VISUAL COGNITION, 2010, 18 (1), 110–159

Psychology Press
Taylor & Francis Group

Object Perception, Attention, and Memory (OPAM) 2009 Conference Report 17th Annual Meeting, Boston, MA, USA

Organized by Joseph Brooks, Artem Belopolsky, Michi Matsukura, and Melanie Palomares

http://www.psypress.com/viscog DOI: 10.1080/13506280903314433

Variations in the magnitude of attentional capture: Testing a two-process model

Brian A. Anderson and Charles L. Folk
Villanova University, Radnor, PA, USA

One relatively neglected way in which the results from attentional capture experiments have tended to diverge is in terms of the magnitude of capture effects. In all of the major attentional capture paradigms, the capture of attention is measured in terms of a continuous metric (response time or accuracy), but typically defined in terms of a significant difference between discrete conditions. A recent review of the attentional capture literature demonstrates that the many documented instances of captured attention vary considerably in terms of the magnitude of measures used (Burnham, 2007).

In attempting to account for such large variations in the magnitude of capture effects, there are two related issues to consider. First, there is the issue of what kinds of factors influence the magnitude of attentional capture. According to contingency-based models (e.g., Folk, Remington, & Johnston, 1992), one might expect the magnitude of capture to vary with the similarity between the defining features of the eliciting stimulus and the current top-down set of the observer. Indeed, there is some evidence for such similarity effects in attentional capture. For example, Ansorge and Heumann (2003) found significant cue validity effects with cues that were similar to, but did not exactly match, the target colour. One purpose of the reported study was to explore such colour similarity effects systematically.

We tested whether systematic manipulation of the similarity between cue and target in the modified spatial cueing paradigm used by Folk and Remington (1998) would yield systematic variation in the magnitude of attentional capture as measured by the size of cue validity effects (see Figure 1a). A target defined by one of two possible colours (red or green)

Please address all correspondence to Brian A. Anderson, Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218, USA. E-mail: bander33@jhu.edu

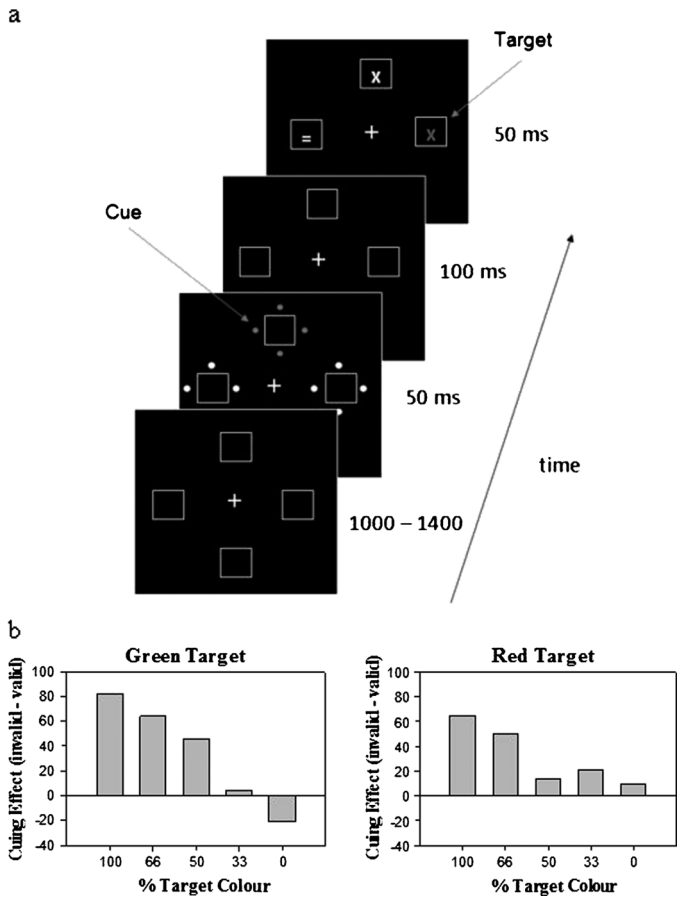

Figure 1. (a) Sequence of events and time course of a trial. Target colour was manipulated between participants, and cue colour within participants. (b) Mean cueing effects (in ms) as a function of percentage of target colour contained in the cue for the green and red target conditions.

was preceded by uninformative peripheral cues that varied in similarity to the target in terms of the percentage of the two colours (100% red/0% green, 66% red/33% green, 50% red/50% green, 33% red/66% green, and 0% red/100% green). Systematic variation was assessed through a linear trend analysis on cue validity effects, which revealed a significant trend for both the green and red target colour conditions, $F(1, 45) = 58.160$, $p < .001$; $F(1, 45) = 38.766$, $p < .001$, respectively (see Figure 1b).

A second issue to consider is the nature of the mechanism by which factors such as cue–target similarity result in variations in the magnitude of attentional capture. One possibility is that these factors produce continuous variations in the amount of resources allocated to the eliciting stimulus, such

that some stimuli receive more attentional processing than others and thus produce a larger magnitude of capture. In this sense, different degrees of the same process are occurring consistently over trials. This can be termed a *continuous* model of attentional capture. Another possibility is that the process by which attention is captured is "all-or-none" in that either all available resources are allocated or none are allocated, but the frequency (i.e., the number of trials) with which the process occurs can vary with changes in similarity. That is, variations in the magnitude of attentional capture might reflect various combinations of trials on which attention is captured and trials on which attention is diffuse and not captured, resulting in *mixture distributions.* This can be termed a *two-process* model of attention allocation (Jonides, 1983). A second purpose of the reported study was to distinguish between continuous and two-process accounts.

We tested whether the RT distributions specific to the mixed colour cues reflect a mixture of trials on which attention is fully captured and trials on which attention is not captured at all (i.e., a two-process model). Research on contingent attentional capture (CAC) suggests a means of identifying distributions specific to fully captured and uncaptured attention. The CAC hypothesis posits that a cue matching the target on a defining property produces the optimal circumstance for attention to be captured; likewise, cues whose defining feature value is orthogonal to the defining feature of the target should produce no capture at all. The results from our first experiment were replicated using only a single mixed colour cue (50%/50%) and a single target (green), and a mixture analysis was conducted on the obtained RT distributions for invalid trials.

A mixture analysis generates an intermediate distribution predicted by a best-fitting combination of samplings from two distributions specific to two discrete states, in the present case maximally captured and uncaptured attention (provided by green and red cue trials, respectively), and compares the statistical properties of this predicted distribution to those of an obtained intermediate distribution (provided by mixed cue trials). When the variance of the obtained intermediate RT distribution (11,073 ms) was compared to the variance predicted by the best-fitting mixture of the RT trials specific to maximally captured and uncaptured attention (12,374 ms) using a paired samples *t*-test, a significant difference emerged, $t(31) = -2.422$, $p = .021$. Thus, the data were found to be inconsistent with a two-process model and the mixture distributions that it predicts.

With respect to the influence of cue–target similarity, our results provide strong evidence that the magnitude of attentional capture varies systematically as a function of the percentage of the target colour present in the cue. Although previous studies have shown that cues similar to a target can produce attentional capture (e.g., Ansorge & Heumann, 2003), ours is the first to demonstrate that the magnitude of attentional capture, as measured by cue

validity effects, varies systematically as a function of variations in similarity. Importantly, this type of systematic variation also provides the necessary conditions for assessing the presence of a mixture distribution, allowing the opportunity to test whether or not the two-process model of attention allocation applies to instances of attentional capture. A mixture analysis applied to the results of the second experiment found that the obtained variance for the intermediate distributions was significantly less than that predicted by a mixture of capture and no capture trials, disconfirming a two-process model.

REFERENCES

Ansorge, U., & Heumann, M. (2003). Top-down contingencies in peripheral cuing: The roles of color and location. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 937–948.

Burnham, B. R. (2007). Displaywide visual features associated with a search display's appearance can mediate attentional capture. *Psychonomic Bulletin and Review*, *14*, 392–422.

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 847–858.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030–1044.

Jonides, J. (1983). Further toward a model of the mind's eye's movement. *Bulletin of the Psychonomic Society*, *21*, 247–250.

