What vs Where: Which Direction Is Faster?

Hongbin Wang (Hongbin.Wang@uth.tmc.edu)
Todd R. Johnson (Todd.R.Johnson@uth.tmc.edu)
Ruijun Bao (Ruijun.Bao@uth.tmc.edu)
School of Health Information Sciences, University of Texas at Houston
7000 Fannin Suite 600, Houston, TX 77030 USA

Abstract

It has been well documented that where a visual stimulus is located and what it is are represented and processed through different neural pathways. This paper reports on an experiment that investigated how the where pathway and the what pathway interact by evaluating and comparing the relative efficiency of retrieval in two directions: from what to where and from where to what. Our results show that retrieving from what to where is faster than retrieving from where to what, quite contradictory to previous results. The implications of our findings are discussed.

Introduction

A large body of evidence in visual perception and attention has shown that different dimensions of a visual stimulus are processed in parallel by different specialized neural systems, especially in early vision (see Farah, 2000). The well documented distinction between the what and where pathways reflects this general principle of modularity in the brain's information processing. Specifically, Ungerleider & Mishkin (1982) suggested that there were two cortical visual systems in the brain, with a ventral pathway through inferior temporal cortex processing information about features that are critical for object recognition, such as shape and color, and a dorsal pathway through posterior parietal cortex processing information about object location and spatial relations among objects. This distinction has later been summarized as "what" versus "where", respectively.

Despite the enormous evidence supporting the segregation of what and where processing, the implications of such segregation on perception, attention, and working memory have been actively debated (see Farah, 2000). One essential issue is how the two pathways interact. For example, while it has been suggested that what and where information is integrated in prefrontal cortex (e.g., Rao, Rainer, & Miller, 1997), Ungerleider, Courtney, & Haxby (1998) show that the segregation of ventral what and dorsal where processing extends from visual cortex to prefrontal cortex, forming a distributed neural system for visual working memory. In the attention literature, although the role of location (where information) in shifting attention and binding other visual features (what information) for object identification has been generally emphasized (e.g., Treisman & Gelade, 1980; Posner, 1980; Lamy & Tsal, 2001), various forms of object-based attention have also been advocated (e.g., Pylyshyn, 2001; Scholl, 2001).

Nissen (1985) reported a study investigating how spatial information processing interacts with visual feature processing. In her Experiment 2, four colored shapes (e.g., blue triangle, red circle, black square, and green diamond) were briefly presented, each at a unique position relative to a center fixation (e.g., top, right, bottom, left, respectively). Subjects were then asked to perform a partial report task in two conditions. In the location-cue condition, subjects were presented a location word (e.g., "top") and asked to report the color and shape of the object that appeared at that location (i.e., "blue" & "triangle"). In the color-cue condition, subjects were presented a color word (e.g., "red") and asked to report the location and shape of the object with that color (i.e., "red" & "circle"). Nissen found that when the cue was a location, correct recall of color and shape were statistically independent; however, when the cue was a color, correct recall of shape depended on correct recall of location.

Based on these results, Nissen suggested that spatial locations played a unique and special role in visual selective attention – it is location that mediates visual feature integration and retrieval but not the other way around. In particular, she suggested that there existed multiple maps, each representing a different visual feature. A color map registered the spatial layout of presented colors, and a shape map registered the spatial layout of presented shapes. Since these maps were co-registered relative to spatial locations, locations became special in that they allowed cross-reference between maps. Therefore, retrieving the shape (or color) of an object given its location as a cue could be done using the single shape map (or color map), resulting in a statistical independence in accuracy. On the contrary, retrieving the shape of an object given its color as a cue required access to two maps – one had to first use the color map to retrieve the location containing an object with that color, followed by using the shape map to retrieve the shape at that location. The crucial mediation role of location in cross-referencing maps led to the statistical dependence in performance.

Though Nissen's analysis has been questioned (e.g., Monheit & Johnston, 1994; van der Velde & van der Heijden, 1997), her claim that spatial location plays a particularly important role in visual perception and selective attention has generally been supported (Isenberg, Nissen, & Marchak, 1990; Tsal & Lavie, 1993; Tsal & Lamy, 2000). Based on Nissen's results, a representational scheme that emphasizes spatial location's function in bridging and binding other visual features was proposed (see Figure 1) and a computational model was developed using the ACT-R cognitive architecture (Johnson, Wang, Zhang, & Wang, 2002). The modeling results matched Nissen's experimental results remarkably well.

