Title
Objects that Work Together May be Perceptually Grouped

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN
1069-7977

Authors
Green, Collin
Hummel, John E.

Publication Date
2005

Peer reviewed

Objects That Work Together May Be Perceptually Grouped

Collin Green (cgreen@arc.nasa.gov)
NASA Ames Research Center, Moffett Field, CA 94035

John E. Hummel (jhummel@psych.ucla.edu)
Department of Psychology, University of California, Los Angeles CA 90095

Abstract

Perceiving functional relationships between objects may be fundamental to understanding visual environments. Even the identification of objects in a scene may be influenced by the functional relations that exist among those objects. In two experiments, normal observers identified a briefly-exposed target object presented with a semantically associated or unassociated distractor object. The target object and distractor were either arranged to work together, or not to do so. Identification was more accurate when target objects were arranged to work with an associated distractor than when they were arranged so as not to work with the distractor. In contrast, identification was worse when target objects worked with unassociated distractors than when they did not. A sensitivity to change in stimulus onset asynchrony suggested that the former (facilitatory) effect was perceptual in nature. We propose this as evidence that sets of objects engaged in familiar functional relationships are perceptually or attentionally grouped.

Keywords: Scene perception; object recognition; function; perceptual grouping.

Introduction

Real world visual scenes are full of information. At the lowest level, the retina encodes a scene as a set of brightness and color values. Later, systems specialized for object and space representation encode the identities and locations of objects in the environment. The result of processing is a representation that permits the observer to recognize the type of place being viewed, and to understand what activities can and should be performed. Clearly, a scene is represented very differently at the retinal level than at this higher level. While the properties of retinal representations are relatively well understood, the properties of the later, more abstract representations (those connected to behavior) are less clear.

Work by Mandler and colleagues (Mandler & Parker, 1976; Mandler & Ritchey, 1977) demonstrated that certain types of information are preferentially encoded while viewing organized scenes. Participants were asked to discriminate between studied (target) scenes and unstudied (lure) scenes. Lure scenes were constructed by making subtle alterations to target scenes and were used to test participant sensitivities to a variety of changes in object and relational information. Results indicated that lures containing object type changes were more easily rejected than those containing object token changes (for example, changing a mug to a plate was more noticeable than changing a mug to a different mug). Additionally, qualitative changes to relations were more easily noticed than metric changes (e.g., when a chair was turned toward a table in the target scene, subjects successfully rejected lures in which the chair was turned away from the table but not in which the chair was farther from [but still facing] the table).

One interpretation of these findings is that scene representations include general information about the semantics of objects (but not much visual detail) and information about meaningful relations between objects (but little about the specific metrics of those relations). In other words, scene-level representations may emphasize the information in the perceptual stimulus which is useful for understanding and interacting with an environment, omitting specific visual details: information that is specifically *functional*.

Building on this interpretation, we recently proposed an account of scene processing based on the representation of functional groupings of objects within larger visual scenes (Green & Hummel, 2004a). In short, groups of interacting objects within a visual scene may serve as a basis for recognizing scenes and for connecting visual information to goal-relevant actions. We hypothesized that functional groups are explicitly represented mental entities, and that these representations affect the allocation of visual attention; preliminary data support this hypothesis (see Green & Hummel, 2004b). The current work further investigated whether functional groups affect object identification in simple visual scenes.

The hypothesis that functional interactions between objects influence visual processing has prior empirical support. Evidence suggests that functional relations affect the identification of visual objects in neuropsychological patients with parietal damage (Riddoch, Humphreys, Edwards, Baker, & Willson, 2003; Humphreys, Riddoch, Forti, & Ackroyd, 2004). Riddoch et al., (2003) studied patients who showed extinction when trying to report the names of two simultaneously-presented objects. When objects were presented together but were not positioned to interact (i.e. were not working together to accomplish some larger goal), patients could report the name of one object, but not both. When the objects were positioned so that they interacted, accuracy in reporting the names of both objects increased markedly. Control conditions indicated that semantic associations between objects were not sufficient to explain the improved performance; rather, it was functional information that facilitated the simultaneous selection of the interacting objects.