Ensemble statistics of a display influence the representation of items in visual working memory

Timothy F. Brady

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

George A. Alvarez

Department of Psychology, Harvard University, Cambridge, MA

In just a glance, observers extract a great deal of perceptual and semantic information from a real-world scene (Oliva, 2005). When later recalling the

Please address all correspondence to Timothy Brady, Department of Brain and Cognitive Sciences, 46-4078, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. E-mail: tfbrady@mit.edu

details of the scene, they are influenced by this gist, tending to remember objects that are consistent with the scene but were not in fact present (Lampinen, Copeland, & Neuschatz, 2001). Thus, memory for individual items is influenced by memory for the structure of the entire scene: The gist creates expectations about what objects were present and where they were located.

Most studies of visual working memory use randomized displays that are, as best as possible, prevented from having any overarching structure or gist (Luck & Vogel, 1997). Even in displays with simple coloured circles, however, there are ensemble statistics such as the mean size that can be computed quickly and easily by observers (Ariely, 2001). Such information, when task relevant, could allow observers to encode the displays more efficiently by providing expectations that reduce the uncertainty in memory for individual items. In fact, because of their simplicity, these displays provide an interesting test case for examining when and how observers make use of such hierarchical constraints in short-term memory.

We thus sought to examine whether the ensemble statistics of a display would guide memory for individual items in a task involving memory for the size of coloured circles. We hypothesized that on displays with several different colours of circles, observers would be biased in reporting the size of a given circle by the size of the other circles of that colour.

METHOD

Six observers were presented with 400 individual displays consisting of three red, three blue, and three green circles of varying size and told to remember the size of all of the red and blue circles, but to ignore the green circles. Each display appeared for 1.5 s, followed by a 1 s blank, after which a single circle reappeared in black at the location that a red or blue dot had occupied. Observers had to slide the mouse up or down to resize this new black circle to the size of the red or blue dot they had previously seen, and then click to lock in their answer and start the next trial. Green distractor items were present in the display because we believed they would encourage observers to encode the items by colour, rather than selecting all of the items into memory at once (Halberda, Sires, & Feigenson, 2006; Huang, Treisman, & Pashler, 2007).

The sizes of the circles were drawn from separate normal distributions for each colour, each with a mean diameter chosen uniformly on each trial from the interval [0.625°,3.125°] (degrees visual angle) and with standard deviation equal to one-eighth of their mean. Thus, on a given trial, the three red dots could be sampled from around 1°, the blue dots from 2.5° and the green dots from 0.6°. However, on the next trial it could be the red dots that were largest and the blue dots smallest; thus, long-term memory for the sizes of the dots could not be responsible for any results.

To allow a direct test of the hypothesis of bias towards the mean, the displays were generated in matched pairs: 200 displays were generated as described; another 200 were created by swapping the colour of the to-be-tested item with a dot of the other nondistractor colour (either red or blue). These 400 displays were then randomly interwoven, with the constraint that paired displays could not appear one after the other. This resulted in 200 pairs of displays, each matched in the size of all of the circles present, with a difference only in the colour of the circle that would later be tested. By comparing the size reported by observers when the tested circle was in one colour with the size reported when it was in another, we were able to directly test the hypothesis that observers would be biased towards the mean of the items in the same colour as the tested item.

RESULTS

Observers' performance was quite good, with an error of 0.5° on average (SEM: 0.06°), significantly less than our empirical measure of chance ($p = .00005$; empirical chance: 0.825°; SEM: 0.04°).

Next, we examined whether observers tended to be biased towards the size of the same-coloured dots. For each matched pair of trials, we selected the display in which the mean of all of the other dots of the same colour as the tested item was smallest and used this as a baseline. If observers were not biased, the ratio between the size observers reported on these displays and the size on the other half of the displays should be 1.0; they should be equally likely to report a larger or smaller size. However, if observers are biased toward the mean size of the same coloured dots, this ratio should be greater than 1.0 (Figure 1).

Observers reported a size on average 1.2 times greater (SEM: $+/-$ 0.05) on the half of the displays with larger same-coloured dots (the dots were on average 1.4 times larger in these displays than their matched counterparts). This ratio was significantly greater than 1.0, $t(5) = 3.76$, $p = .01$, indicating that observers reported a larger size on the display in which the other items of the same colour were larger, despite the fact that all of the dots—including the tested item—were exactly the same size.

CONCLUSION

We find that observers are biased by the ensemble statistics of the display when representing items in visual working memory. This suggests that, for displays with higher order structure like real-world scenes, items in visual working memory are not represented in isolation. Instead, observers use the

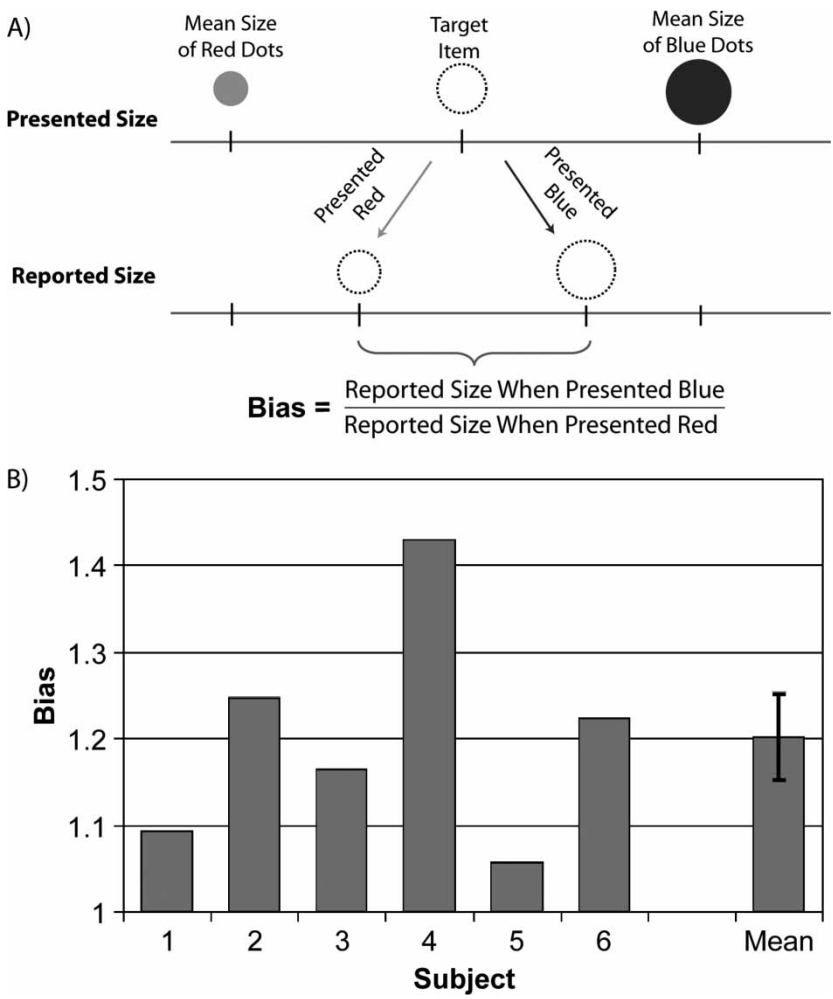

Figure 1. (A) Observers had to remember the size of each red and blue dot. In this example, the blue dots were larger than the red dots. Observers saw each display twice, with the target dot's colour changed for the second presentation. We then measured whether observers reported different sizes for the tested dot when it was red versus when it was blue (the dot was in fact the same size in both presentations). Which colour was larger was counterbalanced across trials in the actual experiment, but bias was always calculated by dividing the size reported for the larger colour by size reported for the smaller colour. (B) The bias for each of our six observers, plus the mean (error bar represents SEM). If observers showed no bias this index would be 1.0; instead, it is significantly greater than 1.0, suggesting a bias to report a larger size when items of the same colour as the tested item were larger in size.

constraints that more abstract information like the gist of the scene or the ensemble size of the set of items provides to reduce their uncertainty about the size of individual items and encode the items more efficiently.

REFERENCES

Ariely, D. (2001). Seeing sets: Representation by statistical properties, *Psychological Science*, *12*(2), 157–162.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially-overlapping sets can be enumerated in parallel. *Psychological Science*, *17*(7), 572–576.