This type of location-indexed multi-map theory of visual perception has interesting implications on the nature of interaction between what and where pathways. On the one hand, it suggests that while each visual feature of a multidimensional visual stimulus is processed and represented separately, each feature representation (what information) is fundamentally intermingled with the corresponding spatial location (where information) in a form of map that directly links visual features and their locations (see Figure 1). This is inconsistent with the general principle of what and where segregation. On the other hand, if one indeed possesses these types of maps and can use them for retrieval, we would expect that in a single map situation retrieving a visual feature (what) from a location (where) is no different from retrieving a location (where) from a visual feature (what). This is just what Nissen suspected. She predicted that when "subjects were cued with a color and reported the location of the cued color, or they were cued with a location and reported the color at the cued location … selection by location would hold no special advantage" (p. 208).

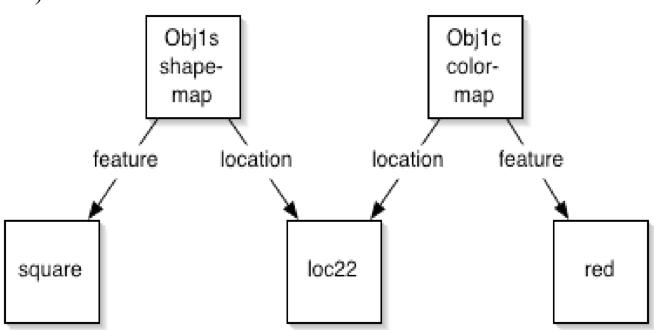

Figure 1: Johnson et al. (2002)'s location-indexed multi-map representations of visual stimuli.

Is from-what-to-where retrieval truly as efficient as from-where-to-what retrieval? Though Nissen's experimental results supported this claim, only accuracy data were provided. Due to possible speed-accuracy tradeoffs, accuracy data alone may not tell the whole story. Yet another way to test the claim is to collect and analyze the reaction time (RT) data as well. If retrievals in the two directions were closely coupled and equally efficient, as suggested by the location-indexed multi-map theory, one would expect similar RTs in either direction. On the contrary, if two directions are not equally efficient, different RTs would be expected.

Experiment

The purpose of the experiment was to explore the interaction of what and where processing by comparing the relative efficiency of retrieval in two directions: from what to where and from where to what. Though Nissen (1985)'s experimental results and Johnson et al. (2002)'s computational model both suggest that the two directions would be similar (see Figure 1), different views exist. O'Reilly and Munakata (2000) reported a connectionist model of spatial attention that involves specific claims of what and where interaction. In that model, a bi-directional link is included to allow a quite direct mutual influence between the spatial where pathway and the object what pathway (see Figure 2). However, the model maintains that the link strengths are not equal for the two directions. In particular, the influence of spatial processing on the object pathway is stronger than the opposite direction, indicated by the thicker arrow in Figure 2. The bi-directional but asymmetric link permits a more flexible balance of multiple factors such as location-based versus object-based attention and top-down versus bottom-up control. It also leads to the prediction that from-where-to-what retrieval should be easier than from-what-to-where retrieval.

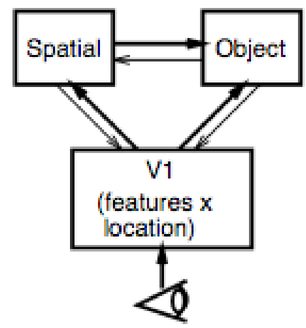

Figure 2: O'Reilly & Munakata (2000)'s model of spatial attention.

The experiment adopted a study-then-test paradigm. In the study phase, an array of objects (drawings), each with a unique location, was presented on a computer screen and the subject was asked to study the array. In the test phase, one of the two retrieval conditions was used. In the from-where-to-what condition, a location was marked and the subject was asked to report the object that had appeared at that location in the study phase. In the from-what-to-where location, a studied object was presented and the subject was asked to report the location where the object had been studied. In either condition, the subject was required to respond as quickly and accurately as possible and the RT was recorded.

One key difficulty in the above design was how to mark or record object locations. To minimize various undesired influences on the RT measures, we adopted a labeling technique. In the from-where-to-what condition, all relevant locations were clearly marked with black squares, except that there was also a question mark appearing in the square of the target location. In the from-what-to-where condition, all relevant locations were again marked with black squares. However, each square was now labeled by a random unique number. The subject had only to report the number that identified the to-be-reported location.

Method

Subjects Twelve graduate students at the University of Texas Health Science Center at Houston were paid to participate in the experiment**.**

Apparatus and Materials Forty black line drawings of common objects were selected from the database developed by Snodgrass and Vanderwart (1980) and randomly assigned to five groups. There was no significant difference among different groups in several major semantic characteristics such as name agreement, image agreement, familiarity, and frequency. Each drawing is 100x100 pixels in size. A windows PC with a 17'' VGA monitor (640x480 resolution) was used to present the stimuli. E-prime was adopted to control the experiment and collect the subject's RT data via a voice key. The subject's verbal response (either a location number or an object name) was recorded by an experimenter sitting next to the subject.

