorial explosion of existing representations for the recognition of single objects (Biederman, 1987; Hayward, 2003; Hummel, 2000; Tarr & Bulthoff, 1998), but this problem would be compounded for relations among multiple objects, which must consider combined identities of *two* objects, not to mention the angle between them. Despite such pessimism, this mechanism does almost certainly exist for *inflexible* processing of some simple and frequently encountered relations that merit efficient long-term representations (Biederman, Lescroart, & Hayworth, 2007; see General Discussion).

For more flexible relational processing, there are at least two ways to reduce this combinatorial explosion to manageable levels. Long-term detectors might 'cheat' by detecting relations between more abstracted properties such as relative differences in brightness (the brighter object is on the right), size (the larger object is on the left), or in the case of Figure 2b, orientation (the object with more diagonal segments is on the right). Knowing that the small object is to the right of the large object could be enough information to conclude that your bicycle is to the right of your garage.

An intriguing second way to cheat would be to delete one object from the long-term recognition network, by exploiting networks (presumably in the Lateral Occipital Complex, or LOC) whose receptive fields contain response biases that depend on where in the field an object appears (Biederman, Lescroart, & Hayworth, 2007). Figure 2c depicts an example of a network that prefers (+), a "+" on the left side of the *current window of selection*. Another network might prefer the "+" to be on the right or top side of the window. More complex relationships (e.g. diagonal) could be coded via combinations of other dimensions (e.g., 'above' paired with

'right'). This account is consistent with evidence that when presenting a pair of objects twice over time, fMRI measures of the LOC show a greater release from adaptation (that is, activation is higher on a subsequent trial) when the two objects flip their respective positions, relative to when they translate the same distance while maintaining their original relation. This result is consistent with the possibility that a new set of long-term representations represents the group when the relation is changed (Biederman, Lescroart, & Hayworth, 2007).

The mechanisms within this simultaneous class require that the observer simultaneously select both objects, because *the window of selection establishes the reference frame for the relations*. The "+" can be judged as to the left of the "x" because it is on the left side of the currently selected region of the visual field.

The networks described above are plausible, and likely underlie our perception of some types of relations. But such mechanisms require long-term representations, and would have difficulty representing relations between novel objects, relations between objects that are only subtly visually different, or even visually identical (e.g., on the dinner table, which fork was mine?), or relations among objects in crowded environments, making it difficult to select the relevant ones simultaneously. At minimum, another more flexible mechanism is needed in such instances.

The mechanism depicted in Figure 2d does exhibit this type of flexibility. Both objects in a relation are selected, and the spatial relationship between the objects is represented by a dynamic network where feature units (e.g. +, x) fire in temporal synchrony (Gray & Singer, 1989; Milner, 1974) with spatial units (e.g., position 4, position 6) that describe their locations (Hummel & Biederman, 1992). Later stages of processing extract explicit spatial relations, and temporal synchrony links each relation (e.g. left-of) with the proper object (e.g., +). Unlike the mechanisms shown in Figure 2b/c, this mechanism uses separate units to represent object identity and object location, allowing flexible representation of any simple relation (see Hummel & Biederman, 1992, and Hummel, 2000, for discussion of the benefits of such *disjunctive* coding). Though this mechanism is importantly different than the previously described long-term

representations, for present purposes it shares a common characteristic - it also requires simultaneous selection of both objects in a relation.

Sequential shifts

We propose a new class of mechanism that might allow flexible processing of spatial relationships. According to the shift account, only one object within the pair is selected at a time. A significant motivation for this mechanism is that many aspects of processing *single* objects, a prerequisite to processing relations between objects, appear to require selection of that individual object. Identifying objects in many cases appears to require amplifying relevant signals from those objects, while suppressing irrelevant noise from other objects (Luck et al., 1997; Treisman, 1996; Treisman & Gelade, 1980; but see VanRullen, 2009 for proposed exceptions). Localizing objects within the visual field may also require individual selection, perhaps due to the coarse coding of location by the ventral visual stream (Hyun, Woodman, & Luck, 2009; Luck & Ford, 1998).

