Cognition 122 (2012) 210–227

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Flexible visual processing of spatial relationships

Steven L. Franconeri *, Jason M. Scimeca, Jessica C. Roth, Sarah A. Helseth, Lauren E. Kahn

Department of Psychology, Northwestern University, United States

ARTICLE INFO

Article history:
Received 23 March 2010
Revised 21 October 2011
Accepted 2 November 2011
Available online 26 November 2011

Keywords:
Attention
Selection
Spatial relationships
Spatial language
Binding
Comparison

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 spatial relationships within structures with existing representations, such as faces, common objects, or prototypical scenes. But strikingly, there is little work on the perceptual mechanisms that allow us to *flexibly* represent arbitrary 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 such judgments. In the *simultaneous* class, both objects are selected 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. We demonstrate that shifts of selection do occur during spatial relationship judgments that feel simultaneous, by tracking selection with an electrophysiological correlate. We speculate that 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 or numerosity.

1. Introduction

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 existing 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 existing representations allow for fast and efficient processing of frequently encountered patterns. However, they have the disadvantage of being inflexible, responding only to particular stimuli.

When predefined representations are not available for a given pattern, a more flexible system supports 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

* Corresponding author. Address: Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208, United States. Tel.: +1 847 491 1259; fax: +1 847 491 7859.

E-mail address: franconeri@northwestern.edu (S.L. Franconeri).

0010-0277/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2011.11.002

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 strikingly little 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), how memory for these positions can become biased by the frame's structure (Holden, Curby, Newcombe, & Shipley, 2010; Lipinsky, Spencer, & Samuelson, 2009) 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 one object is, and we know where the other 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 retinae. 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 their relations simultaneously. This is unlikely, as the number of spatial relations among a set of objects expands at an increasing rate given the number of objects. For a given type of spatial relation (e.g. left/right), two objects have one relation, but a row of four objects has six relations, five objects ten relations – to skip ahead, ten objects have 45 relations. An important constraint that we will place on the flexible relation processing mechanism is that our intuitions about its scope are unreliable. Any sense of detail *must* be an illusion, and actual relational representations could be extremely impoverished. Our impression of broad access to the details of the visual world is frequently wrong in other cases, such as the resolution and color content of the visual periphery. This illusion of detail might rest on processes that seamlessly retrieve needed details 'on demand' (Noë & O'Reagan, 2000; Rensink, 2000).

Indeed, a number of studies reveal strong limits on our ability to judge spatial relationships within visual search tasks (see Fig. 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 relationship 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, 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 relationships, the visual system must *select* the relevant subset of objects for further processing, and relatively inhibit other aspects of a scene.

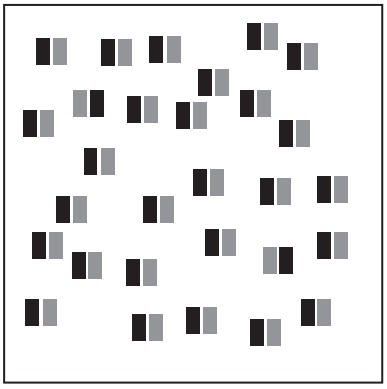

Fig. 1. A difficult spatial relationship search task. Find the target pair with the gray object on the left. Now find the second one.

But what happens under this selection? Thus far, this process represents a 'black box' (Franconeri, in press). We outline two classes of potential mechanisms that might allow the visual system to compare the relative spatial relationships between objects, based on emerging ideas from multiple laboratories. For simplicity, we will consider a single left–right judgment between just two objects. We first review a *simultaneous* class of mechanism, where both objects must be concurrently selected. We then propose a novel *sequential* class, where at least one object within the pair must be selected during the judgment.

1.1. Simultaneous selection

This class of mechanism compares the relative positions of two objects (see Fig. 2a) by treating them as a single

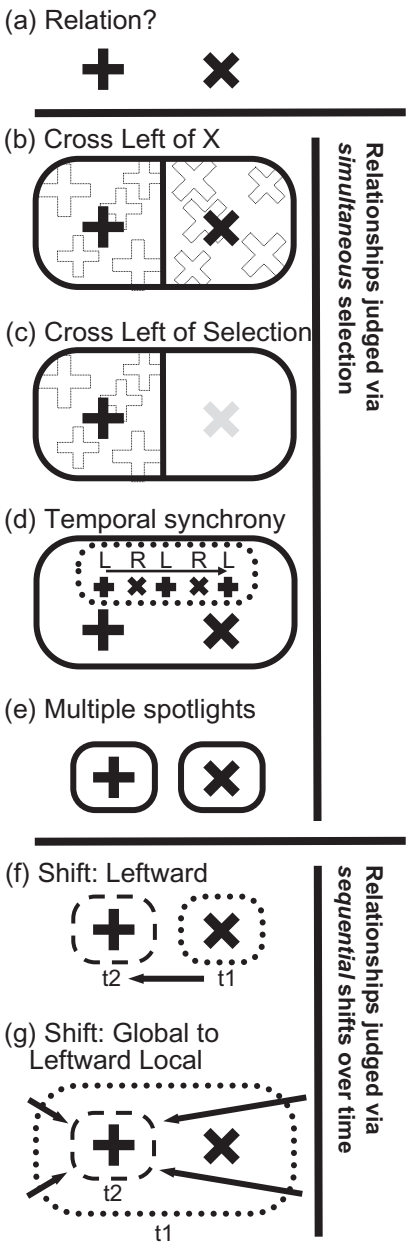

Fig. 2. Potential mechanisms for processing visual spatial relationships. (a) Two objects with an implicit relation. Examples (b)–(e): Mechanisms that require simultaneous selection. Examples (f)–(g): Mechanisms that require shifts of selection over time.

unit, 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 (Tanaka, 2003). Fig. 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, in press; 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 (see Section 9).

For more flexible relational processing, there are at least two ways to reduce this combinatorial explosion to manageable levels. Long-term detectors might detect 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 Fig. 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 simplification 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: Hayworth, Leseroart & Biederman, 2008). Fig. 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.

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 might 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, a more flexible mechanism is needed in such instances.

The mechanism depicted in Fig. 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 Fig. 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.