Design Each subject performed both from-where-to-what and from-what-to-where conditions. The order was counter-balanced among subjects. The study phase was the same for both conditions: eight object drawings were presented at eight locations in the center region of the screen (see Figure 3a) and the subject was required to study them. In each trial of the test phase, the subject either had to report the object corresponding to the location indicated by the question mark (from-where-to-what condition, see figure 3b) or report the number (1-8) that appeared at a location corresponding to a centrally presented object (from-what-to-where condition, see figure 3c).

Each subject performed five blocks of each condition, with each block using a unique group of object drawings for studying and testing. In each block, after studying the object array, the subject proceeded to the test phase, in which each just studied object drawing was tested three times, resulting in 24 testing trials in each block (and 24x5=120 testing trials in each condition).

An extra baseline block was performed in the very beginning of each condition. For the from-where-to-what condition, the baseline block consisted of 40 trials in each of which a single object drawing was presented in the center of the screen and the subject just had to report the name of the object as quickly as possible. For the from-what-to-where condition, the baseline condition consists of 24 trials in each of which eight numbers (1-8) were presented at eight locations with only one appearing in red background and the subject had to report that special number as quickly as possible. The RTs in these trials were recorded as baseline for later data analysis.

Procedure The subject was first provided a piece of paper with the forty object drawings and their corresponding names on it and was asked to read them 3 times to get familiar with them. The subject was then led to the testing room where he or she performed the two experimental conditions in a pre-assigned order. There was a 2-minute break between conditions.

In the study phase of each block, the subject was instructed to study and memorize the eight presented object drawings and their locations, at his/her own pace. In each trial of the test phase, a fixation mark "+" was first presented in the center of the screen for 1.5s, accompanied by a brief beep, at which point the condition-specific retrieval cues were presented and the subject was required to make a corresponding response as quickly and accurately as possible. Once a response was made (or 5s has passed with no response), the next testing trial began until all 24 trials were finished. No feedback of the response correctness was provided for each testing trial.

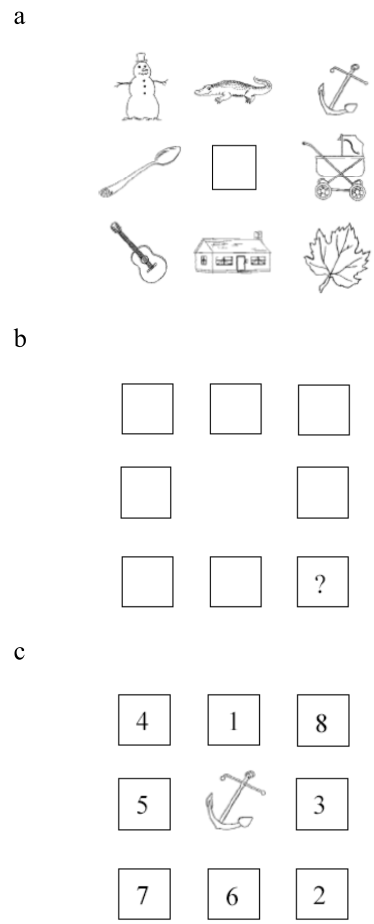

Figure 3: Experimental design. a) study phase; b) from-where-to-what retrieval; c) from-what-to-where retrieval

Results

Accuracy Data The average accuracy was 95.2% for the from-where-to-what condition and 96.3% for the from-where-to-what condition, indicating that subjects could achieve relatively high retrieval accuracy in both conditions.

RT Data Only those RTs from correct trials were used for further analyses. The main results are shown in Table 1.

Several paired t-tests were carried out to compare the RTs in the two conditions. We found a significant difference for the baseline RTs (Diff=-205.6ms, $t(11)=-4.38$, $p<0.001$), a significant difference for the retrieval RTs (Diff=-463.9ms, $t(11)=-5.39$, $p<0.001$), and a significant difference for the RTs of retrieval minus baseline (Diff=-258.3ms, $t(11)=-2.56$, $p<0.02$). While the baseline RT difference is expected due to the well practiced nature of number reading than object naming, the latter two differences were quite surprising. They suggest that retrieval from what to where is faster than retrieval from where to what, contradictory to both predictions described previously.

Table 1: Average RTs in each condition (in ms). The numbers in parentheses are standard errors.