Even if individual object selection can provide a way to resolve these problems, how could individual selection allow explicit recovery of relative position? We propose a mechanism that could recover relation information from the spatial pattern of shifts, while adding only a trivial amount of computational overhead. The solution would be to record the *direction of the attentional shift* from one object to the other. This shift could be briefly held in heightened activation (see Figure 2e), by a circuit similar to a detector for low-level motion (Reichardt, 1969). These detectors could be placed over representations of visual selection and salience, which may be subserved by the lateral intraparietal area (Gottlieb, 2007; Serences & Yantis, 2006) or inferior intraparietal sulcus (Todd & Marois, 2004; Xu & Chun, 2009). Or, instead of detecting shifts, there could also be an efference copy of the shift direction taken from the shift command itself.

In the example in Figure 2e, the locus of selection could be moved to the right object. Selection could then shift toward the left object, leading to a heightened activation of the "+". This heightened activation, combined with the high activation of the representation of a leftward shift, provides the information necessary to conclude that the "+" is on the left of whatever object was last selected. The starting point for the shift might not be on one of the objects, but the center point between the objects. Or the sequence might start at a global scope (Figure 2f), with the locus of selection 'zooming in' in a leftward direction, giving an initial summary representation of the objects ("There's a + and an x, and a horizontal arrangement"), followed by the relational information ("The + is on the left of that arrangement"). The shift might even need to occur multiple times, from one focus and back again, to gain redundancy in the coding of the relation. Such 'back and forth' shifts might even be necessary to encode the relation symmetrically. A single shift might produce a representation of the "+ on the left", and a second shift might be necessary to see an "x on the right".

Using shifts of selection as a *source of information* would be an unusual role for selection, which often is thought to amplify relevant information at the expense of irrelevant information (Hillyard, Vogel, & Luck, 1998; Luck, et al., 1997). 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; Jolicoeur, Ullman, & Mackay, 1986; Logan & Zbrodoff, 1999; Ullman, 1984). The shift account would be compatible with studies showing that when viewing or visualizing a previously viewed scene, the sequence of eye movements across objects is often similar to the sequence observed in the previous view (Brandt & Stark, 1997; Noton & Stark, 1971), or similar to the order in which an experimenter presented the objects (Ryan & Villate, 2009). Such results suggest that memory for spatial information in a scene is accompanied by temporal information for the sequence in which objects were processed.

Most importantly, in contrast to the simultaneous class of models, the shift mechanism requires that the locus of spatial selection shift sequentially *at least once* during spatial

relationship judgments. Under this mechanism, no relational information can be recovered unless this shift occurs.

In summary, the visual system might represent spatial relationships among objects using processes that involve either simultaneous or sequential selection of the judged objects. The simultaneous mechanism almost certainly exists for some types of judgments. We argue that the sequential mechanism is necessary for flexible judgments. Because the sequential mechanism violates our conscious experience of simultaneous selection of both objects in a simple relation, below we offer empirical evidence that during a simple relational judgment, the locus of selection does shift between the objects over time.

Detecting shifts of spatial attention with an electrophysiological correlate

There has long been interest in tracking 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 demonstrated 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, 2009). The signal likely originates in

the lateral extrastriate and inferotemporal cortex (Hopf, et al., 2000), and appears to be controlled by more frontal structures such as the frontal eye fields (Cohen et. al., 2009).

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.

Experiment

In the experiment we only measure these shifts with an object-relative analysis. That is, we always ask whether the shifts were toward or away from a given object. 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, it is difficult to detect these types of shifts with the N2pc technique, 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.

Because the N2pc technique requires averaging the pattern of shifts across many trials, if participants do not adopt a consistent sequence then the average may reflect a mixture of shift

patterns. As an extreme example, if a participant first shifted toward the "+" on odd trials, and the "x" on even trials, the average across trials would show no evidence for shifts. We therefore use a display organization that biases the shift direction toward one object. A 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 visual search experiments (Carrasco, Evert, Change, & Katz, 1995; Wolfe, O'Neill, & Bennett, 1998), including one that used posterior contralateral negativity signals to track shifts of attention (Woodman & Luck, 2003).

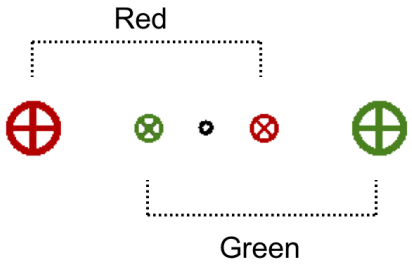

Figure 3: Sample stimulus for experiment. Participants were instructed to report the spatial relationship between shapes of the relevant color while ignoring shapes of the other color.