The work by Riddoch and colleagues is evidence in favor of our hypothesis. If functional groupings of objects are explicitly represented mental entities, then one can imagine that the constituents of such groups would not compete for selection; the entire group could be simultaneously selected. Yet, it is important to determine whether the effects described by Riddoch et al. (2003) were the result of the deficit suffered by the patient population studied, or a property of normal cognition that became more apparent in the presence of parietal damage. In addition, if it is the case that sensitivity to functional information is a property of normal scene processing, then it is worthwhile to determine whether these effects are strong enough to manifest when observers are otherwise unimpaired.

Here, we present two experiments that explored the influence of functional interactions on object identification in normal observers. Our experiments examined whether or not functional interactions between objects affected their identification, and whether or not such effects were sensitive to the semantic associations between objects. de Graef, Christaens, & d'Ydewalle (1990) noted that some scene context effects can be explained as consequences of post-perceptual decision processes. Accordingly, using a manipulation of stimulus onset asynchrony (SOA) we tested whether any such effects of functional information were due to enhanced perception or to post-perceptual processes.

Experiment 1

Experiment 1 required observers to verify whether the second object in a two-object sequence matched a label presented prior to the trial. A target object appeared to the left or right of fixation shortly after a distractor object appeared at fixation. We manipulated the semantic relationship between the distractor the label, and whether the distractor was arranged to interact with the target object (see Figure 1).

Method and Materials

Participants Ten UCLA undergraduates participated to fulfill a requirement for a psychology course.

Stimuli Twenty black and white line drawings of common objects (approximately 2.3° visual angle in width) served as stimuli. The objects consisted of ten semantically-associated pairs (e.g., pitcher-glass, hammer-nail, etc.) which could be arranged to form a familiar interactive relation (see Figure 1). Within a pair, one object was designated the target object and one the distractor object. The distractor was always functionally asymmetric, operating primarily in one direction (e.g., a pitcher may pour from only one side). On each trial, one target object and one distractor object were presented, though these were not necessarily from the same semantically-associated pair. Eight of the stimulus objects were taken from Snodgrass and Vanderwort (1980) and 12 were created specifically for this work.

Each of the ten object pairs was associated with a label that named the target object in the pair. Labels were displayed on the computer screen in black, 24-point, Arial font on a white background.

Stimuli were presented on MacIntosh personal computers with observers seated approximately 66 cm from the computer monitor. SuperLab (version 1.5) was used to manage stimulus presentation and data collection in all experiments.

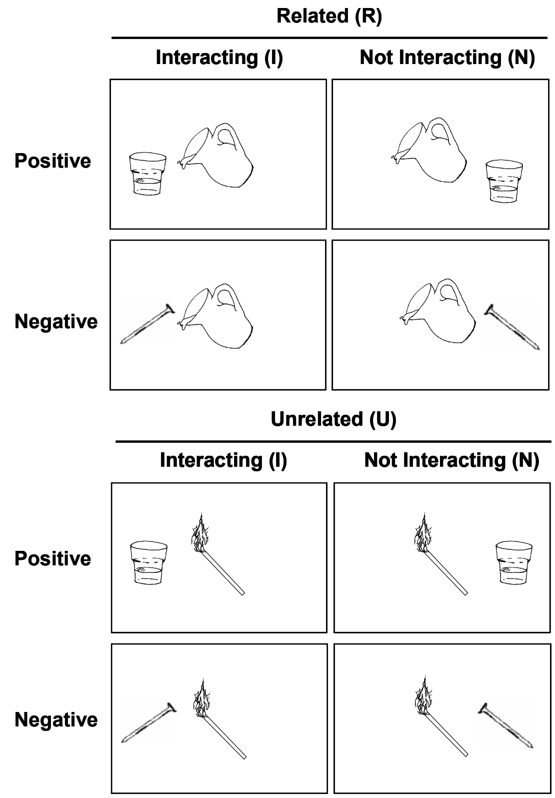

Figure 1: Examples of stimuli in each condition. Here, the label was "glass". Distractors could be Related (R) or Unrelated (U) to the label, and could be oriented to Interact (I) or Not Interact (N) with the target object (the target object was either the target [glass], or a lure [nail]). The same set of stimuli was used in both experiments.