Another type of mechanism that may exhibit such flexibility is one that employs multiple 'spotlights' of selection simultaneously. One variant of this idea is the FINST system (Pylyshyn, 1989). According to this account, the visual system can 'index' the locations of a limited number of objects at once. Because these indexes mark objects without encoding their identities, they can be used to flexibly encode relationships among objects. Recent work suggests a concrete implementation of this idea within spatial relationship judgment tasks (Hayworth, 2009). The visual system might deploy (at least) two attentional 'spotlights' simultaneously, inhibiting each other so that they do not select the same location. A predefined network computes the relative position of spotlights 1–2, allowing recognition of the relative spatial relationship of objects 1–2, separately from processes that allow recognition of the object identities (see Fig. 2e). Under the Hayworth (2009) implementation of this account, these spotlights direct activity within functionally separated object recognition hierarchies within the ventral visual stream. In effect, this account achieves flexible spatial relationship processing by coordinating activity between *two separate visual systems*, with each processing one object within the relation. This is a bold and exciting suggestion. If true, it would require re-evaluation of substantial past work on visual cognition. But for present purposes, we simply note that to judge a spatial relation between two objects, this account also requires simultaneous selection of both objects.

2. 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. First, some past work suggests that to recover the location of an object in the visual field, we must first select that location in a focused way, perhaps due to the coarse coding of location by the ventral visual stream (Hyun, Woodman, & Luck, 2009; Luck & Ford, 1998). Second, identifying objects in many cases appears to require amplifying relevant signals from those objects, while suppressing irrelevant noise from other objects (Luck, Girelli, McDermott, & Ford, 1997; Moran & Desimone, 1985; Reynolds & Desimone, 1999; Treisman, 1996; Treisman & Gelade, 1980; but see VanRullen, 2009 for proposed exceptions).

A solution to both of these problems is to select one object at a time (Treisman & Gelade, 1980). Fig. 2f and 2g illustrate two ways to extract spatial relationships using this principle. In the example in Fig. 2f, the locus of selection could start at the right object, loading its position into a form of spatial working memory. Selection could then shift toward the left object, so that the left object's position could be encoded into a similar spatial working memory representation. Critically, this representation must be able to distinguish whether the second position is to the left of, to the right of, above, etc., the first position. This would be a simple computational problem to solve, given that this mechanism would only need to process relative positions, and could ignore the object identities (see the Section 9 for a concrete proposal for how this mechanism might work, involving encoding the *direction* of the shift). Once this mechanism produced a relative position judgment (e.g. "second selection to the left"), this relative position (left) would be co-activated with the identity of the currently selected object ("+").

The starting point for the shift might not be on one of the objects, but the center point between the objects. The shift might 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 the "x on the right". The sequence could also start at a global scope (Fig. 2g), with the locus of selection 'zooming in' toward a point that is left of the center of the original global selection window, giving an initial summary representation of the objects ("There's a + and an x, and a horizontal arrangement"), followed by the relational term ("second selection to the left") co-activated with the identity of the currently selected object ("+"). Extensions of this mechanism would also serve to process spatial relationships of objects that are not available in the same glance, either due to spatial or temporal separation. It is more difficult to imagine solutions to these problems for many of the simultaneous accounts.

Most importantly, in contrast to the simultaneous class of models, the sequential shift mechanism requires that the locus of spatial selection shift *at least once* during spatial relationship judgments. Under this mechanism, no relational information can be recovered unless this shift occurs.

3. Experiments

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 a sequential mechanism is plausible, and could underlie 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.

In Experiment 1, we ask observers to judge the left/right relation between two categorically different colors, and measure left/right selection over time with an electrophysiological correlate of selection. Even though we urged participants to complete the task by attending to *both* objects, the electrophysiological measure revealed shifts. This was also true in Experiment 1 when we added a difficult dual task intended to prevent shifts of selection. Experiments 2a and 2b use visual search experiments to verify that the colors used in Experiment 1 did not require selection in order to be identified. Instead, it is likely that the need to bind the colors to their respective locations required shifts of selection. Experiment 3 shows that when spatial relationship judgments require not only binding of objects to their relative locations, but additionally present a more challenging identification task (shapes instead of colors), these shifts are even more evident. *We argue that for many real world spatial relationship judgments, addressing known object-location binding and object identification limitations requires sequential selection of objects within a relation. For cases where binding/identification problems prohibit simultaneous selection, then the visual system must have a mechanism that allows recovery of spatial relations from this type of sequence over time.*

4. 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 (see Luck, in press, for review). This *N2pc* component, first demonstrated as negativity at 200–300 ms (N2) (though sometimes as early as 175 ms), 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 et al., 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, or even a combination of the two (Eimer, 1996; 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, Heitz, Schall, & Woodman, 2009).

The N2pc allows an experimenter to track the relative allocation of spatial attention between visual hemifields at a high temporal resolution, by comparing the relative signal strength of electrodes contralateral to one side of the visual field (recall that the right hemisphere primarily processes the left visual field, and vice versa), to those contralateral to the other side of the visual field. Subtracting these two signals reveals a difference wave that shows the relative strength of selection between the two visual fields. When using this technique to 'track' shifts of selection, it is not generally possible to tell whether participants are biased to shift toward the left or right side of space. 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 differences across the cerebral hemispheres. Instead, analyses are typically collapsed across visual hemifield, according to some other display property that predicts which side a participant will select. In Experiments 1 and 3, we use a object proximity manipulation to bias shifts toward a given side of a display. In other tasks, 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 identification priority in visual search experiments (Carrasco, Evert, Chang, & Katz, 1995; Wolfe, O'Neill, & Bennett, 1998), including one that used posterior contralateral negativity signals to track shifts of attention (Woodman & Luck, 1999, 2003). Thus, we used this proximity manipulation to predict which object (and therefore which side of the display) a participant would select. While this manipulation should bias any shifts toward one side of the display, remember that the simultaneous class predicts that selection must encompass both object within a display. All analyses average across visual hemifield (and cerebral hemisphere) by presenting object types equally often on either side of the display, and collapsing results across electrodes contralateral and ipsilateral to a given object proximity type (e.g. the near object).