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 of 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 task-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. If the relation is judged by selecting objects simultaneously, then no pattern of shifts should be evident, and there should be no difference between the signal from electrodes contralateral to the near and far objects.

Methods

Participants

15 Northwestern University undergraduates participated 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, 33.6 pixels per degree. In the stimulus display, a fixation point (always visible during trials) was flanked by two red or green shapes on each side. Each shape was either an "+" or a "x" surrounded by a circular border. The far shapes were 3.57° from the fixation point, 1.13° in diameter, and had 0.15° thick segments, and the near shapes were 1.19° from the fixation point, 0.60° in diameter, and had 2 pixel thick segments. Within each color pair, one shape was a "+" and the other was an "x", and one shape was green (24 cd/m^2) and the other was red (14 cd/m^2) (see Figure 3). The color values were approximately 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 grand average of each subject's median value.

Procedure:

Before starting the experiment, all subjects were given fixation training using a flickering pattern that 'jumps' when fixation is broken, which has been shown to drastically improve fixation performance (Guzman-Martinez, Leung, Franconeri, Grabowecky, & Suzuki, 2009). 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 shapes with a specified color, 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. When responding, participants were instructed to make accuracy their first priority, and speed their second priority.

EEG Recording

ERP was recorded using a Biosemi Active 2 EEG/ERP system. The DC recording was made at 512Hz with a hardware low-pass filter, and then was decimated in software to 256Hz. 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 2 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

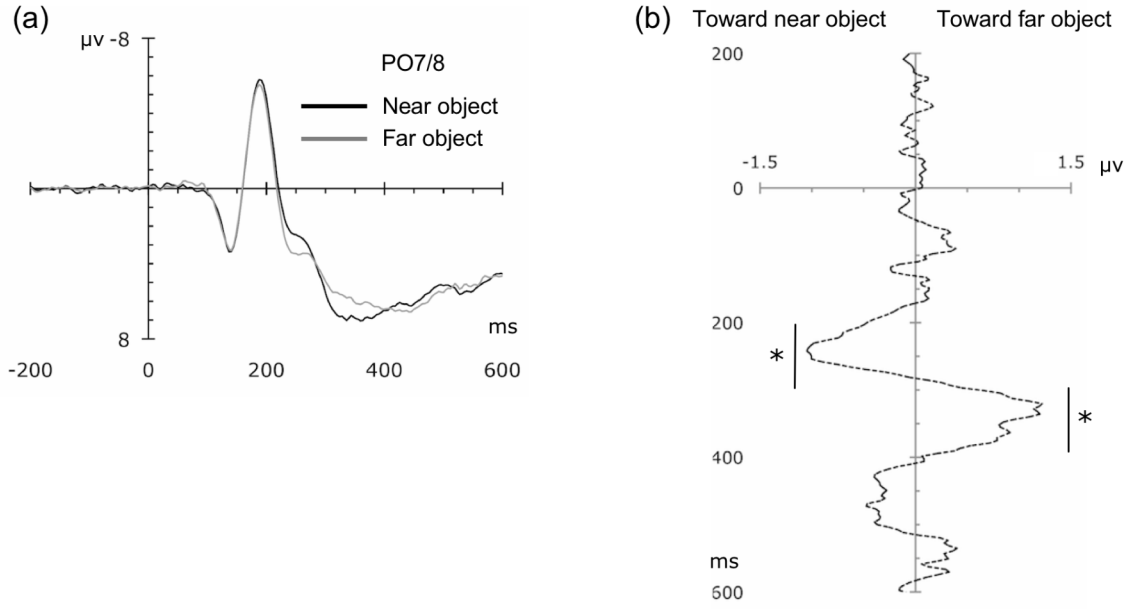

Figure 4: (a) Average ERPs from PO7/8 electrodes contralateral to the closer object of relevant color (dark line) or contralateral to the farther object (gray line). More negative values (plotted upward) indicate shifts of attention toward that object. (b) Difference waves between the lines in Figure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

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 (at most a small fraction of a degree; Hillyard & Galambos, 1970). 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).