Procedure Each subject completed 320 trials (see Figure 2). Each trial began with the presentation of a label. The label was displayed in the center of the screen until the observer pressed a key. Upon key press, a fixation cross replaced the label and remained on the screen for 750 ms. A distractor object was then presented for 50 ms, followed by an inter-stimulus interval (ISI) of 50 ms (SOA = 100 ms) consisting of a blank screen. A target object then appeared for 50 ms, followed by a blank screen, which remained until the observer pressed the "Z" key (present) or the "/" key (absent) indicating whether or not the target object matched the label presented prior to the trial. Target objects appeared lateralized approximately 4.5° to the left or right of fixation. Observers did not know whether the target object would appear to the left or right, and the locations were used equally often. The trial timed out if no response was made within 2500 ms of the onset of the target object. The next trial began after a 1000 ms inter-trial interval. Observers were instructed to respond as quickly as possible without making mistakes.

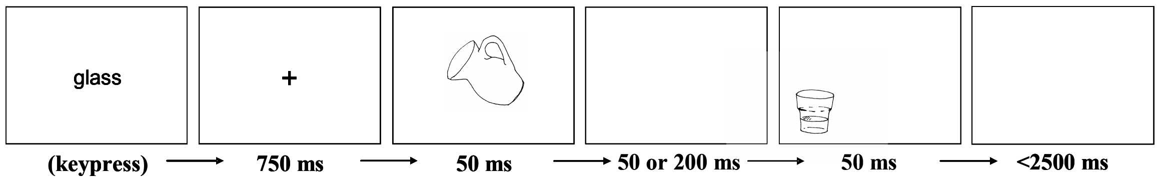

Figure 2: Schematic of trials in Experiments 1 and 2. Each trial began with the presentation of a label. Following a keypress, observers saw a fixation cross and then brief presentations of the distractor and target object. Distractor and target objects were presented with an SOA of 100 ms (Experiment 1) or 250 ms (Experiment 2). Observers had up to 2500 ms to indicate (by keypress) whether or not the target object matched the label.

Design Three within-subjects factors were orthogonally crossed: Label-Distractor Relatedness (Related or Unrelated), Functional Interaction (Interacting or Not Interacting) and Trial Type (Positive or Negative). On Related (R) trials the distractor object came from the stimulus pair associated with the label; on Unrelated (U) trials, the distractor came from a different pair. On Interacting (I) trials the distractor was oriented to function toward the target object; on Not Interacting (N) trials, the distractor was oriented to function away from the target object (see Figure 1). On Positive trials, the target object matched the label; on Negative trials, the target object did not match the label.

It is important to note that Label-Distractor Semantic Relatedness describes the relationship between the distractor object and the label, not the relationship between the distractor and the target object. For example, in the RI/Negative trial depicted in Figure 1 (lower left corner), the label was "glass" and the distractor object (pitcher) came from the object pair associated with that label. However, the target object that was presented (nail) and the distractor object (pitcher) were unrelated. We manipulated the relationship between the distractor object and label instead of the relationship between the distractor and target objects so that we might better observe any bias produced by the presence of a distractor object that was semantically related to the label.

