Green, C. & Hummel, J.E. (2004). Functional relations affect object detection in non-scene displays. In K. Forbus, D. Gentner, and T. Reiger (Eds.): *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, 488-493. Mahwah, NJ: Erlbaum.

Functional Interactions Affect Object Detection in Non-Scene Displays

Collin Green (cbgreen@ucla.edu)
John E. Hummel (jhummel@psych.ucla.edu)
Department of Psychology, 1285 Franz Hall
University of California, Los Angeles
Los Angeles, CA 90095

Abstract

Two experiments suggest that functional relations influence the processing of visual stimuli. Experiment 1 demonstrated that participants are more accurate to detect targets engaged in functional interactions with related items than when they are simply surrounded by those items. Experiment 2 demonstrated that the accuracy of visual search in a non-scene display is affected when distractor items can be grouped functionally versus when distractor items are simply semantically related to each other. Overall, these data suggest that functional relations between objects affect the allocation of visual attention and by consequence, the processing of natural scenes and other structured visual stimuli.

Introduction

An important aspect of semantic knowledge about objects concerns function. The very identity of an object often hinges upon its intended use. The experiments presented here explore the possibility that participants performing object search tasks may be sensitive to functional relations among the objects being searched. This work is based on the idea that natural scenes are mentally represented in terms of the functional groups they comprise (Green & Hummel, 2004). For example, a coffee shop may be defined as a place where it is possible to make, buy, sell, and drink coffee. The objects associated with these activities (a table and chair in certain arrangement suggest dining) form the basic units of the scene definition.

While scene categories are difficult to define in terms of the objects present (the same objects may form different types of scenes by virtue of different arrangements), or in terms of the spatial layout only (the identities and meanings of objects have bearing on scene categorization), a function-based scene representation may provide a consistent, flexible, and useful definition (see Green & Hummel, 2004, for a more thorough discussion).

In both experiments presented here, the presence of functional relations (the presence of meaningful structure) in the stimulus was expected to improve performance: In Experiment 1, we expected the processing of a target object in a functional relation to be facilitated (relative to a target adjacent to the same objects, but not interacting with any of them). In Experiment 2, we expected that functionally meaningful relations would effectively unitize pairs of distractor objects, making search more efficient than when such objects must be rejected one by one.

Experiment 1

Experiment 1 investigated whether functional interactions would affect observers' ability to detect and locate target objects in non-scene displays. The experiment required observers to indicate whether a named target object was present in a masked, briefly-presented array of twelve line-drawn objects. We manipulated whether the search array contained a distractor object semantically associated to the named target, and whether the target and associated distractor (if both were present) were interacting.

In general, the addition of an associated distractor object to a search array impairs performance in visual search. Moores, Laiti, & Chelazzi (2003) found that when participants searched for a target, distractor objects semantically associated with the target had the effect of reducing accuracy and increasing latency relative to when distractors were not associated with the target.

In the current experiment, we expected a similar result. Overall performance should be lower when an associated distractor object is present in the search array relative to when no associated distractor is present (though Auckland, Cave, & Donnelly, 2004, find evidence for the opposite effect). However, it remains unclear whether such effects interact with relational information in guiding visual search. Specifically, there is reason to believe that the introduction of functional interactions between targets and associated distractors will modulate the impairment caused by target-distractor associations, to some degree.

Riddoch, Humphreys, Edwards, Baker, & Willson (2003) found that functional interactions facilitated the processing of the interacting objects. Their subjects were parietal patients who showed extinction when trying to report the names of two simultaneously-presented objects. When objects were presented together but were not interacting, the patients could reliably report the name of one object, but not both. When the objects were positioned to interact, patients showed increased ability to accurately report the name of the second object. This suggests that functional relations may in fact play a special role in the processing of visual stimuli.

Experiment 1 brought together the two results mentioned above, combining semantic associations between targets and distractors with functional interactions between targets and distractors in a single experiment. Based on the results of Moores, et al. (2003) and Riddoch, et al. (2003)

we expected that introducing a semantically associated distractor to a search array would impair target detection, but that this effect would be reduced when the target interacted with the associated distractor.

Method and Materials

Stimuli All materials were presented on a Macintosh iMac personal computer running the SuperLab application. Stimuli were composed of black and white line drawings (some taken from Snodgrass & Vanderwart (1980), others created specifically for this work) that depicted everyday objects.