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 similar to one used by Woodman and Luck (2003). By including *two* sets of objects (see Fig. 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 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, in Experiment 3 the farther object is slightly larger, scaled according to the cortical magnification factor (see Woodman & Luck, 2003). According to the sequential account, we predict that during the spatial relationship judgment, the locus of spatial attention will shift to

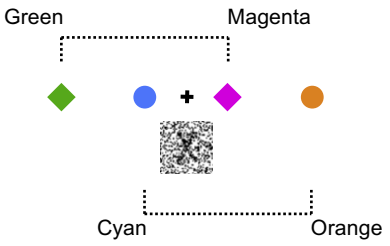

Fig. 3. Sample display for Experiment 1. Participants reported the relative relationships between the colors of the relevant shape set (either diamonds or circles). In the dual-task condition, they additionally reported whether the noise-degraded letter was a vowel or consonant.

the near 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.

5. Experiment 1

In Experiment 1's *relation only* condition, participants judged the color relationship between two objects. Displays were balanced such that there were always two objects on each side of fixation. Participants judged the relation between the diamonds or the circles. Within the relevant set, one object was on each side of fixation, with one was closer to fixation. If judging the spatial relationship between the two colors requires selecting one or more of the objects, a shift should occur at the early post-stimulus time range (200–300 ms) toward the near object, and potentially also toward the farther object at the later post-stimulus time range (300–400 ms) (Woodman & Luck, 2003). To determine whether this shift occurs even when it is discouraged, in addition to the relation only condition, we also included a *dual task* condition where participants were additionally required to identify a noise-degraded letter at fixation (see Luck & Ford, 1998, for a similar manipulation). In the dual task condition, we predicted that the shift of attention related to the spatial relationship judgment would occur later in time. Pilot results using different dual tasks suggest that this shift, which usually occurs initially at 200–300 ms, will occur later (approximately 400 ms post-stimulus). Despite the difference in task requirements, the single- and dual-task conditions used identical displays, and participants simply ignored the noise-degraded letter in the relation-only condition.

5.1. Methods

5.1.1. Participants

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

5.1.2. 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° × 24.4° at an approximate viewing distance of 56 cm, with a 1024 × 768 pixel resolution, 33.6 pixels per degree. Stimuli are depicted in Fig. 3. Fixation displays had a grey background (13 cd/m^2) with a small light grey fixation cross (30 cd/m^2) 0.36° wide. The fixation remained on screen throughout the trial. The stimulus displays contained four objects and a noise-degraded letter. The objects were colored shapes, either diamonds (0.71° wide) or circles (0.60° 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). The color values were approximately perceptually equiluminant, as determined by a separate experiment where eight observers were asked to minimize perceived flicker as a red[1] and green square alternated at 15 Hz. 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. There was also always a white square 1.49° wide, centered 1.49° below fixation, containing a black letter (A, E, O, U; H, S, X, N) in Helvetica font approximately 0.83° 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 cyan and orange. Displays were fully counterbalanced such that each shape and color appeared at each of the four screen locations equally often.

5.1.3. 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).

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 1500 ms, depicting examples of the currently relevant objects in both possible arrangements (e.g., a cyan circle to the left of an orange circle, and an orange circle to the left of a cyan circle underneath). The instruction screen also specified whether the center letter was relevant, by asking the participant to either ignore 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)

[1] For interpretation of color in Figs. 3, 6 and 7, the reader is referred to the web version of this article.

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 800–1200 ms (rectangular distribution), and were followed by the stimulus display for 120 ms, another fixation display for 680 ms. A response prompt screen then appeared, 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).

Eye movements were monitored by a table-mounted SR-Research Eyelink 1000 Remote eyetracker. If participants moved their eyes outside of a 1° radius around the fixation point, from the time window starting from 800–1200 ms preceding stimulus presentation (depending on the randomly chosen inter-trial jitter value) to 800 ms after stimulus presentation, the trial was rejected. Given the small amount of noise present in the eyetracker's position signal (approximately 0.5°), the effective size of the allowed window was actually smaller than the permitted 1° 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 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 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 five trials of each of the eight 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.

5.1.4. EEG recording

ERP was recorded using a Biosemi Active 2 EEG/ERP system. The DC recording was made at 512 Hz with a hardware low-pass filter, and then was decimated in software to 256 Hz. All sites were re-referenced to the post-recording average of the left and right mastoids and low-pass filtered at 80 Hz. 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 200 ms before to 800 ms after stimulus presentation. Participation in the experiment took 2 h, including ERP cap preparation, breaks, and task practice. Inter-trial delays include randomized timing with at least 400 ms of jitter (rectangular distribution) to minimize the impact of previous trials on the EEG signal.

5.2. Results and 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. Every remaining 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).

Fig. 4a depicts activity at PO7/8 for electrodes contralateral to the near and far targets in the single task condition. We predicted that activity would be more negative to the near object between 200–300 ms, and perhaps more negative for the far object between 300–400 ms, suggesting a shift of attention from the near object to the far object over time. The earlier pattern emerged (Difference $M = 0.33$uv), $t(9) = 3.0$, $p = 0.014$, but there was no difference at the 300–400 ms time window. Although not predicted a priori, there was a trend from 500–600 ms for a return to more negativity for electrodes contralateral to the near object (Difference $M = 0.35$uv), $t(9) = 2.0$, $p = 0.08$. 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. Fig. 4b depicts this pattern as a difference wave, expressed as more negativity for electrodes contralateral to the near vs. far object.

Fig. 5 depicts activity at PO7/8 for electrodes contralateral to the near and far targets in the dual task condition. Pilot experiments using a different dual-task manipulation suggested that this shift would occur later in time (approximately 400 ms), 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 contralateral to the *near* object at a late time window, there was more negativity in electrodes contralateral to the *far* object between 400 and 500 ms (Difference $M = 0.5$uv), $t(9) = 2.6$, $p = .026$. Although not predicted a priori, this negativity continued into the 500–600 ms time window ($M = 0.31$uv), $t(9) = 2.3$, $p = .047$. This pattern is consistent with a shift of attention toward the relevant object *farther* from fixation at a later time window, presumably after discrimination of the letter at fixation.