Predictions Our functional groups hypothesis predicts that objects engaged in familiar functional interactions will be better identified than objects not engaged in such interactions. In the context of Experiment 1, we predicted a two-way interaction of Label-Distractor Semantic Relatedness and Functional Interaction. Specifically, we predicted a simple main effect of Functional Interaction on target object identification for Related trials (such that RI performance would exceed RN performance), and further, that this effect would be larger than the simple main effect of Functional Interaction for Unrelated trials (i.e., we predicted [RI – RN] > [UI – UN]). Failing to find this two-way interaction would contradict our hypothesis that functional groups are explicitly represented and influence visual processing.

Analysis In both experiments, Response Time (RT [ms]) and accuracy data (d') were analyzed using within-subject ANOVAs. Trials for which RT was longer than 2500 ms were counted as errors.

RTs were analyzed only for trials to which observers responded correctly. RTs did not differ reliably across conditions in any experiment. In addition, RT data did not suggest any speed-accuracy tradeoffs. Consequently, we report only accuracy results in the remainder of this article.

Results

Data from Experiment 1 are presented in Figure 3. The main effect of Label-Distractor Relatedness on identification was significant ($F[1, 9] = 12.792$, $MSE = 0.146$, $p < 0.05$). There was no main effect of Functional Interaction ($F[1, 9] = 2.037$, $MSE = 0.177$, $p > 0.15$). However, there was a significant interaction between Label-Distractor Relatedness and Functional Interaction ($F[1, 9] = 51.234$, $MSE = 0.070$, $p < 0.05$).

Simple main effect analyses indicated that mean d' was significantly higher in the RI condition (mean d' = 3.22) than in the RN condition (2.82) ($t[9] = 3.303$, $SE = 0.124$, $p < 0.05$). In contrast, mean d' was significantly lower in the UI condition (2.19) than in the UN condition (2.98) ($t[9] = 4.277$, $SE = 0.185$, $p < 0.05$).

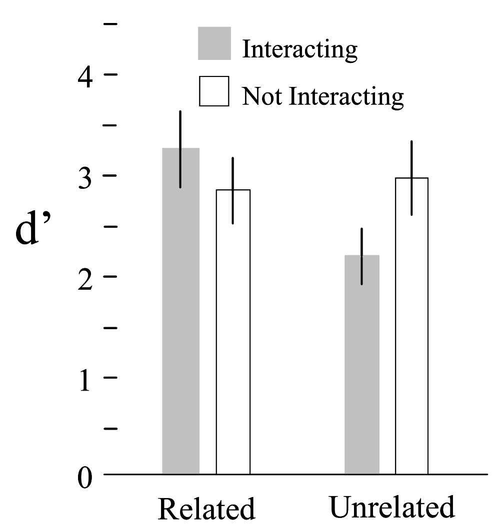

Figure 3: Accuracy data from Experiment 1.

Discussion

Results from Experiment 1 showed the predicted interaction of Label-Distractor Relatedness and Functional Interaction with respect to object identification. Identification of the target object was better for trials where it interacted with a distractor that was semantically related to the label than trials where it did not interact with a distractor that was semantically related to the label (illustrated by the difference between the two left-most bars in Figure 3). For example, it was easier for participants to determine whether an object was a glass when it interacted with a pitcher than when it did not. This effect contrasted with an impairment when the target object interacted with a distractor that was unrelated to the label (illustrated by the difference between the two right-most bars in Figure 3). For example, it was harder for participants to determine whether an object was a glass when it interacted with a chair than when it did not. Together, these results suggest that functional interactions influence object identification, and that the familiarity of object pairings (here, their semantic association) is important in determining the direction of the effect.

Experiment 2

The results of Experiment 1 suggest that familiar functional groups influenced observers' perceptual and/or attentional processes. However, as pointed out by de Graef, Christaens, and d'Ydewalle (1990), context effects in scene processing and object identification are sometimes attributable to post-perceptual processes. Specifically, those authors pointed out that in the presence of incomplete or uncertain perceptual information, observers may adopt educated guessing strategies that yield context effects. As such, it is possible that the observed advantage for familiar interacting object pairs in Experiment 1 is attributable to such a strategy. Experiment 2 sought to determine whether or not the effects observed in Experiment 1 were perceptual in nature, or due to post-perceptual processes such as educated guessing.