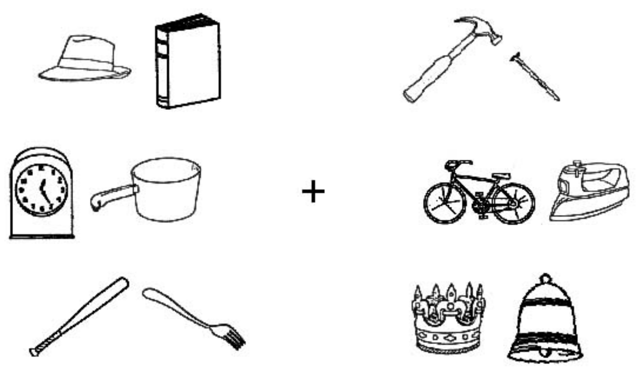

Figure 1: Typical stimulus from Experiment 1. Here the target (hammer) is interacting with a related distractor item (nail).

All search arrays employed the same basic layout (see Figure 1). A fixation cross was centered in the stimulus array. Twelve objects (each approximately 2.3° visual angle in width/height) were arranged around the fixation cross in two concentric circles. The inner circle had a radius of approximately 4.5° visual angle, and the outer circle had a radius of approximately 7.9° visual angle. Six objects were centered on the inner circle, with objects located at 45°, 90°, 135°, 225°, 270° and 315° from vertical. The six remaining objects were centered on the outer circle. Objects on the outer circle were placed horizontally in line with objects on the inner circle. In this way, the twelve objects made up six pairs. This layout placed paired objects closer to each other than to any other object in the array.

Critically, we manipulated the presence and position of a related distractor item in the search display. On target-present trials, the target could appear with no semantically-related distractors (the *target-only* condition), with a semantically-related distractor in a location not adjacent to the target (*non-adjacent*), paired with, but not interacting with the target (*adjacent*), or paired with and interacting with the target (*interacting*). Catch trials were presented in which the related distractor was present without the target (*distractor-only*), and in which neither the related distractor nor the target were present (*none*).

Each participant completed 24 randomly-ordered trials, with each trial using a different target object. Each participant saw one of 24 counterbalanced sets of stimuli. Across counterbalancing sets, every target object appeared in every condition equally often.

Participants Participants were 40 undergraduate psychology students at the University of California, Los Angeles. Participants took part in the experiment as part of a research requirement for a psychology course.

Procedure Participants were instructed to look for named target objects and indicate (a) whether the target object was present and if so, (b) its location in the search display. The participant was given a description of how each trial would proceed and what responses were required. The participant viewed a single practice trial, with the experimenter providing a verbal description of what was happening at each step and how the participant should respond. The experimenter emphasized that accuracy was important in all responses, but that the speed of response mattered only during the detection task.

Each trial proceeded as follows: First, a word naming the target object appeared in the center of the computer screen in black 24-point Arial font and remained on the screen until the participant pressed a key. Then, a fixation cross appeared in the center of the screen. After 750ms, the fixation cross was replaced by a search array. The search array was visible for 250ms and was subsequently masked until response or until the trial timed out (2500ms after search array onset).

The participant indicated whether or not the target object was present in the search array by making a key press (yes or no) as quickly and accurately as possible. After response, the participant was presented with an labeled layout of the search array and was asked to indicate the location in which the target object appeared (or to verify that the target object did not appear) by pressing a letter on the keyboard. This response was not speeded. A 1000ms inter-trial interval during which the computer screen was blank preceded the next trial.

Results

Accuracy and response time (RT) data were analyzed using within-subjects ANOVAs. Trials upon which detection RT exceeded 2500ms were counted as errors. Error trials were excluded from all RT analyses.

Detection Accuracy Accuracy data (***d'***) from the detection task are presented in Table 1. The main effect of stimulus condition on detection accuracy only approached significance ($F(3,117) = 1.927$, $MSE = 0.621$, $p = 0.129$). However, planned comparison indicated that mean ***d'*** in the Interacting condition was significantly higher than mean ***d'*** in the Adjacent condition ($t(39) = 3.242$, $SE = 0.126$, $p = 0.002$). This comparison is the most revealing with respect to the effect of functional interactions, as the only difference between the Interacting and Adjacent conditions is the orientation of the associated distractor object. Of the four conditions, only the Interacting condition produced performance significantly different than chance ($t(39) = 1.895$, $SE = 0.168$, $p = 0.0325$ one-tailed).