These results demonstrate that participants shifted attention in systematic ways during spatial relationship

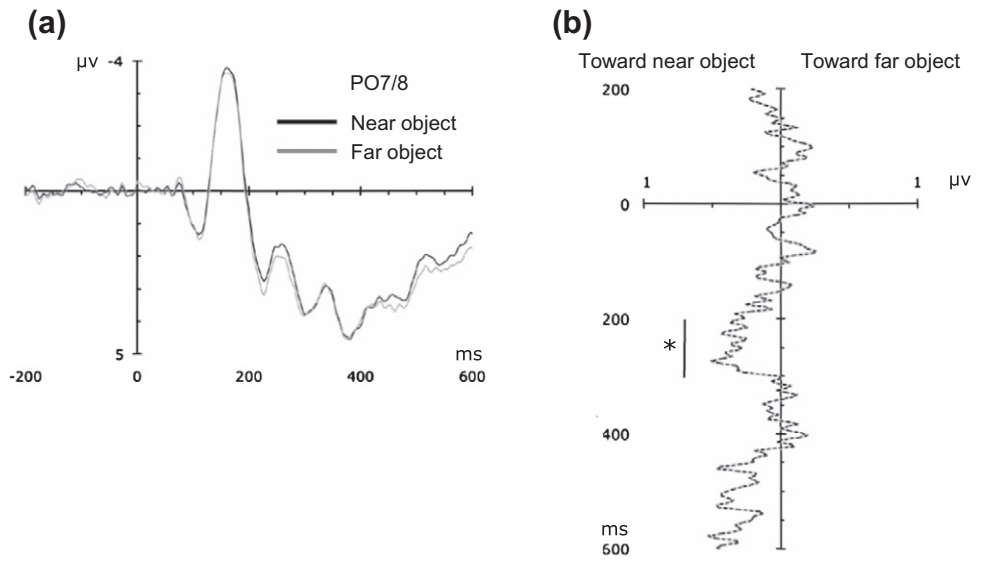

Fig. 4. For the *single-task* condition of Experiment 1. (a) Average ERPs from PO7/8 for electrodes contralateral to the closer object of relevant color (dark line) or the electrodes contralateral to the farther object (grey line). More negative values (plotted upward) indicate shifts of attention toward that object. (b) Difference waves between the lines, indicating shifts toward the near object (leftward deviation), or far object (rightward deviation).

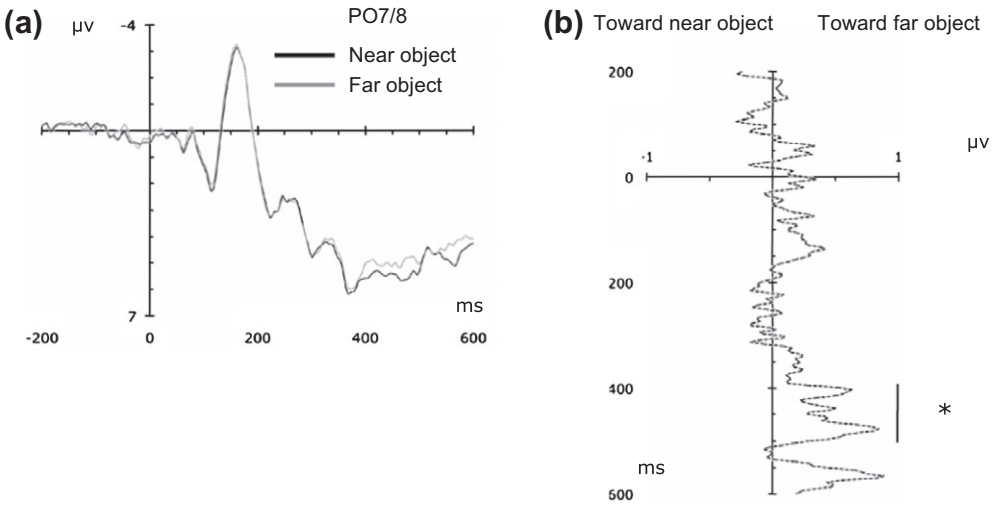

Fig. 5. (a) Identical analysis as shown in Fig. 4, but for the *dual-task* condition of Experiment 1, including (b) a difference wave.

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 Experiments 2a and 2b 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 objects.

There are two other intriguing aspects of the results of Experiment 1. First, both the single- and dual-task conditions show only one shift toward one object, instead of shifts toward both the near and far objects. This pattern of results is consistent with our claim that such shifts are required for spatial relationship processing, and that shifting to both objects should not be necessary. For example, to know that magenta is to the left of green, it is sufficient to know that magenta is on the left. Thus, the important aspect of the results is not whether there were one or two shifts, but that there were any shifts at all. The single-shift pattern could also be taken as support for the 'global to local' shift mechanism

described in Fig. 2g, which explicitly predicts a single shift after global selection of both objects.

The second intriguing finding is that 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 (e.g., Bahcall & Kowler, 1999; Cave & Zimmerman, 1997; Hopf et al., 2006), making the far object a more attractive target. Another intriguing possibility is that, if selection began at the degraded letter and then shifted to the near object, then because the letter was placed below the fixation point, the position difference between these points would create a diagonal, a poor exemplar of a horizontal relationship (Logan & Sadler, 1996). In contrast, the position difference between the letter and the far object would have a much stronger horizontal component.

6. Experiment 2a

Based on previous work using visual search paradigms (e.g., Treisman & Gelade, 1980), identifying the colors used in Experiment 1 should not have required that they be individually selected. However, because the set of colors was slightly more heterogeneous than in many past experiments, we conducted a visual search task to ensure that a singleton color could be efficiently identified in this type of visual search display. If adding additional distractors to the display does not substantially increase response times, then the color identification requirements of Experiment 2 should not require focused attention.

6.1. Methods

6.1.1. Participants

Ten Northwestern University undergraduates participated in a 30-min session in exchange for payment or course credit.

6.1.2. Stimuli

Stimuli were similar to those used in Experiment 1, except that up to 3, 6, or 12 colored shapes were distributed across the display (see Fig. 6a). 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 12-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 1. Inner shapes were placed 45° off the display's vertical or horizontal axes (22.5° for outer objects).