Experiment 2 replicated Experiment 1 with a longer SOA (250 ms instead of 100 ms). Post-perceptual effects should grow in magnitude (or at least remain present) when observers are given additional time to process stimulus objects and utilize strategies in response. However, to the extent that the effects in Experiment 1 were perceptual rather than strategic, a lengthened SOA would be expected to diminish or eliminate them (e.g., by decreasing observers' tendency to perceptually integrate the objects in each pair; di Lollo, Hogben, & Dixon, 1994).

A post-perceptual account of the effects observed in Experiment 1 predicts effects of the same size (or larger) in Experiment 2 as observed in Experiment 1. On the other hand, a perceptual account of those effects predicts that these effects should diminish or disappear with the longer SOA used in Experiment 2. The latter prediction is consistent with our hypothesis that the effects in Experiment 1 were perceptual in nature. As such, we predicted that the differences observed in Experiment 1 would be diminished or eliminated in Experiment 2 (that is, we predicted RI – RN = UI – UN = 0).

Method and Materials

Ten UCLA undergraduates participated to fulfill a requirement for a psychology course. These participants were from the same subject pool as those in Experiment 1, but were not the same individuals. In Experiment 2, target objects were presented after the distractors with a 250 ms SOA. Otherwise, the methods and materials used in Experiment 2 were identical to those of Experiment 1.

Results

Data from Experiment 2 are presented in Figure 4. Participants generally performed slightly better in Experiment 2 than in Experiment 1, suggesting that the task was easier with a longer SOA. There was a significant main effect of Distractor-Target Semantic Relatedness on detection accuracy. Mean d' was higher on Related trials (mean d' = 3.14) than on Unrelated trials (2.89) ($F[1, 9] = 33.774$, $MSE = 0.061$, $p < 0.05$). There was also a main effect of Functional Interaction: mean d' was lower on Interacting trials (2.75) than on Non-Interacting trials (3.29) ($F[1, 9] = 39.386$, $MSE = 0.129$, $p < 0.05$). In addition, there was a significant interaction between Distractor-Target Semantic Relatedness and Functional Interaction ($F[1, 9] = 13.617$, $MSE = 0.250$, $p < 0.05$).

Analyses of simple main effects indicated that the mean d' in the RI condition (3.07), was not different from that of the RN condition (3.20) ($t[9] = 0.657$, $SE = 0.199$, $p > 0.50$). However, mean d' was significantly lower in the UI condition (2.42) than in the UN condition (3.37), ($t[9] = 6.806$, $SE = 0.191$, $p < 0.05$), as in Experiment 1.

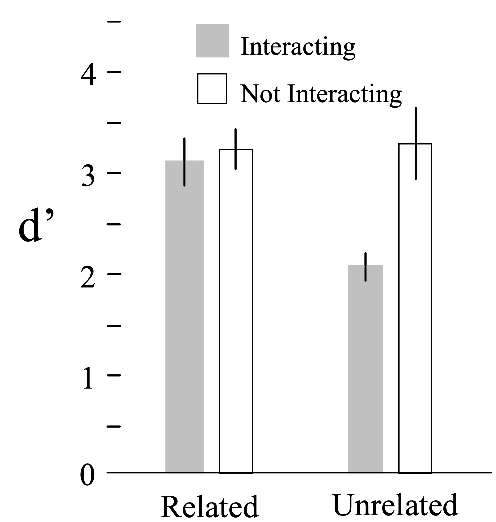

Figure 4: Accuracy data from Experiment 2.