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 this 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 pattern appeared where amplitudes were more negative contralateral to the far target (Difference M=0.82uv), $t(8)=4.4$, $p=0.002$. This pattern of activity supports our prediction that participants would first shift to the near object and then shift to the far object. However, we note that the activity seen at times 300-400ms post-stimulus does not always reflect a second shift but instead may reflect a separate positive component (Ptc) that can result from the first shift (Hilimire, Mounts, Parks, & Corballis, 2009). Without additional control experiments we cannot conclusively determine whether our later activity reflects a second shift or is the result of the first shift. But critically, we can conclude that there was at least one shift.

Participants in the experiment shifted attention between two shapes during a spatial relationship judgment, suggesting that such shifts may typically accompany relational judgments. These results cannot directly show that the shift itself carries relational information. Instead, serial inspection of each shape may have been necessary in order to encode the identity and/or location of each object. An ideal demonstration would be to show, within the same display, that shifts occur during tasks that require spatial relationship information, but not when tasks simply require identity information. Due to the high level of automaticity of attentional shifts, this demonstration may be difficult to construct. Imagine noting whether a red and a green circle were the same or different color. After seeing the display, even if the spatial relationship between

the objects were irrelevant, you would still be able to report it. If the relationship can be reported, the relationship was recognized, and the shift account predicts that selection must have shifted among the judged objects. Demonstrating the need for shifts to construct spatial relationships would therefore require displays in which object identities can be reported, but the spatial relationship among them cannot. One display where this occurs is an 'illusory conjunction', where under dual task conditions, briefly presented displays of simple colored shapes or letters can lead to incorrect reports of which colors were paired with which object locations (Treisman & Schmidt, 1982). Because these incorrect reports only occur on a small percentage of trials for a subset of observers, it is difficult to collect sufficient ERP data to examine attention shifts with such displays. Our laboratory is currently testing whether it is possible to examine correlations between trials containing this illusion and the presence of attentional shifts.

But critically, even if the observed shifts are due to the need to identify and/or localize each individual object, then this effect presents a strong challenge to any model of spatial relationship processing that relies on simultaneous selection. If participants had been able to use, e.g., a long-term representation of vertical/horizontal lines to the left or diagonal lines to judge the relationship among the objects, then the shifts should not have been necessary. One might also point to the interleaved arrangement of the objects as making such simultaneous selection more difficult. But the visual world requires exactly this type of 'interleaved' judgments all the time, in both the natural environment (e.g. scenes) and constructed displays (e.g. diagrams). Any potential mechanism for spatial relationship processing must to be able to reconstruct relations among objects that are inspected sequentially over time.

A similar potential critique is that in the present displays, once the positions of the objects are known, the observer can 'cheat' by using the relative position of just a single object to complete the task. But again, this is true more broadly in any real world task. Relative spatial relationship judgments have only one degree of freedom, and once the relevant objects are identified, only one object needs to be inspected to resolve the relation (see the 'global to local' mechanism in Figure

2f for an example). Using shifts of selection would allow the visual system to efficiently exploit this property, even if there is no conscious trace of the sequential nature of the underlying mechanism. Furthermore, if we already constantly shift attention among objects of primary interest, then the shifts themselves can serve as a carrier for the relative position of those objects, with a minimum of additional computational overhead.

General Discussion

When we judge visual spatial relationships among objects, we may feel as though we attend to both objects in the relation simultaneously. Indeed, one class of mechanism allows relation judgment when selection of objects is simultaneous. We argue for the existence of a novel class of relation judgment mechanism where selection can shift among the judged objects over time, and offer electrophysiological evidence that shifts do occur in a simple judgment that gives the impression of simultaneous selection. We suggest that the visual system could use dynamic information about these shifts to represent the spatial relationship among the objects in a computationally feasible way. The shifts require simple circuitry to detect and store, and shifts of attention could automatically produce relational information as a by-product of natural attentional exploration of a scene.

Relation to other models of spatial relationship judgment

This sequential shift mechanism presents specific ways to implement 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 'spatial 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 additional horizontal displacement. Another stage binds the

objects to their correct roles in the relation. The present account shares some characteristics with this model, and specifies many steps at a lower level of implementation. However, some characteristics are different. For example, evaluating how well a set of objects matches a spatial template for a given relation would not involve a separate stage. Instead, the 'typicality' of the relation would be determined by how well the direction of the shift (the vector itself) matches the prototypical shift vector orientation for that relation.