Detection Response Time RT data are presented in Table 1. Mean RT on the detection task did not vary across conditions ($F(3,111) = 0.537$, $MSE = 48600$, $p = 0.658$). No pair-wise comparisons yielded significant differences.

Table 1: Summary of response time and accuracy data from Experiment 1.

Condition	Detection d'	Detection RT	Localization accuracy
Interacting	0.319	1340 ms	0.558
	(SE = 0.168)	(55)	(0.05)
Adjacent	-0.091	1388	0.488
	(0.17)	(61)	(0.06)
Non-Adjacent	0.024	1398	0.431
	(0.19)	(54)	(0.06)
Target Only	0.078	1362	0.502
	(0.14)	(57)	(0.06)

Localization Accuracy Accuracy data for the localization task are presented in Table 1. As a measure of localization accuracy, we report the probability that the correct location would be chosen given that a target was present and the observer attempted to localize the target. That is, we excluded target-absent trials, and trials where the observer made a localization response indicating that the target did not appear in the search array.

There was no main effect of stimulus condition on localization accuracy ($F(3,105) = 0.911$, $MSE = 0.108$, $p = 0.439$). No pair-wise comparisons were significant.

Discussion

Though weak, these results do suggest that functional relations influence the processing of objects in non-scene displays. This effect obtained even though the task did not require participants to use (or even notice) the functional relations in the stimuli. Indeed, it may be argued that functional information is not useful in this task. Only one sixth of the trials each participant saw contained a meaningful functional relation between the target and related distractor item. One would not expect the pattern of results observed were the processing of functional relations effortful.

If functional information influences the allocation of visual attention during simple search tasks, then it seems plausible that the guidance of visual attention during search of natural, structured scenes is also influenced by such information. Heuristics about what kinds of objects should appear together in scene could help the visual system to efficiently deploy attention, and facilitate the processing of scene-consistent stimuli. The finding that functional relations affect the processing of simple visual stimuli also suggests that visual representations may include abstract, functional information.

Experiment 2

Experiment 1 suggests that functional relations do influence the processing of visual stimuli during a search task. The presence of a functional relation involving the target object and an associated distractor object increased detection accuracy relative to when an associated distractor was adjacent to, but not interacting with the target. That result does little to discriminate between the possibility that interacting objects are processed more efficiently than other objects and the possibility that functional interactions capture visual attention.

Experiment 2 sought to decide between these explanations. In this experiment, distractor objects engaged in functional interactions, and the number of functional groupings in the search array was varied parametrically. If functional groups capture attention, then one would expect the addition of interacting distractor pairs to impair performance, performance suffering increasingly as more interacting pairs are added. On the other hand, if objects engaged in functional relations are processed more efficiently than objects not engaged in functional interactions, then one would expect performance to improve as more interactions are introduced to distractor objects. Search time in non-scene displays is a function of the number of distractor items present (Biederman, et al., 1988). If functionally interacting objects form perceptual groups, then adding interactions among distractors while holding the total number of display objects constant should effectively reduce the number of perceptual units that must be searched. As a result, displays with more interactions should yield superior search performance.

Method and Materials

Stimuli All materials were presented on a Macintosh iMac personal computer running the SuperLab application. Stimuli were composed of a subset of the black and white line drawings used in Experiment 1.

The experimental trials were divided into four conditions: zero functional interactions (the *0i* condition), one interaction (*1i*), two interactions (*2i*), or three interactions (*3i*). In addition to the four experimental conditions, two control conditions were run to provide baseline search performance measures. All search arrays in the four experimental conditions employed the same basic layout (see Figure 2). A fixation cross was centered in the stimulus array. Eight objects (each approx. 2.3° visual angle in width/height) were arranged around the fixation cross in two concentric circles. The inner circle had a radius of approximately 4.5° visual angle, and the outer circle had a radius of approximately 7.9° visual angle. Four objects were centered on the inner circle, located at 45°, 135°, 225°, and 315° from vertical. The four remaining objects were centered on the outer circle. Each object on the outer circle was placed horizontally in line with an object on the inner circle. In this way, the eight objects made up four pairs. As before, objects within a pair were closer to each other than to any other object in the array.