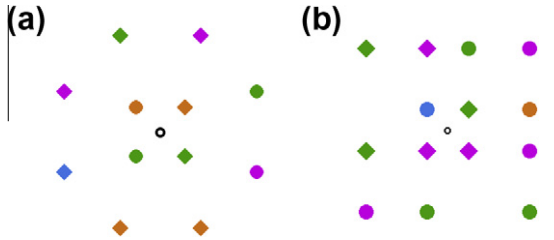

Fig. 6. (a) Sample display for Experiment 2a. Participants reported whether the target shape (in this case, cyan) was a diamond or circle. (b) Sample display for Experiment 2b. Participants reported whether the target shapes (in this case, cyan/orange) were arranged horizontally or vertically. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.1.3. 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 1000 ms 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-s delay.

6.2. Results and discussion

Accuracy was 97% at each set size. Mean response times were 597, 625, and 643 ms for the 3, 6, and 12 shape displays, respectively, showing a positive ($t(9) = 3.6$, $p = .006$) but effectively flat slope (5.1 ms/item). Slopes were 11 ms/item or less for every participant.

Even though the displays in Experiment 2a were more populated and more crowded than those used in the spatial relationship judgment task in Experiment 1, there was virtually no cost in identifying a target color, suggesting that identifying the colors used in Experiment 1 did not require that they be serially selected.

7. Experiment 2b

Experiment 2b was conducted to address two potential critiques of Experiment 2a. First, while Experiment 2a required observers to identify one target color, Experiment 1 required observers to find a relationship between two target colors. Second, while shape was the reported feature in Experiment 2a, in Experiment 1 shape was a property that signaled the locations of the target objects, serving as a redundant cue beyond color for target identity. Combining shape and color in this way might remind some readers of a 'conjunction' visual search task, where observers are asked to find a target defined by two properties (e.g.

a red circle among red diamonds and green circles). Because such searches lead to inefficient visual search in many cases (Treisman & Gelade, 1980; Wolfe, 1994), at first glance the shifts may seem to be due to the need to 'resolve' these conjunctions. But, importantly, in Experiment 1, color was perfectly diagnostic of the locations of the target objects. As an analogy to the conjunction search example, there were no red diamond distractors, and red alone was sufficient to identify the target. Instead, the shape differences existed to maintain a conservative test of the shift hypothesis. By instructing the subject to focus attention on *both* objects with a given shape (e.g. circles), the shared property encourages observers to simultaneously select both objects, giving the best chance of disproving the shift hypothesis.

Experiment 2b presents a new search task that addresses each of these potential concerns. Participants were required to locate *both* (e.g.) cyan and orange circles within a display of green and magenta diamonds and circles, and then decide whether the two targets were arranged vertically or horizontally. This task retained the requirements to simultaneously identify two colors, as well as the characteristic of having those two targets redundantly signaled by a set shape (e.g. circles).

7.1. Methods

7.1.1. Participants

Eight Northwestern University undergraduates participated in a 30-min session in exchange for payment or course credit.

7.1.2. Stimuli

Stimuli were similar to those used in Experiment 2a, but used 6, 10, or 14 colored shapes distributed across positions within an invisible 4×4 grid. Each object's horizontal and vertical eccentricity was scaled in the same manner as in Experiment 3a, such that objects near fixation along either axis were pushed closer to fixation along that axis (see Fig. 6b). The target color pair and shape were randomly chosen from those used in Experiment 1, and distractors were chosen from the remaining colors, with an equal number of objects of each color present in the display. Half of the distractors were diamonds and half were circles.

7.1.3. Procedure

There were 360 trials, divided by target colors and shape (e.g., aqua and orange circles) into 90 trial blocks presented in random order for each participant. Each sequence began with an instruction screen depicting the target colors and shape, followed by 30 trials of each set size, in random order. In each trial there was a 800 ms fixation display, followed by the search display until response. Participants used a keypress response to report the arrangement of the target objects. Incorrect responses were followed by an 'incorrect' message and a 5-s delay.

7.2. Results and discussion

Accuracy was high at all set sizes (M = 95%, 96%, and 96% for set sizes 6, 10 and 14, respectively). Response times actually decreased slightly (M = 688, 679, and 686 ms), but slopes were not significantly different from zero ($t(7) = 0.46$, $p = 0.66$).

The results of Experiment 2b confirm that when a spatial relationship judgment is not required, shifts of selection are also not required to identify two target colors and their arrangement, in a display containing heterogeneous shapes. Instead, the shifts are more likely due to a need to select each object in order to bind its identity to a location within the relation. One caveat to this conclusion could be that while Experiment 1 relied on electrophysiological measures of attention shifts, Experiments 2a and 2b rely on behavioral measures, and care should be taken in assuming that they measure the same underlying mechanisms. We are more confident about our conclusions because of past electrophysiological results (using the same component) confirming that object identification tasks do not necessarily lead to shifts of attention to the identified object (Luck & Ford, 1998).

8. Experiment 3

We argue that spatial relationship judgments should require shifts of selection between objects in order to bind object identities to locations, and identify each object. Experiment 1 tests the binding prediction by using categorically different colors that do not require selection in order to be identified. But real-world relation judgments require identification of more complex objects. Experiment 3 therefore tests binding as well as identification, by using slightly more complex objects (shapes) that present a more challenging identification task. The results show that in this more realistic task, the evidence for shifts is even stronger.

8.1. Methods

8.1.1. Participants

Fifteen Northwestern University undergraduates participated in exchange for payment or course credit.

8.1.2. Stimuli

Stimuli were similar to Experiment 1, with the following changes. 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°

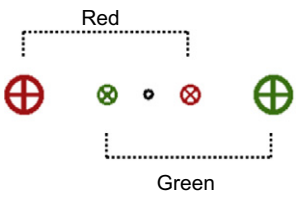

Fig. 7. Sample stimulus for experiment. Participants were instructed to report the spatial relationship between shapes of the relevant color while ignoring shapes of the other color.

thick segments, and the near shapes were 1.19° from the fixation point, 0.60° in diameter, and had 2 pixel (0.06°) 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 Fig. 7). The color values were approximately perceptually equiluminant, as described in Experiment 1.