The sequential shift account also shares characteristics with the Attention Vector Sum (AVS) model of Regier & Carlson (2001). The AVS 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 on 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 AVS such that instead of summing vectors, only one vector is created and evaluated. The starting point of this vector could be 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 specified by Regier & Carlson (2001).

'Inflexible' relational judgments

Here we divide the taxonomy of flexible spatial relationship judgments into those that require simultaneous vs. sequential selection. But some types of relation detectors may not require selection in the first place, and instead they may operate broadly across the visual field. The tradeoff is that these detectors may be extremely inflexible, and respond only to highly specific patterns in the environment (see VanRullen, 2009 for discussion of similar detectors for other

types of visual features). This possibility is demonstrated by a few visual search tasks for within-object relations that are surprisingly efficient. 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). 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 allowed participants to group and reject them efficiently, leading to quick access to the one remaining object (Wang, Cavanagh, & Green, 1994). In a similar example, observers quickly find a shaded circle that appears convex among shaded circles that appear concave, when the convexity is signaled by shading cues that exploit the visual system's assumption of overhead lighting (e.g., Ramachandran, 1988). But the patterns that these detectors process appears to be highly specific, such that subtle changes to the stimuli (such as slightly 'breaking apart' the faces of the cube) can sharply impair processing efficiency (Enns & Rensink, 1990; Ramachandran, 1988).

Categorical vs. Coordinate spatial relationship judgments.

The type of spatial relationship judgment that we consider here, where objects identities and locations must be matched with coarse categories such as "left of" or "above", while ignoring precise details such as the distance between objects, have previously been called *categorical* spatial relationships (e,g., Kosslyn, 1987). This label differentiates categorical judgments from another type of 'spatial relationship' judgment with substantially different processing requirements. *Coordinate* judgments allow observers to ignore the identities of objects, and instead make precise judgments about the metric distance between them, or the shape of their global configuration (see Chabris & Kosslyn, 1998). There are other judgments that may be similar in their processing requirements, such as vernier acuity tasks, where observers are asked to discriminate fine differences in the alignment of two lines (Shiu & Pashler, 1995; Yeshurun & Carrasco, 1999), or line bisection tasks, where observers are asked to mark the precise midpoint

of a line (Jewell & McCourt, 2000; McCourt & Jewell, 1999). Behavioral, neuroimaging, and neuropsychological evidence suggest that categorical and coordinate judgments are dissociable processes (Chabris & Kosslyn, 1998; Jager & Postma, 2003; Kosslyn, 1987).

In a particularly relevant example of this dissociation, two recent studies show that the speed of categorical and coordinate spatial relationship judgments interacts with the size of the window of selection (Borst & Kosslyn, 2010; Laeng, Okubo, Saneyoshi, & Michimata, 2010). Encouraging observers to select smaller areas of the visual field (either with small flashing cues or priming with a task requiring 'local' processing) gives a relative benefit to categorical judgments, while pre-cueing a large area surrounding both objects gives a relative benefit to coordinate judgments. We propose that this effect occurs because categorical spatial relationship judgments require the selection of one object at a time, matching a smaller processing scope, while coordinate spatial relationship judgments require the selection of multiple objects simultaneously (allowing evaluation of the size or shape of the envelope surrounding them), matching a larger processing scope.

Beyond left and right

There are types of relations that may be harder to explain with an attentional shifting mechanism. For example, how would this account deal with front-back relations, which also lead to inefficient visual search (Moore, Elsinger, & Lleras, 2001)? 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. One possibility is that as we make eye movements between the near and far objects,

a similar motion detector could signal the direction of *changes* in the vergence angle of the eyes - when this angle becomes more acute, the currently fixated object is the farther one. A second intriguing possibility is that the visual system might exploit correlations with depth, such as the tendency for farther objects to appear retinotopically higher than other objects on the same ground plane.

Inside-outside relationships could be supported by this shifting mechanism. There is a large body of existing work on shifting the locus of selection between global and local scales (e.g. Kimchi, 1992). To use a shift of attention to perceive an inside-outside relation, we would only need to add a detector circuit that fired whenever the scale switched from local to global (expanding), or vice-versa (narrowing). 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 from the local to the global scale and were now attending to the basket.