Discussion

An important qualitative change in the data pattern resulted from the increased SOA in Experiment 2: the facilitatory effect of familiar interacting pairs (the advantage for RI over RN trials) observed at the 100 ms SOA did not persist at the 250 ms SOA. This finding suggested that the facilitation in Experiment 1 was a perceptual effect. By contrast, the difference between the UI and UN conditions observed in Experiment 1 also obtained in Experiment 2, so it is not possible to rule out a post-perceptual effect in this case.5

General Discussion

Experiment 1 demonstrated that both the semantics of objects and their arrangement influence object identification, and that these factors interact. When distractor objects were semantically related to the label, identification was better when the target and distractor were arranged to work together than when they were not arranged to work together (a facilitatory effect). When distractor objects were unrelated to the label, arranging the target and distractor to work together made identification worse than when they did not work together (an impairment effect). These result support the conclusion that knowledge about object functions can influence object identification.

Experiment 2 sought to establish whether the effects observed in Experiment 1 were perceptual or post-perceptual. An extended SOA between distractor and target objects was predicted to eliminate the effects observed in the prior experiment. As expected, the facilitatory effect (RI > RN) disappeared in Experiment 2, suggesting a perceptual basis for this effect. However, the impairment effect (UI < UN) remained in Experiment 2, suggesting that it had a post-perceptual component. In addition, the fact that the facilitatory effect disappeared while the impairment effect remained suggests that these effects may have separate causes and thus warrant further investigation independently.

Jointly, Experiments 1 and 2 provide evidence that familiar interacting object pairs are better processed than familiar, non-interacting object pairs and that this advantage is perceptual. There are at least two (not mutually exclusive) explanations for these effects. It could be that functional relations affect perceptual processing by directing visual attention (i.e., cuing) in the direction of the distractor's function. Alternatively, familiar functional interactions may serve to group or integrate the interacting objects into larger perceptual units, which might reduce competition for selection between the objects, or perhaps protect them from decay and interference.

To some degree, the high error rates on UI trials lead us to suspect that attentional cuing did not underlie the advantage for familiar interacting object pairs. If observers were cued by the distractor on all trials, then performance in the UI condition should have been better than observed (as the target object appeared in the "cued" location). A perceptual grouping account is more consistent with our findings. However the hypothesis that distractor objects serve to orient attention to the location of target objects should be tested directly in future experiments.

Green and Hummel (2004a) suggested that scene comprehension is based on representations that incorporate the general semantics or features of visual objects, as well as meaningful spatial relations between them. The experiments reported here demonstrate empirically that these factors (object semantics and object relations) do interact during object identification (at least, when object identification is made difficult by brief exposure).

An important theoretical implication of these results concerns the nature of the perceptual-cognitive interface, and the ability of learned knowledge to influence perception. While these data do not speak to the question of whether or not the actual percept generated from a visual scene is affected by knowledge about object semantics and functional interactions, the data do provide evidence that visual attention and perceptual grouping processes are influenced by such knowledge. Similar effects exist elsewhere in cognitive science, a notable instance being the word superiority effect. Letters are better identified when they are presented as part of a familiar word than when they are presented within a nonsense string, or alone. At least one account of the word superiority effect attributes this difference to the existence of word-level mental representations that are selectively activated by the presence of familiar groupings of letters (words) (Johnston & McClelland, 1980; Johnston, 1981). Functional group representations might play a role in the perception of individual objects similar to that of word representations in letter identification. That is, there may be a "scene superiority effect" for scenes that depict objects engaged in familiar interactions.

Riddoch et al. (2003) concluded that action-based representations serve to reduce competition for selection among visual objects. Access to familiar functional (or action-related) object group representations that enable simultaneous selection of multiple objects might underlie observers' advantage for identifying target objects that are part of a familiar functional group rather than a familiar but non-interacting group.