8.1.3. Procedure

Similar to Experiment 1, with the following differences. Eye movements were not monitored with an eyetracker, and trials with eye movements were instead eliminated at the analysis stage by inspection of EOG channels. At the beginning of each trial a fixation point was displayed for 1800–2200 ms, followed by the stimulus display for 1500 ms. The fixation remained on screen throughout the trial. 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 × 2 green shape orderings × 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.

8.2. Results and 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 nine 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 (at most a small fraction of a degree; Hillyard & Galambos, 1970). Trials with incorrect responses or responses of over 1500 ms were also removed from the analysis. Accuracy was high (*M* = 96.6%, SD = 3.4%). Response time was 741 ms on average (SD = 82 ms).

We predicted a priori that activity would be more negative contralateral to the near shape between 200–300 ms post-stimulus, and more negative contralateral to the far shape after 300–400 ms post-stimulus (Woodman & Luck, 2003). The results confirm this prediction. Fig. 8a depicts waveforms for electrodes contralateral to the near and far shapes, and Fig. 8b depicts the difference between these two expressed as signals consistent with attentional shifts toward either shape. At earlier times, 200–300 ms post-stimulus, PO7/8 amplitudes were more negative contralateral to the near target compared to the far target (Difference $M = 0.78$uv), $t(8) = 4.2$, $p = 0.003$. At later times 300–400 ms post-stimulus, the reverse pattern appeared where amplitudes were more negative contralateral to the far target (Difference $M = 0.82$uv), $t(8) = 4.4$, $p = 0.002$. This pattern of activity supports our prediction that participants would first shift to the near object and then shift to the far object.

We note that the activity seen at 300–400 ms 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. However, the Ptc is typically substantially smaller than the initial

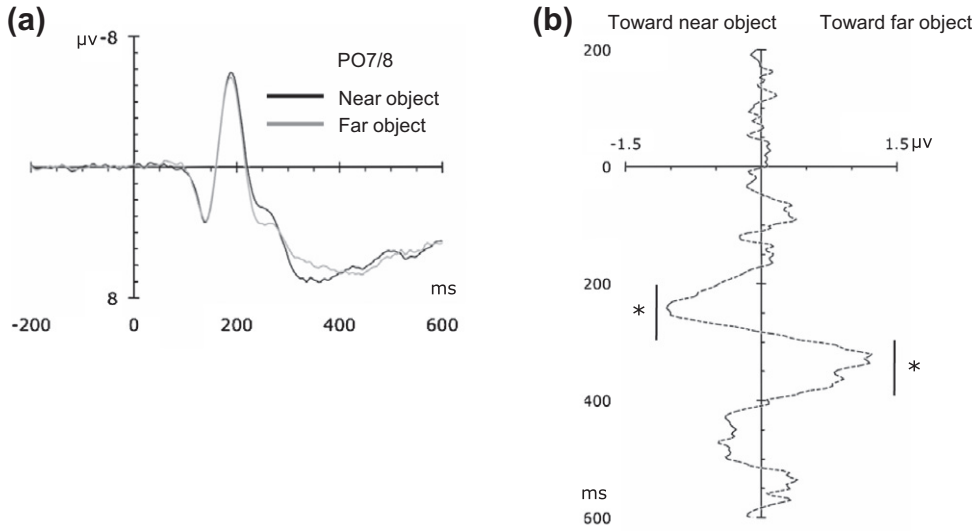

Fig. 8. (a) Average ERPs from PO7/8 electrodes contralateral to the closer object of relevant color (dark line) or contralateral to the farther object (gray line). More negative values (plotted upward) indicate shifts of attention toward that object. (b) Difference waves between the lines in Fig. 4a,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

n2pc (e.g., Luck, Fan, & Hillyard, 1993; Luck et al., 1997), while true second shifts are at least as large (Woodman & Luck, 1999, 2003). In the present data, the second shift is at least as large as the first. But critically, even if the second 'shift' did not occur, we could still conclude that there was at least one shift.

One potential criticism of Experiments 1 and 3 is that the interleaved arrangement of the objects made simultaneous selection more difficult. But the visual world typically requires exactly this type of 'interleaved' judgments, 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 1° 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 Fig. 2g 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 positions of those objects.

A final critique is that the present experiments can only show that shifts occur, but these results cannot completely disambiguate the computational role that they serve. We claim that these shifts are needed to (a) localize/identify/bind one of the objects (e.g. Exp 1) or both of the objects (Experiment 3), and then (b) compute the relative location of the objects (by examining the direction of the attentional shift). However, it is possible that the shifts seen in our results are only needed for (a) and not (b). Particularly for Experiment 3, where there were shifts to both objects, it is possible that these shifts were needed to load two objects into a working memory store (requiring that they first be localized/identified/bound), but then some other unknown process computed the relation in a 'simultaneous' way. While this explanation is logically possible and the present results cannot exclude it, we prefer the attention-shift explanation. The 'simultaneous relation extraction within working memory' solves the problem of flexible relation processing by invoking a homunculus – it does not specify *how* the relation is extracted. Given the visual processing constraints we reviewed in the introduction, we do not see how this mechanism *could* operate without specifying a mechanism similar to the one suggested here (extracting the direction of attention shift), but at a higher level of processing (e.g., shifting the equivalent of 'attention' within the memory representation).

An important demonstration for future research will be to show 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, such demonstrations are 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. According to the shift account, if the relationship can be reported, then the relationship was recognized, and selection must have shifted among the judged objects. Demonstrating the need for shifts to construct spatial relationships therefore requires 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). Our laboratory has completed initial ERP studies similar to Experiment 1 suggesting that, under dual task conditions, spatial relationship judgments are associated with shifts of selection, but these shifts disappear during same-different color judgment trials where participants report an illusion of 'unbound' colors.

9. 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 requires simultaneous selection in order to make relation judgments. But because simultaneous selection is known to bring processing difficulties associated with both object identification and binding of those identities to specific locations, we argue for the existence of a novel class of relation judgment mechanism where selection can shift among the judged objects over time. We offer electrophysiological evidence that shifts do occur in a simple judgment that gives the impression of simultaneous selection. Experiment 1 showed evidence of shifts in a simple color relation judgment task, even when a dual task manipulation was added to discourage shifts. Experiments 2a and 2b verified that the colors used did not require selection to be identified. Instead, selection likely shifted in order to bind these colors to their locations (Hyun et al., 2009; Luck & Ford, 1998). Experiment 3 demonstrated that when object identification was made slightly more difficult (shape judgment instead of color), simulating a more realistic judgment, the pattern of shifts was even more salient.