Finally, the relationships discussed here have all been object-relative judgments made within a retinotopic frame of reference. But relational judgments can be made relative to other reference frames, such as the head, the body, the ground plane, or other external objects (Carlson, 2000; Mou & McNamara, 2002; Rieser, 1989; Shelton & McNamara, 2001). For the present work, we cannot distinguish a retinotopic frame from any other frame. If the direction of the attentional shift is coded in a retinal reference frame, there would need to be a translation mechanism between the coordinate space of this frame and the frame needed for a given task. This translation could either be of the locations that the shift mechanism operates over, or of the shift direction itself after it has been made over a retinotopic representation. The latter option seems more computationally simple.

Connections between visual space and spatial language

There are strong similarities between such visuospatial representations of relations, and linguistic descriptions with the perceptual processing of spatial relations between objects (Carlson

& Logan, 2005; Logan, 1995; Logan & Sadler, 1996). There are similarities in the use of reference frames between visuospatial representations and spatial language. Linguistic descriptions refer to 'target' objects relative to 'reference' objects (e.g., "The x is to the right of the +"). In visuo-spatial representations, multiple factors influencing the choice of reference frame (Culjpers, Kappers, & Koenderink, 2001; Mou & McNamara, 2002). In spatial language, there are similar semantic, action, and other experience-based properties of objects that can guide the assignment of this target and reference status (Taylor & Tversky, 1992; 1996). Finally, the spatial layouts that are considered 'acceptable' for a given relationship (e.g., is an object that is to the upper left of another as good an example of 'above' as is an object directly above another?) are similar between the two domains (Hayward & Tarr, 1995; Logan & Sadler, 1996; Regier & Carlson, 2001).

The strength of these similarities has led some to propose that spatial language is grounded by an underlying perceptual representation (Crawford, Regier, & Huttenlocher, 2000; Regier & Carlson, 2001). An attractive quality of our sequential representation of visual spatial relationships is that (a) it could serve as this underlying perceptual representation, and (b) it is in a similar representational format to language. Because linguistic descriptions of space require that only one object be verbalized at once, the structure of linguistically specified spatial relationships is necessarily sequential. The sequential shift account proposes that the perception of a relationship between two objects requires the sequential selection of the two objects, paired with a relational term consisting of the shift direction. Thus, the signal over time within the visual system would be "object 1, right shift, object 2". The reference object might be the starting point of the attentional shift, and the target object the ending point. The relationship "object 1 is to the left of object 2" similarly collapses spatial structure into a message over time. This link would be consistent with the close ties between the dynamics of sequential eye movements across scenes and the comprehension of linguistic descriptions of those scenes (Altmann & Kamide, 2009; 2007), as well as the production of descriptions of those scenes (Griffin & Bock, 2000). Such

similar patterns over time could help translate between visual and linguistic representations of scene structure (Clark & Chase, 1972).

The way that the information is visuospatially depicted can also have a strong effect on the linguistic descriptions that people produce to describe them (Shah, Mayer, & Hegarty, 1999; Zacks & Tversky, 1999). Representing the same information with a bar graph can lead to conclusions about the relation between two discrete data points, with one being (e.g.) "higher" than the other. For a line graph, the same data might be described by a participant as showing a trend, involving a value (e.g.) "rising". The association also works in reverse, where different linguistic descriptions of data can lead participants to produce the associated graph type (Zacks & Tversky, 1999). Such differing conclusions at linguistic or other 'cognitive' levels may be driven by differences in the way that the relations in a graph are encoded by the visual system. Line graphs might encourage simultaneous selection, leading to conclusions of trends, while bar graphs might require sequential selection, leading to conclusions of discrete comparisons.

This link to language could present a solution to a problem encountered by any account of flexible spatial relationship representation - how do we judge or store relations among more than two objects? While chunking objects into hierarchically organized groups might suffice in some case (e.g. object A is to the left of group BC), other cases might require a more complex conjunction of relations (e.g. object A is to the left of B which is to the left of C). There may be memory representations that can store the results of recent relational judgments. But language may also play a key role in guiding attentional sequences, and storing the information that they reveal. Linguistic representations are already known to buffer visuospatial representations. Among children who have difficulty remembering visual left-right spatial relations between simple shapes, cueing relations linguistically (e.g., "Look - the red one is on the left") creates a more robust representation that leads to higher performance (Dessalegn & Landau, 2008). Several control experiments suggested that this benefit was related to the way that the linguistic description highlights both objects while still specifying a direction of the relation between them.