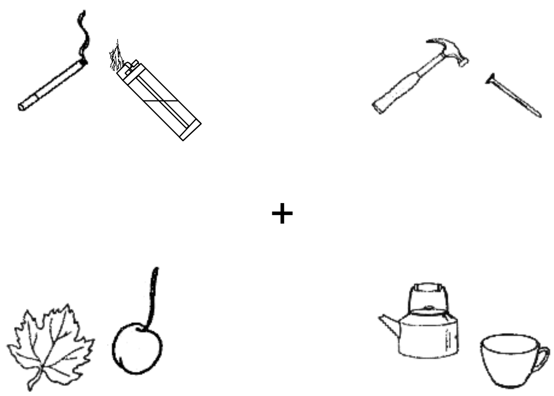

Figure 2: Typical stimulus from Experiment 2. Here, the target (leaf) is accompanied by seven distractor objects, of which four are engaged in interactions (lighter-cigarette, hammer-nail). The other distractor pair (kettle-cup) is related but not interacting. This is an example of a *2i* stimulus.

The identity and orientation of distractor items in the search display were manipulated. In each array, one distractor object was paired with the target object (or a lure). The remaining six distractor objects were organized into three pairs. The distractor objects in each pair were semantically associated and capable of entering into a functional interaction. In this experiment, participants performed trials in which there were zero, one, two, or three of these distractor pairs were actually arranged to interact.

In the control-five object condition (the *5c* condition), the target (or lure) was accompanied by four distractor objects that were unrelated to the target, and unrelated to each other, for a total of five objects in each array. In the control-eight object condition (*8c*), there were seven distractor objects unrelated to the target and each other, for a total of eight objects in each search array. Objects in the *8c* condition were arranged in accordance with the layout depicted in Figure 2. Objects in the *5c* condition occupied five of the eight positions (randomly selected) in the standard search array for this experiment.

Each participant completed 228 randomly-ordered trials. Each participant saw every target object in every condition, but no one target appeared with the same distractor objects in more than one array.

Participants 40 undergraduate psychology students at the University of California, Los Angeles participated in the experiment as part of a research requirement for a psychology course.

Procedure The procedure in Experiment 2 was identical to that of Experiment 1. On each trial participants viewed a target label and a briefly-presented search array which was masked. Participants made a speeded response indicating the presence or absence of the target object, and then a non-speeded location response. Instructions were identical to those in Experiment 1.

Results

Response time (RT) and accuracy data were analyzed using within-subjects ANOVAs. Trials on which detection RT exceeded 2500ms were counted as errors. Error trials of all types were excluded from all RT analyses.

Detection Accuracy Accuracy data (***d'***) are presented in Table 2. There was a main effect of stimulus condition (including control conditions) ($F(5,195) = 4.750$, $MSE = 0.17$, $p < 0.001$). There was also a significant effect of condition among the four experimental conditions (*0i*, *1i*, *2i*, and *3i*) ($F(3,117) = 3.485$, $MSE = 0.164$, $p=0.018$). Detection accuracy was significantly higher in the *5c* condition than in the *8c* condition ($F(1,36) = 4.152$, $p = 0.04$).

Planned comparisons indicated that accuracy in the *1i* condition was significantly worse than in the *0i*, and *3i* conditions, but not the *2i* condition. The *0i*, *2i*, and *3i* conditions were not significantly different than each other.

Trend analysis indicated that there was a significant increasing linear trend in detection accuracy across the *1i*, *2i*, and *3i* conditions ($F(1,39) = 8.684$, $MSE = 0.169$, $p = 0.005$). In addition, there was a significant quadratic trend across the *0i*, *1i*, *2i*, and *3i* conditions ($F(1,39) = 6.085$, $MSE = 0.151$, $p = 0.018$).

Detection Response Time RT data are presented in Table 2. There was no significant difference in RT across the six experimental conditions, ($F(5,200) = 1.199$, $p=0.311$).

No pairwise contrasts were significant, but participants were marginally faster to accurately respond in the *5c* condition, than in the *8c* condition ($F(1,40) = 3.137$, $p = 0.084$).

Localization Accuracy Accuracy data for the localization task are presented in Table 2. As in Experiment 1, we report the probability that the correct location would be chosen given that the target was present and the observer attempted to localize the target.