Another recent study uses behavioral techniques to suggest that selection shifts between objects during spatial relationship judgments (Holcombe, Linares, & Vaziri-Pashkam, 2011). The task consisted of determining spatial relationships between pairs of colored circles contained within two concentric rings. For example, a participant might report that the red object (on the inner ring) was 'inside' the green object (on the outer ring). This decision is trivially easy, until the experimenters begin to spin concentric ring stimulus at increasing rates. The first result of the study is that detecting color identities can be performed at fast speeds, but judging relations between those colors requires far slower speeds. Similar in spirit to the visual search studies described in the introduction (e.g. Logan, 1994, 1995), this result suggests a capacity-limited resource for spatial relationship judgment.

The second result supports the idea that this limited resource is a mechanism that we would classify as 'sequential'. When participants make errors in their relational judgments, they typically choose relations between an inner object and an outer object that 'trails' that correct outer object in time. That is, instead of choosing the object directly outside of the red, they choose the object that appears outside the red at a later time. This suggests that participants start their shift sequence at one of the inner colors (as in our proximity manipulations), and as they shift toward the outer object over time, they often choose an outer object that appears later.

Such results are inconsistent with accounts predicting that selection *must* simultaneously envelop both objects in order to judge the spatial relationship between them. Even if the observed shifts are due to the need to localize and/or identify each individual object, then this effect presents a strong challenge to any model of spatial relationship processing that relies on simultaneous selection. Instead, the visual system must have access to a mechanism that allows *reconstruction of relations in space from this dynamic process that unfolds over time*.

How might this reconstruction process work? In the introduction, we described this process at an abstract level, noting that the first selection location could be compared to the second location. To demonstrate that this comparison could be computationally relatively simple, we offer a one account for how this might be implemented. A solution would be to *record the direction of the attentional shift* from one object to the other. This shift direction could be briefly held in heightened activation (see Fig. 2f), by a circuit similar to a detector for low-level motion (Reichardt, 1969). Note that this mechanism could still detect 'motion' of selection regardless of whether the shift is analogue or discrete. Arrays of these detectors could be placed in parallel 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 after they occur, the visual system could also compute a shift direction using an 'efference copy' of the shift command itself.

It would also be computationally simple to employ a similar set of detectors (or efference copy signal) for the global-to-local 'zoom' system depicted in Fig. 2g. We feel that this mechanism is more likely than the object-to-object shift account. It is consistent with a bias toward initial global processing (Navon, 1977), and even if such simultaneous selection did not provide relations between objects, it could still provide a summary of the distribution of features present (Treisman & Schmidt, 1982). It would explain the feeling that we apprehend spatial relationships via simultaneous selection, because that would be true for the initial stage of processing. It also provides a way to perceive relations within more complex sets of more than two objects. After initially selecting the whole group, there could be parallel representation of all individual object locations within that group (Franconeri, Bemis, & Alvarez, 2009). Then the 'zoom' operation could provide the identity of any given object within the group's reference frame.

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, Girelli, McDermott, & Ford, 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; see Levinthal & Franconeri, 2011, and Xu & Franconeri, in press, for new examples of attention serving similar active roles for visual grouping and within-object structure assignment). 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. More broadly, applications of similar shift accounts to other visual tasks might serve to create combinatorially expressive representations (Fodor & Pylyshyn, 1988).

9.1. Relation to other models of spatial relationship judgment

This sequential shift mechanism presents specific ways to implement stages of Logan and 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 and 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 and Carlson (2001).

There is also related work on the origins of the 'Simon Effect' suggesting a role for attention shifts in spatial relationship judgment. In the Simon Effect, responses are faster to a stimulus when the stimulus and response share the same relative spatial position (e.g. the left response button for the left object). Some accounts of this effect suggest that these spatial positions are coded not by location of each object within a reference frame, but by the relative direction of shifts of attention within a display (Stoffer, 1991). Support for this account typically comes from experiments that manipulate shifts of attention with various types of cues, while keeping the locations of the stimulus objects constant. Many of these studies find results suggesting that the directions of these shifts mediate the Simon Effect, instead of the object locations per se (Abrahamse & Van der Lubbe, 2008; Nicoletti & Umilta, 1994; Proctor & Lu, 1994; Rubichi, Nicoletti, Iani, & Umilta, 1997; Stoffer & Umilta, 1997).

9.2. '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).

9.3. Categorical vs. coordinate spatial relationship judgments

The type of spatial relationship judgment that we consider here, where object 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). The control task used in Experiment 2b should be considered a 'coordinate' decision, because participants were asked to judge the group orientation of a pair of objects that were featurally identical. 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), line bisection tasks, where observers are asked to mark the precise midpoint of a line (Jewell & McCourt, 2000; McCourt & Jewell, 1999), or displacement detection tasks, where observers must decide if a pair of points have increased in separation (Palmer, 1986a,b). 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. The object-to-object shift account (Fig. 2f) predicts this result, 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.

9.4. 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). 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 also 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 were 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 the shift direction itself after it has been made over a retinotopic representation. The latter option may be more computationally efficient.

9.5. Connections between visual space and spatial language

There are strong similarities between visuospatial representations of relations and linguistic descriptions of relations (Carlson & Logan, 2005; Logan, 1995; Logan & Sadler, 1996). For example, there are similar semantic, action, and other experience-based properties that help determine the choice of reference frames (Cuijpers, Kappers, & Koenderink, 2001; Mou & McNamara, 2002; Taylor & Tversky, 1992, 1996). In addition, the spatial layouts that are considered 'acceptable' for a given relationship (e.g., an object directly above another is a better example of 'above' than an object to the upper left of another) are also 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 shift account proposes that the perception of a relationship between two objects requires the sequential selection of at least one object, 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, 2007; Altmann & Kamide, 2009), as well as the production of descriptions of those scenes (Gleitman, January, Nappa, & Trueswell, 2007; 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 two values 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 values 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 values 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.