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, *317*(5839), 823–825.

Lampinen, J. M., Copeland, S., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1211–1222.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *The encyclopedia of neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.

Action influences figure–ground assignment

Joshua D. Cosman and Shaun P. Vecera

Departments of Neuroscience and Psychology, University of Iowa, Iowa City, IA, USA

Our ability to segregate objects from one another in a scene is a fundamental perceptual mechanism. Central to this process is figure–ground assignment, or the segregation of candidate objects from their backgrounds. Although hierarchical models of vision (e.g., Julesz, 1984; Pylyshyn, 1999) posit that figure–ground assignment occurs prior to higher level visual processing such as focal attention, a recent study demonstrated that attention can influence which regions of a scene are assigned figural status (Vecera, Flevaris, & Filapek, 2004). As a result, it has been posited that figure–ground segregation is an interactive process, in which both bottom-up and top-down cues compete to bias which items in a scene are perceived as figures and grounds (Vecera et al., 2004; Vecera & O'Reilly, 1998).

Such an interactive account raises the possibility that other high-level factors can influence figure–ground segregation. Specifically, it is possible that *acting* upon an object can increase the likelihood that it will be assigned figural status. In other words, action may act as a cue to figure–ground assignment. Such a possibility is suggested by the presence of populations of bimodal visuotactile neurons that respond exclusively to tactile and visual

Please address all correspondence to Joshua Cosman, Department of Neuroscience, University of Iowa, E305 Seashore Hall, University of Iowa, Iowa City, IA 52242-1407, USA. E-mail: joshua-cosman@uiowa.edu

stimulation in peripersonal space around the hand (di Pelligrino, Làdavas, & Farnè, 1997; Graziano & Gross, 1993). It has been posited that such bimodal representations are responsible for integrating visual and tactile space, supporting the control of reaching/grasping and visual processing of objects near the hand (Làdavas, di Pellegrino, Farnè, & Zeloni, 1998; Reed, Grubb, & Steele, 2006). It is possible that the presence of such bimodal representations may complement the unimodal visual representations provided by early visual processing, and as a result may have the effect of biasing or strengthening perceptual processing of objects near the hand. The focus of the current study was to determine whether such bimodal representations exert an effect on figure–ground assignment, a process typically thought to rely on preattentive, unimodal visual processing.

EXPERIMENT 1

Twenty-four observers performed a figure-memory task in which they viewed ambiguous figure–ground displays and then performed a matching task where they were asked which of two probe regions matched one presented in the initial display (see Vecera et al., 2004; Figure 1a). A visual anchor, either the observer's hand or a wooden dowel (manipulated between subjects), was present in one region of the bipartite display allowing us to compare the effects of actively reaching towards one region of the display with those of simply having a visual anchor present in one region.

Critically, on half of the trials the matching probe region had contained the visual anchor during the presentation of the figure–ground display. If reaching towards an object increases its likelihood of being seen as figure, we would expect faster RTs to matching probes that had contained the hand during presentation of the figure–ground display. Participants were told anchor position would not predict which region would be tested, and were told to focus on the contour at fixation to maximize their ability to recognize the shape of the matching region during the memory task. Eye position was monitored to ensure that observers did not preferentially fixate either region of the display.

Analyses revealed a significant interaction between anchor type (hand vs. dowel) and region probed (anchor present vs. anchor absent), $F(1, 22) = 5.9$, $p = .03$. Planned comparisons showed that the interaction was driven by significantly faster reaction times in the "hand" anchor condition when the probe matched the region containing the observer's hand, $t(11) = 2.7$, $p = .02$, whereas there was no such difference in the dowel anchor condition, $t(11) = 0.38$, $p = .90$ (see Figure 1c). Neither main effect was significant, all Fs < 1.8, ps $< .22$. Analyses of accuracy data revealed no significant main effects or interactions, all Fs < 1, ps $> .73$. Thus, it appears that when

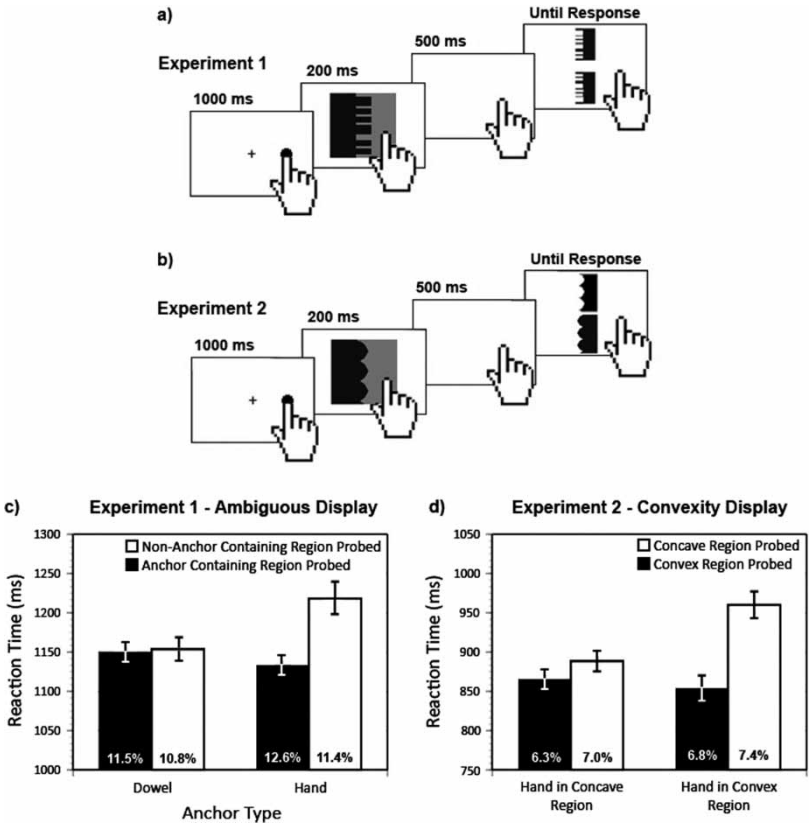

Figure 1. Task sequence and reaction time data. (a) Trial sequence for Experiment 1. (b) Trial sequence for Experiment 2. (c) Mean reaction times to probe trials in Experiment 1. (d) Mean reaction times to probe trials in Experiment 2. Error bars represent 95% confidence intervals, and error rates are present in white at the base of each bar.

image-based cues to figure–ground assignment are absent, the presence of an observer's outstretched hand influences figure–ground assignment.

EXPERIMENT 2

A follow-up experiment was performed with 12 new participants to examine whether the position of an observer's hand can compete with image-based, configural cues to figure–ground assignment when both are present in a scene. This experiment was identical to Experiment 1, but instead of using ambiguous figure–ground displays we used displays with strong convexity cues known to influence figural assignment (Kanisza & Gerbino, 1976; see Figure 1b). Typically, when observers view such displays they are more likely

to see the convex region as figure. Participants were asked to place their outstretched hand into either the convex figure or the concave ground on a given trial, allowing us to see whether the presence of a hand in a region could compete with a strong image-based cue to figure–ground assignment.

Analyses revealed a significant main effect of region probed (convex vs. concave), $F(1, 11) = 14.9$, $p = .001$, with participants responding faster overall to convex probes. Critically, we also observed a significant interaction between hand region (convex vs. concave) and region probed, $F(1, 11) = 4.9$, $p = .04$. As can be seen in Figure 1b, the difference in RTs to convex and concave probes was reduced when the hand was placed in the concave region of the display, indicating that this region was better able to compete for figural status when an observer's hand was located within it. The main effect of hand region was not significant, $F(1, 11) = 3.2$, $p = .10$ (Figure 1d). Analyses revealed no significant main effects or interactions in the accuracy data, Fs < 1, ps $> .65$.

DISCUSSION

Our results suggest that bimodal neural representations can act as a cue to figure–ground assignment, interacting with image-based cues to bias the assignment of figural status to regions near the hand. The fact that objects of action are more likely to be perceived as figures suggests that other early visual processes may also be influenced by action, and that neural systems involved in the bimodal visuotactile representation of scenes can exert effects on unimodal perceptual processing.