There was a significant main effect of stimulus condition on localization accuracy ($F(5,195) = 38.649$, $MSE = 0.006$, $p < 0.001$). There was also a main effect of stimulus condition across the *0i*, *1i*, *2i*, and *3i* conditions ($F(3,117) = 3.522$, $MSE = 0.005$, $p = 0.017$). Post-hoc analysis indicated that localization was significantly more accurate in the *3i* condition than in the *0i* condition, and that a significant linear trend existed across the *0i*, *1i*, *2i*, and *3i* conditions (F(3,117) = 9.597, MSE = 0.005, p = 0.004).

Table 2. Summary of Reaction Time and Accuracy Data for the Detection Task in Experiment 2.

Condition	Detection d'	Detection RT	Localization accuracy
0i	1.40 (0.09)	812 ms (SE = 30)	0.85 (0.02)
1i	1.17 (0.10)	822 (34)	0.88 (0.02)
2i	1.36 (0.12)	833 (34)	0.87 (0.02)
3i	1.44 (0.09)	820 (32)	0.90 (0.01)
5c	1.61 (0.10)	810 (32)	0.68 (0.01)
8c	1.42 (0.10)	831 (32)	0.83 (0.02)

Discussion

The results of Experiment 2 suggest several possibilities. If the only meaningful difference among the four experimental conditions is the decrease in accuracy observed in the *1i* condition, then the data support an attention-capture account. Specifically, if the presence of a single functional group among distractors impairs performance, then it is possible that the participant's attention was drawn to the functional interaction (which never contained the target) and so the actual target was detected less often. One could explain the disappearance of this effect in the *2i* and *3i* conditions if the ability of a functional group to capture attention is dependent on its uniqueness in the display. Adding multiple functional interactions among distractors may "wash out" such an effect by bringing the average salience of the display elements closer to the maximum salience of any one element (i.e., the salience of a functionally interacting pair is farther from the mean salience of the display when one interaction is present than when multiple interactions are present). This explanation is somewhat unsatisfying, and the trend analyses performed suggest a more interesting alternative.

There was a significant linear trend among the *1i*, *2i*, and *3i* conditions, and a significant quadratic trend among those and the *0i* condition. The shape of these data suggest that the addition of functional interactions among distractor objects did not strictly hurt performance (as predicted by an attention-capture account), nor did the introduction of interactions strictly improve performance (as predicted by a grouping account). It seems possible that functional groups do capture attention, but that they also facilitate the processing of the objects they comprise. The drop in performance from the *0i* to *1i* conditions would indicate that the increase in search efficiency resulting from the inclusion of a single interacting pair of distractors did not outweigh the cost incurred by that pair's tendency to capture attention. However, as more interacting distractor pairs were added, the accumulated gains from more efficient processing of interacting objects improved overall performance. Performance in the *3i* condition was only numerically superior to that in the *0i* condition, but if the linear trend across the *1i* to *3i* conditions is extrapolated, then one can imagine that the continued addition of functional interactions among distractor would produce performance reliably exceeding that in the *0i* condition. In fact, in the extreme case, imagine searching for a random object among a disorganized array of distractors versus searching for the same object among distractors organized into a coherent scene. Both intuition and empirical evidence suggest that search will be more efficient in the latter case (Loftus & Mackworth, 1978; Hollingworth & Henderson, 2000).

Finally, localization accuracy data from Experiment 2 suggest that the extraction of spatial information about objects in a stimulus is more efficient when that stimulus includes functional relations between objects. Notably, it was organization of non-target objects that led to this advantage. In Experiment 1, a similar (but unreliable) advantage was observed for localization of target objects that engaged in functional interactions.

Conclusions and Future Directions

The current work was motivated by the idea that natural scenes are mentally represented in terms of the functional groups they comprise (Green & Hummel, 2004). Experiment 1 suggested that objects are more easily detected or identified when they were interacting with related distractors as compared to when they were not interacting with related distractors, or when no related distractors were present. Experiment 2 showed an interesting non-monotonic pattern associated with the introduction of functional interactions to the display. The addition of a single interaction seemed to impair detection, while performance improved with addition of subsequent interactions.

The data from Experiments 1 and 2 suggest that functional interactions may have two effects on visual search: 1) single functional groups may capture attention; 2) objects in functional groups may be processed more efficiently than objects not engaged in interactions.