9.6. 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 dynamic shifts of selection. This flexible mechanism would complement other long-term representations of visual structure, and language might help construct compositional representations of arbitrary object arrangements.

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, Platt, & Brannon, 2009; Kadosh, Lammertyn, & Izard, 2008; Pinel, Piazza, Le Bihan, & Dehaene, 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; Halford, Wilson, & Phillips, 2010; Hummel & Holyoak, 2003), 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, Alex Holcombe, Todd Horowitz, 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, and to Todd Handy for providing his laboratory for early pilots of this work. This work was supported by NSF SLC Grant SBE-0541957, the Spatial Intelligence and Learning Center (SILC), and an NSF CAREER Grant (S.F.) BCS-1056730.

References

Abrahamse, E. L., & Van der Lubbe, R. H. J. (2008). Endogenous orienting modulates the Simon effect: Critical factors in experimental design. *Psychological Research, 72*, 261–272.

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 Cognitive 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.

Carrasco, M., Evert, D. L., Chang, 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.

Cave, K. R., & Zimmerman, J. M. (1997). Flexibility in spatial attention before and after practice. *Psychological Science, 8*, 399–403.

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 maintaining feature conjunctions. *Psychological Science, 19*(2), 189–195.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology, 99*(3), 225–234.

Enns, J. T., & 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.

Fodor, J., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture. *Cognition, 28*, 3–71.

Franconeri, S. L. (in press). In D. Resiberg (Ed.). *The nature and status of visual resources*. Oxford University Press.

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition, 113*, 1–13.

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.

Gleitman, L., January, D., Nappa, R., & Trueswell, J. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language, 57*(4), 544–569.
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.
Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences, 14*(11), 497–505.
Hayward, W. (2003). After the viewpoint debate: Where next in object recognition? *Trends in Cognitive Sciences, 7*(10), 425–427.
Hayward, W. G., & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition, 55*, 39–84.
Hayworth, K. J. (2009). PhD thesis. *Explicit encoding of spatial relations in the human visual system: Evidence from functional neuroimaging*.
Hayworth, K., Lescroart, M., & Biederman, I. (2008). Explicit relation coding in the Lateral Occipital Complex [Abstract]. *Journal of Vision, 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. (2009). Electrophysiological indices of target and distractor processing in visual search. *Journal of Cognitive Neuroscience, 21*, 760–775.
Hilimire, M. R., Mounts, J. R. W., Parks, N. A., & Corballis, P. M. (2009). Competitive interaction degrades target selection: An ERP study. *Psychophysiology, 46*, 1080–1089.
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.
Holcombe, A. O., Linares, D., & Vaziri-Pashkam, M. (2011). Perceiving spatial relations via attentional tracking and shifting. *Current Biology, 21*(13), 1135–1139.
Holden, M., Curby, K., Newcombe, N. S., & Shipley, T. F. (2010). A category adjustment approach to memory for spatial location in natural scenes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36*, 590–604.
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., et al. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex, 10*(12), 1233–1241.
Huang, L., & Pashler, H. (2005). Attention capacity and task difficulty in visual search. *Cognition, 94*, B101–B111.
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. (in press). Object recognition. In D. Reisberg (Ed.) *Oxford handbook of cognitive psychology*. Oxford, England: Oxford University Press.
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. (2000). Pseudoneglect: A review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia, 38*, 93–110.
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., & Michimata, C. (2010). Processing spatial relations with different apertures of attention. *Cognitive Science, 35*(2), 297–329.
Levinthal, B., & Franconeri, S. L. (2011). Common fate grouping as feature selection. *Psychological Science, 22*(9), 1132–1137.
Lipinsky, J., Spencer, J. P., & Samuelson, L. K. (2009). Corresponding delay-dependent biases in spatial language and spatial memory. *Psychological Research, 74*(3), 337–351.
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. (in press). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related components. In S. J. Luck, & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components*. New York: Oxford University Press.
Luck, S. J., Fan, S., & Hillyard, S. A. (1993). Attention-related modulation of sensory-evoked brain activity in a visual search task. *Journal of Cognitive Neuroscience, 5*, 188–195.
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.
Milner (1974). A model for visual shape recognition. *Psychological Review, 81*, 521–535.
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.
Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782–784.
Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 162–170.
Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353–383.
Nicoletti, R., & Umilta, C. (1994). Attention shifts produce spatial stimulus codes. *Psychological Research, 56*, 144–150.
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. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*(12), 520–527.
Palmer, J. (1986a). Mechanisms of displacement discrimination with and without perceived movement. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 411–421.
Palmer, J. (1986b). Mechanisms of displacement discrimination with a visual reference. *Vision Research, 12*, 1939–1947.
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.
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.
Proctor, R. W., & Lu, C.-H. (1994). Referential coding and attention-shifting accounts of the Simon effect. *Psychological Research, 56*, 185–195.
Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32*, 65–97.
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), 191–201.
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), 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.
Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron, 24*(1), 19–29.
Rieser, J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*(6), 1157–1165.
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.
Rubichi, S., Nicoletti, R., Iani, C., & Umilta, C. (1997). The Simon effect occurs in relation to the direction of an attention shift. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 1353–1364.
Ryan, J. D., & Villate, C. (2009). Building visual representations: The binding of relative spatial relationships across time. *Visual Cognition, 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*, 735–774.
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.
Stoffer, T. H. (1991). Attentional focusing and spatial stimulus–response compatibility. *Psychological Research, 53*, 127–135.
Stoffer, T. H., & Umilta, C. (1997). Spatial coding with reference to the focus of attention in S–R compatibility and the Simon effect. In B. Hommel & W. Prinz (Eds.), *Theoretical issues in S–R compatibility* (pp. 181–208). Amsterdam: North-Holland.
Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex, 13*, 90–99.
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.
Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language, 31*, 261–282.
Taylor, H. A., & 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.
Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202–238.
Wolfe, J. M. (1998). What 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.
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.
Xu, L. & Franconeri, S. L. (in press). The head of the table: The location of the spotlight of attention may determine the 'front' of ambiguous objects. *Journal of Neuroscience*.
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.