REFERENCES

Di Pellegrino, G., Làdavas, E., & Farnè, A. (1997). Seeing where your hands are. *Nature*, *388*, 730.

Graziano, M. S., & Gross, C. G. (1993). A bimodal map of space: Tactile receptive fields in the macaque putamen with corresponding visual receptive fields. *Experimental Brain Research*, *97*, 96–109.

Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neurosciences*, *7*(2), 41–45.

Kanisza, G., & Gerbino, W. (1976). Convexity and symmetry in figure-ground organization. In M. Henle (Ed.), *Vision and artifact* (pp. 25–32). New York: Springer.

Làdavas, E., di Pellegrino, G., Farnè, A., & Zeloni, G. (1998). Neuropsychological evidence of an integrated visuotactile representation of peripersonal space in humans. *Journal of Cognitive Neuroscience*, *10*, 581–589.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 341–423.

Reed, C. L., Grubb, J. D., & Steele, C. (2006). Hands up: Attentional prioritization of space near the hand. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 166–177.

Vecera, S. P., Flevaris, A. V., & Filapek, J. C. (2004). Exogenous spatial attention influences figure-ground assignment. *Psychological Science*, *15*, 20–26.

Vecera, S. P., & O'Reilly, R. C. (1998). Figure-ground organization and object recognition processes: An interactive account. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 441–462.

Controlling stimulus variability reveals stronger face-selective responses near the average face

Nicolas Davidenko and Kalanit Grill-Spector

Department of Psychology, Stanford University, Stanford, CA, USA

The past decade of fMRI research has identified face-selective regions in the human ventral stream that respond more strongly when people observe faces than other objects and are thought to be critically involved in face perception and recognition (Grill-Spector, Knouf, & Kanwisher, 2004; Kanwisher, McDermott, & Chun, 1997). However, the underlying neural representations that subserve humans' remarkable ability to recognize thousands of individual faces are not well understood. A basic question is whether responses in face-selective regions increase or decrease as faces deviate from the average face. In one view, face-selective neural responses are anchored on the average (or mean) face, suggesting responses should increase as faces deviate from the mean face in particular directions (or angles) away from the mean (Leopold, Bondar, & Giese, 2006; Loffler, Yourganov, Wilkinson, & Wilson, 2005). An alternative view posits that neurons are tuned to particular stored exemplar faces, and responses decrease as faces deviate from the preferred face exemplar. Because the distribution of faces is thought to be centrally dense, the latter view predicts higher responses near the mean face. Electrophysiological and fMRI research shows that responses are reduced, or adapted (Grill-Spector et al., 1999; Li, Miller, & Desimone, 1993) to repetitions of similar faces, and thus assessing the strength of responses to faces blocked by their distance from the mean requires the control of stimulus variability within each block. Here, we use a parameterized space of face silhouettes (Davidenko, 2007) and high-resolution fMRI (HR-fMRI) to measure responses in face- and object-selective regions as we manipulate distance from the mean face and control in two ways the variability of stimuli at each distance from the mean.

Please address all correspondence to Nicolas Davidenko, Department of Psychology, 450 Serra Mall, Building 420, Stanford, CA 94107, USA. E-mail: ndaviden@stanford.edu

METHODS AND RESULTS

Stimuli

We defined blocks of parameterized face silhouettes at five different distances from the mean face in silhouette face space (Davidenko, 2007). In the "matched angular variability" (MAV) condition, we matched the number of sampled directions (or angles) sampled in each block of faces (Figure 1a). In the "matched physical variability" (MPV) condition, we matched the physical similarity of faces in each block (Figure 1b).

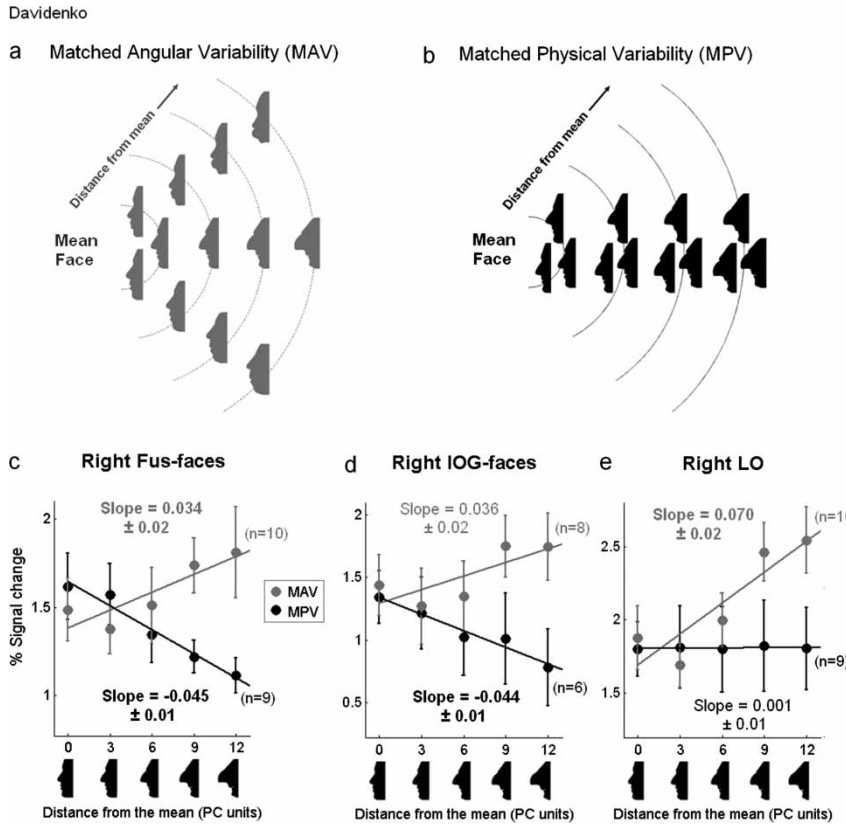

Figure 1. (a–b) Stimulus design for the MAV and MPV conditions. Arcs represent fixed distances from the mean. (c–e) Responses in face- (c–d) and object-selective (e) regions to blocks of face silhouettes in the MAV (grey) and MPV (black) conditions. Percentage signal change is versus fixation + SEM across subjects.

Behavioural measures

We assessed two behavioural measures to determine whether matching angular or physical variability resulted in matched perceptual variability among the stimuli in each block. First, we measured perceptual discrimination (d') from performance on a one-back task where subjects responded to infrequently repeating face stimuli. By examining performance across all blocks, we found d' was correlated more with physical variability ($r = .94$) than with angular variability ($r = .66$). Second, we obtained subjects' judgements of dissimilarity (on a 1–7 scale, with 1 = "identical" and 7 = "maximally dissimilar") between pairs of face silhouettes sampled from within each block. We found that dissimilarity judgements were also more closely coupled to physical ($r = .96$) than angular ($r = .69$) variability. These results support the use of physical variability as a proxy for perceptual variability.

fMRI measures

We scanned 12 subjects at high resolution (1.5 mm isotropic voxels) as they observed blocks of face silhouettes in the MAV (10 subjects) and MPV (9 subjects, 7 overlapping) conditions. We measured responses in two independently localized face-selective regions (Fus-faces and IOG-faces) and one object-selective region (LO) as we manipulated distance from the mean face in the MAV and MPV conditions.

We found that responses as a function of distance from the mean face differed drastically across the two conditions (Figure 1c–e; significant two-way ANOVA interaction between distance from the mean and condition, all Fs > 5.0, $p < .01$). In the MAV condition (where angular variability was matched but physical variability increased with distance from the mean), responses in face-selective regions increased (mean slopes = 0.034 and 0.036, in Fus-faces and IOG-faces, respectively; Figure 1c–d, grey). In the MPV condition (where physical variability was matched but angular variability decreased with distance from the mean), responses in Fus-faces and IOG-faces decreased with distance from the mean (mean slopes = -0.045 and -0.044, respectively; Figure 1c–d, black). In contrast, responses in LO increased in the MAV condition but were constant in the MPV condition across distances from the mean face (Figure 1e).