The linguistic cue may have guided the children to create a sequence, and encouraged the child to use language to store the result of that sequence.

Conclusion

While we may have an intuition that we make visual spatial relationship judgments by simultaneously selecting multiple objects across space, we instead argue that spatial relations may be constructed by a sequence of attentional shifts over space and time. This flexible mechanism would complement other long-term mechanisms for visual structure representation, and language might aid in the construction of compositional representations of arbitrary arrangements of objects.

We dissociate mechanisms that process spatial relations among objects simultaneously from those that process relations sequentially. Could this dissociation apply for relations beyond space? There is little work examining how we process the most simple visual relations or comparisons, along dimensions such as brightness, size, orientation, and number. Which bag contains more grapes? Which building is larger? Such decisions present the same problems found in spatial relationship judgments (which might be rephrased similarly as "which object is righter?").

Relative magnitude judgments might rely on simultaneous selection of two objects, followed by a comparison to a long-term representation (e.g., for a large object to the left of a small object), but we suspect a sequential process. The only architectural change required would be to place the 'motion detectors' not across a topographic representation of space, but across abstract representations of dimensions like brightness, size, orientation, or number. One-dimensional representations may exist for such domains (Cantlon, Brannon, & Platt, 2009; Kadosh, Lammertyn, & Izard, 2008; Pinel et al., 2004; Walsh, 2003) and in the domains of number and time they are often called 'accumulators' (Feigenson, Dehaene, & Spelke, 2004; Meck & Church, 1983). Placing a simple 'motion detection' circuit over such representations could generate

relational information automatically during sequential selection of objects or collections with different values, for any abstracted dimension. This mechanism would not suffice to process more complex relations (e.g. "Mary loves John") (Gentner & Loewenstein, 2002; Hummel & Holyoak, 2003; Halford, Wilson, & Phillips, 2010), but could generate relative magnitude judgments within a single dimension. The existence of this mechanism would suggest an exciting possibility: our ability to judge relative magnitudes could be credited to visual circuitry designed to detect motion in the world, co-opted to detect motion in the mind.

Acknowledgements

We thank the following people for helpful discussion: Irving Biederman, Laura Carlson, Joan Chiao, Heeyoung Choo, Banchiamlack Dessalegn, Todd Handy, Kenneth Hayworth, Jim Hoffman, Jiye Kim, Mark Lescroart, Gordon Logan, Steve Luck, Ken Paller, 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. This work was supported by NSF SLC Grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC).

References

Altmann, G.T.M. & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: eye-movements and mental representation. *Cognition, 111*, 55-71.

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*(1)*,* 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.

Borst, G. & Kosslyn, S. M. (2010). Varying the scope of attention alters the encoding of categorical and coordinate spatial relations. *Neuropsychologia, 48*, 2769-2772.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cogntiive Neuroscience, 9,* 27-38.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences, 13*(2) 83-91.

Carlson, L. A. (2000). Selecting a reference frame. *Spatial Cognition and Computation*, *1*(4), 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.

Carlson-Radvansky, L. A., & Radvansky, G. A. (1996). The influence of functional relations on spatial term selection. *Psychological Science, 7,* 56-60.

Carrasco, M., Evert, D. L., Change, I, & Katz, S.M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics, 57*(8)*,* 1241-1261.

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

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.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472-517.

Cohen, J.Y., Heitz, R.P., Schall J.D., & Woodman, G.F. (2009). On the origin of event-related potentials indexing covert attentional selection during visual search. *Journal of Neurophysiology, 102*, 2375-2386.

Crawford, L. E., Regier, T., & Huttenlocher, J. (2000). Linguistic and non-linguistic spatial categorization. *Cognition*, *75*, 209-235.

Cuijpers, R. H., Kappers, A. M., & Koenderink, J. J. (2001). On the role of external reference frames on visual judgements of parallelity. *Acta Psychologica*, *108*, 283-302.

Dessalegn, B., & Landau, B. (2008). More than meets the eye: the role of language in binding and maintaing feature conjunctions. *Psychological Science, 19(2),* 189-195.

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.

Feigenson, L., Dehaene, S., & Spelke, E.S. (2004). Core systems of number. Trends in Cognitive Sciences (8), 7, 307-314.