Existing data from eye movement studies are not consistent with the first claim. A number of studies (De Graef et al, 1990; Henderson et al., 1999; Loftus & Mackworth, 1978) suggest that visual information (e.g., local contrast, spatial frequency, color) is the main determinant of fixations early in natural scene viewing. Evidence indicates that during natural viewing, scene-consistent objects are fixated more rapidly than inconsistent objects, but that this type of semantic information only mediates eye movements after the first several fixations on a scene (De Graef et al., 1990). In short, semantic information does not seem to influence early fixations, playing a role only later in visual scanning. This suggests that functional groups (which include abstract semantic information) should not capture attention.

Alternatively, the data from these experiments may be explained as an effect of familiarity with canonical arrangements of objects. Because people are routinely exposed to objects arranged in functionally meaningful ways, some other non-attentional influence may be involved. Specifically, the existence mental symbols that represent entire familiar functional groupings may influence the preattentive grouping of a visual stimulus array and lead to faster processing of the objects therein. The existence of perceptual groupings based on functional information (which were not considered in those studies) might affect early attentional guidance in way that is not easily understood if one is looking for effects of semantic consistency only.

Empirical and computational work have been used to study the effects of perceptual grouping on visual search with basic perceptual stimuli (e.g., colored shapes, oriented lines). Some models of search account for effects of perceptual grouping better (and more naturally) than others. For example, the Spatial and Object Search (SOS) model (Grossberg, Mignolla, & Ross, 1994) places perceptual grouping processes at the center of visual search operations. Grouping processes take place pre-attentively in the SOS model, an assumption consistent with a number of empirical results (e.g., Humphreys, et al., 1989).

An important aspect of the SOS model with respect to the functional grouping hypothesis is that its perceptual grouping mechanisms are linked to spatially-invariant representations of objects. SOS allows knowledge about objects to influence perceptual grouping (presumably, so that objects form perceptual units, instead of collections of object parts or features). An extension of SOS might employ representations above the level of objects (e.g. representations of functional groups) to make contact with grouping processes as well. A preattentive grouping mechanism linked to representations of functional groups might yield effects like those observed in Experiments 1 and 2.

Whether functional groups are perceptual groups remains to be established, but the results of Experiments 1 and 2 suggest that familiar functional relations (interactions between objects in a visual scene) may be an important component of visual processing. Current work addresses the possibility that functional groups are in fact perceptual objects.

Acknowledgments

The authors thank Irv Biederman, Steve Engel, Keith Holyoak, Zili Liu, Brian Stankiewicz, members of the LISA lab, and the CogFog group for helpful discussion and comments on this work. Also, thanks to Jerlyn Tolentino and Erica Weiss for many hours of effort in the lab. This work was supported by NIH/NINDS NRSA F31-NS43892-02.

References

Biederman, I., Blickle, T.W., Teitelbaum, R.C. & Klatsky, G.J. (1988). Object search in non-scene displays. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14(3)*, 456-467.

De Graef, P., Christiaens, D., & D'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research/Psychologische Forschung, 52(4)*, 317-329.

Green, C. & Hummel, J.E. (2004). Relational Perception and Cognition: Implications for Cognitive Architecture and the Perceptual-Cognitive Interface. In B.H. Ross (Ed.): *The Psychology of Learning and Motivation, Vol. 44*. San Diego: Academic Press. 201-226.

Grossberg, S., Mignolla, E., & Ross, W.D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review, 101(3)*, 470-489.

Henderson, J. M., Weeks, P. A., & Hollingsworth, A. (1999). The Effects of Semantic Consistency on Eye Movements During Complex Scene Viewing. *Journal of Experimental Psychology: Human Perception & Performance, 25(1)*, 210-228.

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition, 7(1/2/3)*, 213-235.

Humphreys, G.W., Quinlan, P.T., & Riddoch, M.J. (1989). Grouping processed in visual search: Effects with single and combined feature targets. *Journal of Experimental Psychology: General, 118*, 258-279.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception & Performance, 4(4)*, 565-572.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6(2)*, 182-189.

Riddoch, M.J., Humphreys, G.W., Edwards, S., Baker, T. & Willson, K. (2003). Seeing the action: Neuropsychological evidence for action-based effects on object selection. *Nature Neuroscience, 6(1)*, 82-89.

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(2)*, 174-215.