To determine how the three factors—distance from the mean, physical variability, and angular variability—contributed to responses across all conditions, we conducted a step-wise multiple regression analysis on mean responses across subjects in each block of faces. For responses in face-selective regions, physical variability was a significant positive factor

(associated with increased responses) and distance from the mean was a significant negative factor (associated with decreasing responses), together explaining 85% and 93% of the variance in Fus-faces and IOG-faces responses, respectively. Angular variability did not explain any additional variance. LO responses were highly correlated with physical variability, which explained 89% of their variance, whereas other factors were not significant.

DISCUSSION

Our data provide evidence that (1) physical variability drives responses across face- and object-selective regions, and (2) when this factor is controlled, responses in face-selective regions are strongest near the mean face. In contrast, responses in object-selective LO are not modulated by distance from the mean face when physical variability is controlled, suggesting that this is a face-specific effect. We suggest that previous studies that found increasing responses in face-selective regions as a function of distance from the mean (Leopold et al., 2006; Loffler et al., 2005) likely confounded physical variability with distance from the mean.

Stronger responses to faces near the mean face may reflect sensitivity to the distribution of experienced faces. After years perceiving and encoding faces likely drawn from a centrally dense distribution (Valentine, 1991), face-selective neurons may become tuned to best represent this distribution. As a result, faces near the mean may activate more face-selective neurons, and in turn elicit a larger BOLD response, than faces far from the mean. This interpretation is consistent with an exemplar-based neural face space (see Jiang et al., 2006) where responses are strongest to the frequently experienced faces near the mean face.

REFERENCES

Davidenko, N. (2007). Silhouetted face profiles: A new methodology for face perception research. *Journal of Vision*, *7*(4), 1–17.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–562.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*, 187–203.

Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using FMRI and behavioral techniques. *Neuron*, *50*, 159–172.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*(7102), 572–575.

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, *69*(6), 1918–1929.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience*, *8*(10), 1386–1390.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *43A*(2), 161–204.

Target enhancement and distractor suppression in multiple object tracking

Matthew M. Doran and James E. Hoffman

University of Delaware, Newark, DE, USA

In multiple object tracking (MOT), observers keep track of target objects that move haphazardly around a display in the presence of identical distractors. The typical result from this paradigm is that observers can accurately track up to about four objects, with performance declining precipitously beyond this number. However, recent evidence indicates that the number of objects that can be effectively tracked is not fixed but depends on factors such as speed and interobject distance (Shim, Alvarez, & Jiang, 2008).

Decreasing interobject distance reduces tracking performance, which is compatible with the idea that visual attention may be particularly important in MOT in order to maintain individuation of target objects in the face of nearby distractors. Previous research has shown that one source of errors during MOT arises when observers mistakenly begin tracking distractors that pass close to targets (O'Hearn, Landau, & Hoffman, 2005; Pylyshyn, 2004). Therefore, a sensible strategy might be to suppress or inhibit distractors that pass close to and are confusable with targets. Consistent with this idea, Pylyshyn, Haladjian, King, and Reilly (2008) reported that probes appearing on distractors that were located in a different depth plane than the tracked objects were detected more frequently than same-depth plane distractor probes. According to them, objects in different depth planes are preattentively segregated, allowing observers to easily ignore different-depth plane distractors without the need to actively suppress them. This is consistent with

Please address all correspondence to Matthew M. Doran, Department of Psychology, University of Delaware, Newark, Delaware 19716, USA. E-mail: mdoran@psych.udel.edu

the claim that suppression of nearby distractors by visual attention may be useful in preventing nearby distractors from being mistaken for targets.

In the present experiment we attempted to provide converging evidence for the role of inhibition in MOT by examining the N1 component of the event-related brain potential elicited by probes appearing on targets and distractors. Displays were arranged so that some distractors on each trial were relatively close to tracked targets while others were further away (Figure 1A). During each trial, irrelevant probe flashes appeared intermittently on targets, nearby distractors, and far distractors. On average, near distractor probes were approximately 2.5° ($SD = 0.9°$) from targets compared to 10.6° ($SD = 1.7°$) for far distractor probes. Mean eccentricities were similar across conditions: Target, $M = 7.4°$, $SD = 1.3°$; near distractor, $M = 7.5°$, $SD = 1.3°$; far distractor, $M = 7.5°$, $SD = 1.2°$. If distractors that are close to targets are inhibited, probe flashes appearing on them should result in a smaller N1 component compared to probe flashes appearing on targets. However, this pattern is ambiguous as it could also result from enhancement of the target without suppression of the distractor. Probes on far distractor objects can serve as a "neutral baseline" for disambiguating these results since far distractors should not be confusable with target objects and should not, therefore, require suppression. "Pure suppression" of nearby distractors should result in equivalent N1s for probes appearing on targets and far distractors together with smaller N1s for near distractor probes. In contrast, the signature of "pure target enhancement" would be a large target N1 coupled with small and equivalent N1s for distractors regardless of distance.

Two separate N1 components were observed in this experiment. The *posterior* N1, thought to be generated in the lateral occipital complex (LOC; Di Russo Martinez, & Hillyard, 2003; Martinez, Ramanathan, Foxe, Javitt, & Hillyard, 2007), occurred over lateral posterior electrode locations and peaked approximately 175 ms poststimulus for contralateral stimuli (Figure 1B). Comparable posterior N1 components were observed for targets and far distractors and they were both larger than N1s for near distractors (Figure 1B and C), $F(2, 22) = 5.16$, $p < .05$ (see Figure 1C for p-values based on pairwise t-tests). This pattern is consistent with suppression of nearby distractors. The *anterior* N1, thought to be generated in superior parietal areas near the intraparietal sulcus (IPS; Di Russo et al., 2003), occurred over frontocentral electrode locations and peaked approximately 155 ms post-stimulus (Figure 1D). In contrast to the suppression pattern observed in the posterior N1, the anterior N1 showed a pattern consistent with pure enhancement of target objects. N1 amplitude was larger for targets than both near and far distractors which did not differ (Figure 1D and E), $F(2, 22) = 4.96$, $p < .05$ (see Figure 1E for p-values based on pairwise t-tests). Apparently, visual attention can act at different levels of the visual system to both enhance target objects and suppress distractors during MOT.