Gentner, D., & Loewenstein, J. (2002). Relational language and relational thought. In J. Byrnes & E. Amsel (Eds.), Language, Literacy, and Cognitive Development (pp. 87-120). Mahwah, NJ: LEA.

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.

Gottlieb, J. (2007). From thought to action: the parietal cortex as a bridge between perception, action, and cognition. *Neuron, 53*, 9–16.

Gray C. M. &Singer W. (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences USA, 86*, 1698-1702.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274- 279.

Guzman-Martinez, E., Leung, P., Franconeri, S. L., Grabowecky, M., & Suzuki, S. (2009). Rapid eye-fixation training without eye tracking. *Psychonomic Bulletin & Review. 16*, 491-496.

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.

Hayward, W. G., & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition, 55*, 39-84.

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., & Galambos, R. (1970). Eye movement artifact in the CNV. *Electroencephalography and Clinical Neurophysiology, 28*, 173–182.

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.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220-264.

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.

Jewell, G. & McCourt, M. E. Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia, 38*, 93–110 (2000).

Jolicoeur, P., Ullman, S., & Mackay, L. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition, 14*, 129–140.

Kadosh, R. C., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology, 84*(2), 132-147.

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.

Laeng, B., Okubo, M., Saneyoshi, A., and Michimata, C. (2010). Processing spatial relations with different apertures of attention. *Cognitive Science, 1-33*.

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.

McCourt, M. E. & Jewell, G. (1999). Visuospatial attention in line bisection: stimulus modulation of pseudoneglect, *Neuropsychologia, **37*(7)**, 843–855.

Meck, W. H. & Church, R.M. (1983). A mode control model of counting and timing processes.*Journal of Experimental Animal Behavior, 9*, 320-334.

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, MA: Belknap Press of Harvard University Press.

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

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.

Mou, W. & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 162–170.

Noë, A. & O'Reagan, J. (2000). Perception, attention and the grand illusion. *Psyche: An Interdisciplinary Journal of Research on Consciousness, 6*(15).

Noton D, & Stark L. (1971). Eye movements and visual perception. *Scientific American, 224*(6), 35-43.

Oliva, A. & 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.

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research, 13*, 1703-1721.

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

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments, *Neuron, 41*, 983–993.

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

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331*(14), 163-166.

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.

Reddy, L. & VanRullen, R. (2007). Spacing affects some but not all visual searches: implications for theories of attention and crowding. *Journal of Vision, 7*(2):3, 1-17.

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. (2000). The Dynamic Representation of Scenes. *Visual Cognition, 7*(1-3), 17-42.

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.

Reynolds, J. H. & Desimone, R. (1999) The Role of Neural Mechanisms of Attention in Solving the Binding Problem, *Neuron, 24*(1), 19-29.

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.

Ryan, J. D., Villate, C. (2009). Building visual representations: The binding of relative spatial reltionships across time. *Visual Cogntion, 17*, 254-272.

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.

Serences, J. T. & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*, 38-45.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology, 91*, 690-702.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology, 43*, 274-310.

Shiu, L-P. & Pashler, H. (1995). Spatial attention and vernier acuity. *Vision Research*, *35*, 337-343.

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.

Taylor, H. A. and Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, *31*, 261-282.

Taylor, H. A. and Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, *35*, 371-391.

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.

Todd, J., & Marois, R. (2004). Capacity limit of visual sort-term memory in human posterior parietal cortex. *Nature, 428*(15), 751-754.

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 Psychology, 12*, 97-136.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*, 107-141.

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

VanRullen, 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.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483-488.

Wang, R. F. (2003). Spatial representations and spatial updating. In D. E. Irwin & B. H. Ross (Eds.), *The Psychology of Learning and Motivation, 42, Advances in Research and Theory: Cognitive Vision (*pp. 109-156). San Diego, CA: Academic Press.

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.

Wolfe, J. M., O'Neill, P., & Bennett, S. C. (1998). Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Perception & Psychophysics, 60,* 140–156.

Xu, Y. & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences, 13*, 167-174

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

Yeshurun, Y. & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research*, *39*, 293-306.

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory and Cognition*, *27*, 1073-1079.

Figure Captions

Figure 1: A difficult spatial relationship search task. Find the target pair with the gray 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. 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 (gray 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).