Condition	Baseline	Retrieval	Retrieval-Baseline
From-what-to-where	676.1 (40.9)	1333.0 (101.0)	656.9 (89.0)
From-where-to-what	881.7 (46.4)	1796.9 (75.0)	915.2 (90.3)
Difference	-205.6 (46.9)	-463.9 (86.0)	-258.3 (101.0)

Discussion

Segregation of processing is a general principle of how the brain carries out cognitive functions. It has been well documented that different dimensions of a visual stimulus, including it spatial location and various visual features (e.g., color, shape, and texture) are represented and processed through different neural pathways. One critical question is how different pathways interact with each other to give rise to unified human cognition.

This paper reported an experiment that intended to investigate how the where pathway and the what pathway interact by evaluating and comparing the relative efficiency of retrieval in two directions: from what to where and from where to what. Previous results predicted that the two directions were either equally efficient (e.g., Nissen, 1985; Johnson et al., 2002) or that from-where-to-what retrieval is faster than from-what-to-where retrieval (e.g., O'Reilly & Munakata, 2000). Quite surprisingly, our results contradicted either prediction. Showing that retrieving from what to where is faster than retrieving from where to what, our results imply some quite different underlying representations. Specifically, our results suggest that the link strength from object identity (or other visual features) to its location is stronger than the link strength from object location to its identity. It seems that object location, as an important feature of object, is readily represented and strongly bound with the object representation. Therefore, given an object, its location can be quite quickly retrieved. On the other hand, there may not exist readily retrievable location representations that link to the objects that have occupied that location. Such information may have to be computed online when needed, therefore taking longer time (e.g., Hunt & Waller, 1999).

It is important to note that there are multiple factors that may contribute to the pattern of results in our experiment. For example, we allowed subjects to study the object array at their own pace. On average, our subjects used about 2.5 minutes (range = 1.2 to 3.5 minutes) to study the array, which was very different from Nissen's experiments where the stimuli were presented very briefly (~120ms). As a result, we are actually examining the representations underlying a longer-term memory than the perceptual memory Nissen examined. It is likely that the representations underlying perceptual visuospatial memory is quite different from the representations underlying well-studied longer-term visuospatial memory. In addition, our use of object drawings might play a role. While the number of relevant screen locations was quite limited and well defined in our design, the number of potential objects might be numerous and not well defined, resulting in a type of fan effect (e.g., Anderson & Reder, 1999). A general conclusion should not be drawn until these factors are carefully examined.

Acknowledgments

This work is supported by grants from the Office of Naval Research (Grant Nos. N00014-01-1-0074 & N00014-04-1-0132). We thank Dr. Yanlong Sun for his help in experimental design.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Press.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General, 128*, 186-197.

Farah, M. J. (2000). *The cognitive neuroscience of vision*. Malden, MA: Blackwell Publishers.

Isenberg, L., Nissen, M. J., & Marchak, L. C. (1990). Attentional Processing and the Independence of Color and Orientation. *Journal of Experimental Psychology: Human Perception & Performance, 16*(4), 843-856.

Hunt, E., & Waller, D. (1999). *Orientation and wayfinding: A review* (Technical Report to ONR). Arlington, VA.

Johnson, T. R., Wang, H., Zhang, J., & Wang, Y. (2002). A Model of Spatio-Temporal Coding of Memory for Multidimensional Stimuli. In *The Twenty-Fourth Annual Conference of Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.

Lamy, D., & Tsal, Y. (2001). On the status of location in visual attention. *European Journal of Cognitive Psychology, 13*(3), 305-342.

Monheit, M., & Johnston, J. C. (1994). Spatial attention to arrays of multidimensional objects. *Journal of Experimental Psychology: Human Perception & Performance, 20*(4), 691-708.

Nissen, M. J. (1985). Accessing features and objects: Is location special? In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 205-219). Hillsdale, NJ: Erlbaum.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3-25.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal lobe injury on covert orienting of visual attention. *Journal of Neuroscience, 4*, 1863-1874.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127-158.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of What and Where in the Primate Prefrontal Cortex. *Science, 276*, 821-824.

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174-215.

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

Tsal, Y., & Lamy, D. (2000). Attending to an object's color entails attending to its location: Support for location-special views of visual attention. *Perception & Psychophysics, 62*(5), 960-968.

Tsal, Y., & Lavie, N. (1993). Location Dominance in Attending to Color and Shape. *Journal of Experimental Psychology: Human Perception and Performance, 19*(1), 131-139.

Ungerleider, L. G., Courtney, S. M., & Haxby, J. V. (1998). A neural system for human visual working memory. *Proc Natl Acad Sci U S A, 95*, 883-890.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior*. Cambridge, MA: MIT Press.

van der Velde, F., & van der Heijden, A. H. C. (1997). On the Statistical Independence of Color and Shape in Object Identification. *Journal of Experimental Psychology: Human Perception and Performance, 23*(6), 1798-1812.