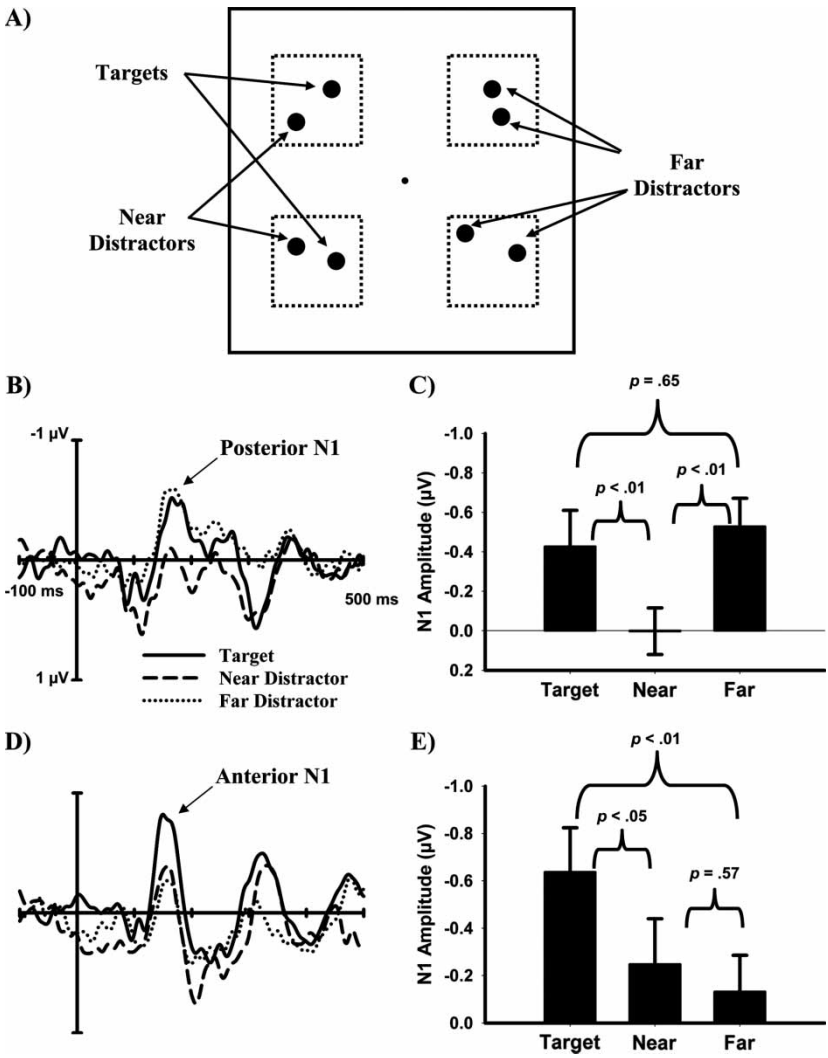

Figure 1. (A) Schematic of the MOT displays in this experiment. Objects were positioned so that two objects were located in each quadrant. During the motion phase of each trial, the objects reflected off of the boundaries of invisible "containers" (illustrated by the dotted squares) so that they remained in separate quadrants. Participants tracked two target objects that were positioned in adjacent quadrants such that there were always near distractors (i.e., distractors within the same quadrant as a target) and far distractors (i.e., distractors in the quadrants without targets). Probes were presented equally often on targets, near distractors, and far distractors. Note that the dotted boxes are for illustration purposes and were not visible to the observers. (B) ERP waveforms measured at occipitoparietal scalp locations illustrating the posterior N1. (C) Mean posterior N1 amplitude. (D) ERP waveforms measured at frontocentral scalp locations illustrating the anterior N1. (E) Mean anterior N1 amplitude. The posterior N1 shows distractor suppression; the anterior N1 shows only target enhancement.

The finding that target enhancement and distractor suppression appeared in different ERP components associated with different generators in the brain suggests that they are separate and distinct processes. The parietal system associated with the anterior N1 seems to primarily enhance attended locations (He, Fan, Zhou, & Chen, 2004). On the other hand, the extrastriate system associated with the posterior N1 seems to primarily suppress distractors perhaps by means of a winner-take-all competition (Desimone & Duncan, 1995). That suppression was only applied to nearby distractors is consistent with this sort of competition since distractors that share receptive fields with targets would be subject to suppression whereas far away distractors would not.

REFERENCES

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual-attention. *Annual Review of Neuroscience*, *18*, 193–222.

Di Russo, F., Martinez, A., & Hillyard, S. A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cerebral Cortex*, *13*(5), 486–499.

He, X., Fan, S. L., Zhou, K., & Chen, L. (2004). Cue validity and object-based attention. *Journal of Cognitive Neuroscience*, *16*(6), 1085–1097.

Martinez, A., Ramanathan, D. S., Foxe, J. J., Javitt, D. C., & Hillyard, S. A. (2007). The role of spatial attention in the selection of real and illusory objects. *Journal of Neuroscience*, *27*(30), 7963–7973.

O'Hearn, K., Landau, B., & Hoffman, J. E. (2005). Multiple object tracking in people with Williams syndrome and in normally developing children. *Psychological Science*, *16*(11), 905–912.

Pylyshyn, Z., Haladjian, H., King, C., & Reilly, J. (2008). Selective nontarget inhibition in multiple object tracking (MOT). *Visual Cognition*, *16*(8), 1011–1021.

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11*(7), 801–822.

Shim, W. M., Alvarez, G. A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin and Review*, *15*(2), 390–397.

Individual differences in overriding attentional capture

Keisuke Fukuda and Edward K. Vogel

University of Oregon, Eugene, OR, USA

Working memory capacity (WMC) is known to be severely limited, yet variable across individuals. Recent evidence suggests that high WMC

Please address all correspondence to Keisuke Fukuda, 1227 University, Eugene, OR 97405, USA. E-mail: keisukef@uoregon.edu

individuals are better at selectively processing target information only, whereas low WMC individuals cannot help but also process distracting information (Cowan & Moray, 2006; Vogel, McCollough & Machizawa, 2005). However, one fundamental ambiguity remains in understanding the relationship between WMC and attention: Is poor attentional control by low capacity individuals the result of weak "top-down" control for selecting task-relevant items or is it the result of an inability to override involuntary attentional capture from irrelevant items? In this study (Fukuda & Vogel, 2009), we tested between these two alternatives by measuring event-related potentials (ERPs) that are sensitive to where attention is allocated both voluntarily and involuntarily by attentional capture.

After measuring individuals' WMC by the standard change detection procedure (see Awh, Barton, & Vogel 2005), we measured voluntary and involuntary attentional control by recording ERPs from healthy young adults while they performed a spatial attention task. In the attention task, a cue was presented for 500 ms at the beginning of each trial to indicate where the target item would be presented (Figure 1a). The cue consisted of two diamond shapes that contained one red and one green dot at a corner of each diamond. For a half of the subjects, the position of the red dot indicated where the target item would come up, and the green dot did for the other half of them. After 200 ms of blank screen, a target Landolt C came up at the cued location along with three distractors on the target side and four on the other side. On two-thirds of the trials, 100 ms after the target array disappeared, a bilateral task-irrelevant probe (i.e., a filled white square) was flashed either at the target location or at the location of one of the distractors. On the other third of trials, no probe was presented. At the end, subjects reported the orientation of the target item by pressing one of four buttons. No response was required for the probe.

RESULT AND DISCUSSION

To measure the allocation of spatial attention we examined the early visually evoked ERPs (i.e., P1 and N1) that reflect sensory processing in extrastriate cortical areas beginning within the first 75 ms of stimulus onset. These ERP components are known to be modulated by spatial attention, with larger amplitudes observed for stimuli that appear in attended locations as compared to greatly reduced amplitudes for stimuli presented in unattended locations (Hillyard, Vogel, & Luck, 1998; Mangun, Hillyard, & Luck, 1993). In lateralized displays, such as those used here, the P1 attention effect is typically observed as a larger positive voltage at electrodes over the hemisphere that is contralateral to the attended side of the display; whereas the N1 attention effect is often observed as a larger negative voltage at

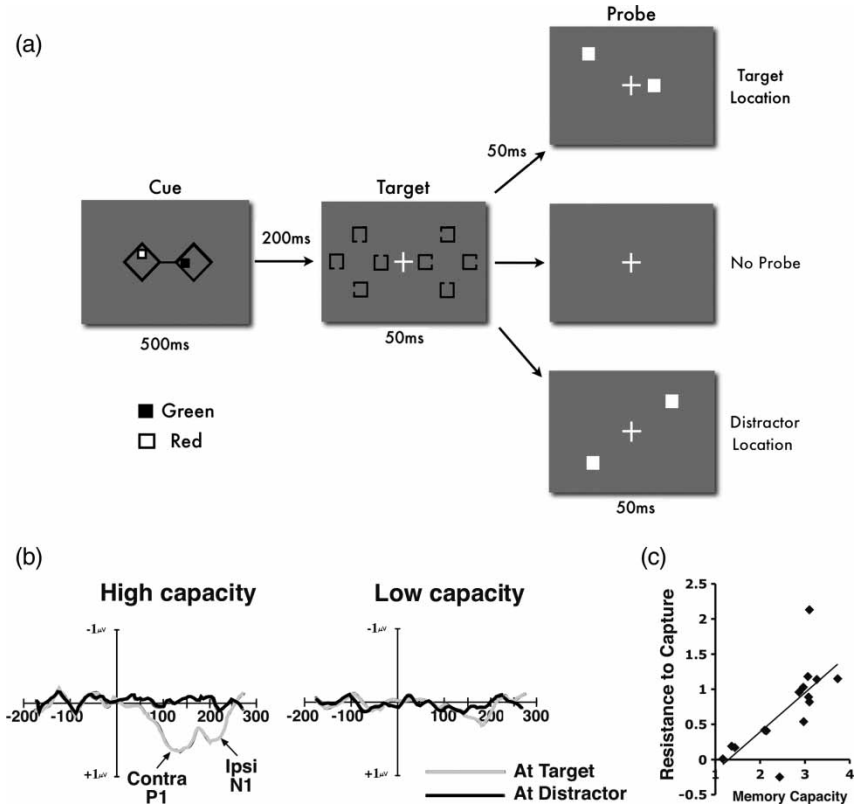

Figure 1. (a) The spatial attention task. 500 ms after participants fixate their vision at the centre of the screen, a cue array that contained two diamonds with a green and a red dot in each was presented for 500 ms. For a half of the participants, the position of the red dot corresponded to the position of the target item, and for the other half, the position of the green dot did. 200 ms after the cue offset, a target array composed of eight Landolt Cs in the configuration shown here was presented for 50 ms. On two thirds of trials, bilateral task irrelevant probes were flashed for 50 ms at either target location or one of the distractor location 100 ms after the onset of the target array. On the other third of trials, there was no probe presented following the target array. At the end of trials, participants were asked to press a button to report where the gap was on the target Landolt C. (b) The P1N1 attention effect evoked by target and distractor probe for high and low working memory capacity individuals. The left panel represents the average waveforms for individuals with higher working memory than the median (median = 2.9), and the right panel shows the average waveforms for those with lower working memory capacity than median. (c) The scatterplot regressing the P1/N1 capture effect against individuals' working memory capacity. As can be seen, there was a strong positive correlation between individual working memory capacity and the P1/N1 capture effect.

electrodes over the hemisphere that is ipsilateral to the attended side (P1/N1 attention effect), we can obtain a highly sensitive measure of where and how well spatial attention is directed at a particular moment during the task (Heinze, Luck, Mangun, & Hillyard, 1990).