Conclusions

Our conclusions suggest an important role for functional information in the processing of visual scenes. Problems of scene categorization (e.g. "Is this an office?") can be recast as problems of category definition (e.g. "What is an office [generally]?"). We propose that definitions (mental representations) of scene categories include functional information, and that this is one source of their flexibility. Scenes from a single category might produce very different visual inputs, and the representations that connect visual information to knowledge about scene categories should allow for such variation. Functional group representations provide a means by which visual information can be connected to abstract knowledge about scene categories as well as actions and goals relevant to the environment. In addition, as our results suggest, knowledge about the typical functional organization of the environment could assist in the processing of action-relevant objects and object groupings.

These results suggest that research on scene perception (especially the role of context in scene perception) must account for both the semantic and relational context in which a target object is identified. In the past, researchers have manipulated the identity or location of objects in a scene without considering the creation and disruption of meaningful relations between objects (e.g. Henderson, Weeks, & Hollingworth, 1999; Hollingworth & Henderson, 2000; Loftus & Mackworth, 1978; Mackworth & Morandi, 1967; Moores, Laiti, & Chelazzi, 2003). Other experiments have differentiated between "meaningful" and "non-meaningful" changes to scenes (e.g., Werner & Thies, 2000). We suggest that "meaningful" changes are those that create or disrupt familiar functional groupings of objects. The results presented here (notably, that functional interactions can be beneficial or detrimental to object identification, depending on object semantics) suggest that changes in functional groupings of objects must be considered in addition to (or as a component of) changes in overall scene context.

In summary, object detection in visual scenes cannot be understood solely in terms of object semantics, nor solely in terms of object relations (layout). Associations between objects and the spatial arrangement of objects both influence the processing objects. Perceptual and attentional grouping processes are affected by observers' knowledge about the uses of object groupings within a scene, and these effects are not restricted to objects or groupings that are expected or goal-relevant. Additional research is needed to clarify the role of functional information in the natural viewing of more complex stimuli. The present results highlight the need for consideration of both object semantics and relations as they jointly pertain to actions and goals relevant to observers in real environments.

Acknowledgments

The research reported in this article was supported by National Research Service Award F31-NS43892-02 to C.G. from the National Institute of Neurological Disorder and Stroke, as well as a grant to J.E.H. from the UCLA Academic Senate. Part of this project was completed while the first author held a National Research Council Research Associateship Award at NASA Ames Research Center.

The authors wish to thank Irv Biederman, Robert Bjork, Steve Engel, Keith Holyoak, Zili Liu, Tom Wickens, Brian Stankiewicz, and Eric Ruthruff for their insights regarding this work. Also, thanks to Erica Weiss and Jerlyn Tolentino for their excellent work in the lab. Finally, we thank Alex Doumas, Derek Devnich, Niki Kittur, Don Kalar, and the CogFog group for support and helpful discussion.

References

di Lollo, V., Hogben, J.H., & Dixon, P. (1994). Temporal integration and segregation of brief visual stimuli: Patterns of correlation in time. *Perception & Psychophysics*, 55(4), 373-386.

Green, C. & Hummel, J.E. (2004a). 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.

Green, C. & Hummel, J.E. (2004b). Functional interactions 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*. Mahwah, NJ: Erlbaum.

Humphreys, G.W., Riddoch, M.J., Forti, S., & Ackroyd, K. (2004). Action influences spatial perception: Neuropsychological evidence. *Visual Cognition*, 11(2/3), 401-427.

Johnston, J.C. & McClelland, J.L. (1980). Experimental tests of a hierarchical model of word identification. *Journal of Verbal Learning and Verbal Behavior*, 19, 503-524.

Johnston, J. C. (1981). Understanding word perception: Clues from studying the word-superiority effect. In O. Tzeng & H. Singer (Eds.), *Perception of Print: Reading Research in Experimental Psychology*. Hillsdale, NJ: Erlbaum.

Mandler, J. M., & Parker, R. E. (1976). Memory for Descriptive and Spatial Information in Complex Pictures. *Journal of Experimental Psychology: Human Learning & Memory*, 2(1), 38-48.

Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, 3(4), 386-396.

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.