We measured both voluntary and involuntary control mechanisms by examining the spatial allocation of attention during two separate moments of task performance. By examining the P1/N1 attention effect to the target array, we could measure how effectively an individual could take information from the cue, voluntarily orient attention towards the target location. It revealed that both high and low WMC individuals showed large P1/N1 attention effect to target array, suggesting that both groups are equally good at voluntarily orienting attention to the cued location, $F < 1$, ns. Thus, we observe no relationship between voluntary control of attention and working memory capacity, $r = .11$, *ns*.

In contrast, the P1/N1 attention effects to the task-irrelevant *probe* that quickly followed the target array provided a measure of how effectively the individual was able to override attentional capture by the distractors. For example, imagine an individual who was able to completely resist against attentional capture by the distractors presented in the target array. When the task-irrelevant probe is presented at the target location, there should be a large P1/N1 attention effect because that location is still attended; however, when the probe is presented at the location of one of the distractors, a negligible P1/N1 response is expected because attention is still focused solely on the target location. Thus, perfect override of attentional capture would produce a large difference between P1/N1 response to target location and to distractor location (P1/N1 capture effect). On the other hand, consider an individual who was incapable of resisting against attentional capture by the distractors surrounding the target and involuntarily reoriented attention to include one or more of the distractor locations. When the probe is presented at the target location, the P1/N1 response may be reduced because the target location may not still be attended exclusively; however, probes presented at the distractor location may show an increased P1/N1 response because one or more of these locations is now attended. Thus, a perfect inability to override capture would produce a small P1/N1 capture effect. Consequently, P1/N1 capture effect provides an estimate of effectiveness at overriding involuntary attentional capture.

As Figure 1b shows, high capacity individuals show a strong attention effect only to target probe ($p < .01$) and not to distractor probe ($p > .25$). On the other hand, low capacity individuals showed much attenuated attentional response to target probe and therefore, we did not observe a significant difference in attentional responses to target and distractor probes ($p > .3$). Furthermore, correlational analysis revealed a highly significant positive correlation between individual WMC and P1/N1 difference, $r = .73$, $p < .001$ (Figure 1c). This supports that low WMC individuals were poorer at resisting against attentional capture.

In conclusion, these results suggest that the poor attentional ability associated with low WMC is not due to the deficit in voluntary orienting of

attention, but rather a consequence of a deficit in adequately overriding involuntary attentional capture from distracting information.

REFERENCES

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items, regardless of complexity. *Psychological Science*, *18*(7), 622–628.

Cowan, N., & Morey, C. C. (2006). Visual working memory depends on attentional filtering. *Trends in Cognitive Sciences*, *10*, 139–141.

Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *Journal of Neuroscience*, *29*, 8726–8733.

Heinze, H. J., Luck, S. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focussed attention with bilateral stimulus arrays. I: Evidence for early selection. *Electroencephalography and Clinical Neurophysiology*, *75*, 511–527.

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*, *353*, 1257–1270.

Mangun, G. R., Hillyard, S. A., & Luck, S. J. (1993). Electrocortical substrates of visual selective attention. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 219–243). Cambridge, MA: MIT Press.

Vogel, E. K., McCollough W. A., & Machizawa, G. M. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503.

Sound–shape congruency affects the multisensory response enhancement

Elena Makovac and Walter Gerbino

Department of Psychology "Gaetano Kanizsa" and BRAIN Centre for Neuroscience, University of Trieste, Trieste, Italy

Perception combines prior knowledge and inputs from different senses. Several mechanisms mediating multisensory integration and its subsequent effects on cognitive and behavioural processes have been studied and three general principles have been suggested: The spatial rule, the temporal rule, and the inverse effectiveness rule (Bolognini & Làdavas, 2005; Calvert, 2001; Spence & Driver, 1997). According to such rules, the spatiotemporal coincidence of two different stimuli (e.g., auditory and visual) is a necessary

Please address all correspondence to Elena Makovac, Department of Psychology "Gaetano Kanizsa", University of Trieste, via Sant'Anastasio 12, 34134 Trieste, Italy. E-mail: elena.makovac@hotmail.it

condition for multisensory integration, resulting in an effect known as the *multisensory response enhancement* (MSE): The response to multisensory stimuli exceeds the response to the most effective unisensory stimulus, whereas spatially and temporally disparate stimuli produce either depression or no change in the response.

Perception is also defined in terms of feature extraction. For instance, we can abstract common properties from sounds and images. A relevant demonstration was discussed by Köhler (1929, pp. 224–225), who asked people to match the nonsense words *takete* and *maluma* to two shapes, one spiky and one curvy, and reported that most answered without hesitation. This result suggests that sounds are not attached to shapes arbitrarily. They share properties that can be abstracted, like the "sharp" and "cutting" quality of a word, or the "soft" quality of a visual shape. Ramachandran and Hubbard (2001) called this the *bouba/kiki* effect.

More recently, Gallace and Spence (2006) provided the first empirical demonstration that a simultaneous irrelevant sound can speed up the explicit response in a visual size judgement task. They showed that reaction times were shorter in congruent (e.g., high frequency sound coupled to a small shape) than incongruent (e.g., low frequency sound coupled to a small shape) conditions.

We argue that the MSE, well documented in the case of spatiotemporal coincidence between two different stimuli, should be stronger in the case of sound–shape congruency. Responses should be faster in the multisensory condition (when a sound and a visual shape are presented together) than in the unisensory condition (when only a visual shape is presented, either spiky or curvy); and the advantage should be larger in congruent (spiky sound coupled with a spiky visual shape, curvy sound coupled with a soft visual shape) than incongruent (spiky sound coupled with a soft visual shape, curvy sound coupled with a spiky visual shape) conditions.

METHOD

Visual stimuli were either red or blue on a yellow background. Stimulus efficiency was high for red (strong luminance contrast relative to the background) versus low for blue (weak luminance contrast relative to the background). Since form perception is mediated mainly by the luminance channel, we expected faster responses to red than blue visual shapes.

Before the experimental session, each participant was required to match the two sounds to the two shapes. All participants ($N = 11$) associated the spiky sound to the spiky figure and the soft sound to the curvy figure.

Participants should respond only to visual stimuli (either spiky or curvy), in unisensory (visual stimuli alone) and multisensory (visual and auditory

stimuli presented at the same time and in the same spatial position) conditions, and not respond to auditory stimuli alone. Since attention should be concentrated on visual stimuli, auditory stimuli were task-irrelevant.

All trials included three events: 500 ms background with central fixation cross; empty background lasting 130–1200 ms; two outline squares, lasting 700–1100 ms, at the left and right of the fixation cross, marking the possible locations of the visual target. Stimulation conditions differed according to the last 67 ms event, which included a spiky/soft sound in *unisensory auditory* (catch) trials; a red/blue spiky/curvy shape *in unisensory visual* (US) trials; a spiky sound coupled with a spiky shape or a soft sound coupled with a curvy shape in *multisensory congruent* (MSC) trials; a soft sound coupled with a spiky shape or a spiky sound coupled with a curvy shape in *multisensory incongruent* (MSI) trials (Figure 1A).

Each participant was shown 480 trials, divided into 15 blocks. Every block contained 32 trials in a different random sequence, including: Eight catch trials (resulting from the $2 \times 2 \times 2$ combination of position, sound and repetition), and eight US, eight MSC, and eight MSI trials (all resulting from the $2 \times 2 \times 2$ combination of colour, position, and shape).

RESULTS

Figure 1B shows the distribution of reaction times. In US trials responses were faster for red than blue targets, $F(1, 10) = 39.79$, $p < .01$, as expected on the basis of luminance contrast. With blue targets we obtained a significant MSE in both congruent and incongruent conditions, $F(1, 10) = 59.33$ and 30.62, respectively, both $p < .01$, and a superiority of MSC over MSI trials, $F(1, 10) = 4.71$, $p < .05$, as expected on the basis of sound–shape congruency. On the contrary, with red targets we obtained neither an MSE, $F(1, 10) = 1.90$ and 1.30 for congruent and incongruent conditions, respectively, both $p > .05$, nor a difference between congruent and incongruent conditions, $F < 1$. Results for red targets can be attributed to a *floor effect*: The simultaneous sound cannot improve the fast response to a high efficiency visual target.

DISCUSSION

Data in Figure 1B are consistent with the so-called *inverse effectiveness rule*. When the efficiency of the unisensory stimulus is low, an MSE is always obtained, but the effect is stronger when the properties of stimuli in the two coupled modalities are linked by a qualitative similarity (the shape–sound congruency in the *takete/maluma* effect), confirming that multisensory integration depends on symbolic, as well as spatiotemporal, proximity. When the efficiency of the unisensory stimulus is high, the MSE is smaller or

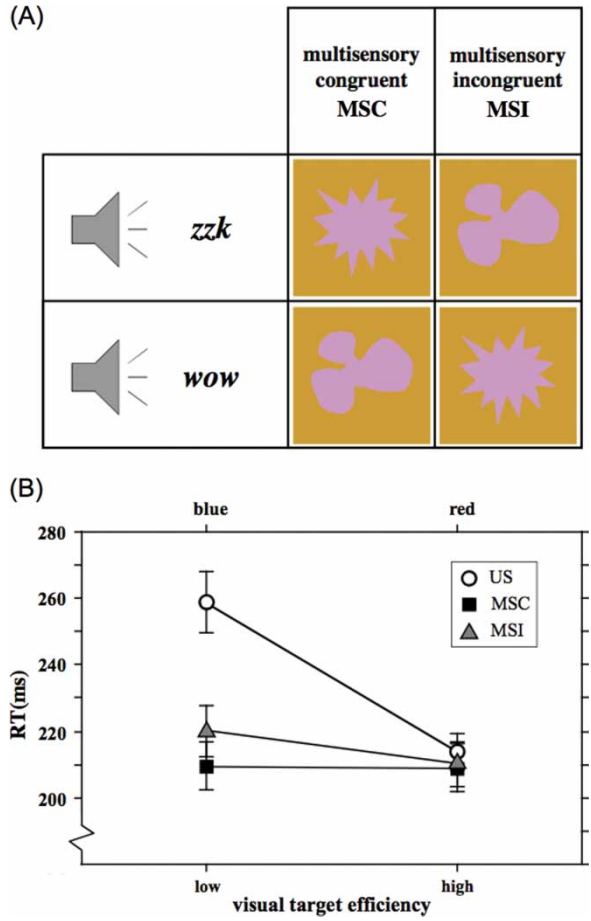

Figure 1. (A) The sound–shape coupling in MSC (spiky shape with spiky sound, curvy shape with soft sound) and MSI (vice versa) trials. (B) Mean reaction times (± 1 SEM) for low-efficiency (blue) and high-efficiency (red) targets in different trials. To view this figure in colour, please see the online issue of the Journal.

even absent, because performance in the unisensory condition is already optimal.

To summarize, responses to low-efficiency visual targets are improved by task-irrelevant simultaneous sounds, better if congruent. The MSE effect and the congruency effect were obtained in an implicit task, in which participants were not required to evaluate the properties of targets and other unattended stimuli, but should simply detect relevant events in the designated modality.

REFERENCES

Bolognini, N., & Làdavas, E. (2005). Visual localization of sounds. *Neuropsychologia*, *43*, 1655–1661.

Calvert, G. A. (2001). Cross-modal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110–1123.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception and Psychophysics*, *68*, 1191–1203.

Köhler, W. (1929). *Gestalt psychology.* New York: Liveright.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies*, *8*, 3–34.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception and Psychophysics*, *59*, 1–22.

Making waves in the stream of consciousness: Eliciting predictable oscillations in visual awareness with pretarget entrainment at 12 Hz

Kyle E. Mathewson, Monica Fabiani, Gabriele Gratton, Diane M. Beck, and Alejandro Lleras

Beckman Institute and Department of Psychology, University of Illinois at Urbana Champaign, Urbana, IL, USA

When stimuli are presented at the threshold of consciousness, identical stimulus presentations can result in very different states of awareness from moment to moment. The well-defined psychophysical laws describing the relationship between stimulus properties and sensitivity are, however, probabilistic in nature, and thus cannot explain why any particular stimulus reaches awareness.

Recently, it has been shown that this variability in conscious awareness may be a function of the phase of natural rhythmic oscillations in the brain (Mathewson, Gratton, Fabiani, Beck, & Ro, 2009). Here, we ask whether these oscillations in awareness can be experimentally controlled by entrainment to rhythmic visual signals in the environment. We tested if visual sensitivity can be synchronized to the expected occurrence of rapid, rhythmic events, with maximal sensitivity to targets presented in phase with the

Please address all correspondence to Kyle E. Mathewson, Beckman Institute, University of Illinois, 405 N. Mathews Ave., Urbana, IL 61801, USA. E-mail: kmathew3@illinois.edu

Supported by a Natural Science and Engineering Research Council of Canada Scholarship to KEM.

preceding entraining stimulation, and reduced sensitivity for out-of-phase targets. This method would therefore represent a novel technique to control the degree to which identical stimuli reach consciousness.

Prior work in the auditory modality has shown an enhancement of auditory processing for targets presented in phase with a regular but irrelevant series of tones, resulting in enhanced sample matching (Jones, Moynihan, MacKenzie, & Puente, 2002). Here, we show that rapid visual entrainment (at 12 Hz) can modulate visual awareness of a subsequently presented target. A metacontrast mask was used (at optimal stimulus–onset asynchrony—SOA) to lower target visibility to near-threshold levels.

METHODS

Figure 1A shows the stimulus dimensions and task sequence completed by 16 University of Illinois at Urbana-Champaign students. After a fixation cross and a blank screen, a variable number of masks were presented as entrainers (0, 2, 4, or 8). Entrainers were presented at 12 Hz (comparable to the frequency of the natural brain oscillations described in Mathewson et al., 2009). After the final entrainer, the SOA before the target onset (tSOA) was either 82 ms (in phase) or 32, 59, 107, or 130 ms. A control condition was included with only the first and last entrainers from the eight-entrainer condition with a blank in-between interval in order to control for foreperiod effects of the entrainers.

A metacontrast mask followed each target after a constant mSOA. A fixation cross then reappeared and subjects indicated whether or not they had seen the target. On some trials, the target was omitted. Blocks randomly mixed trials of each entrainment condition (0, 2, 4, 8, or control) and each tSOA condition.

RESULTS

Figure 1B shows the average detection rate for each entrainment condition as a function of the tSOA between the final entrainer and the target. The average detection rate in the no-entrainers condition ($M = 0.45$) indicated considerable metacontrast masking. Target detection in the control condition was lower ($M = 0.22$), suggesting substantial forward masking by the "last entrainer". As predicted, the presentation of entraining stimuli prior to the onset of the target at 12 Hz elicited a peak in detection at the moment the next entrainer would have occurred (82 ms tSOA), as well as worsening detection for targets appearing before or after this point in time, evidently releasing the target from the effects of forward masking. Significant quadratic effects were found for two, $t(15) = 3.5$, $p = .003$, four, $t(15) = 4.9$,

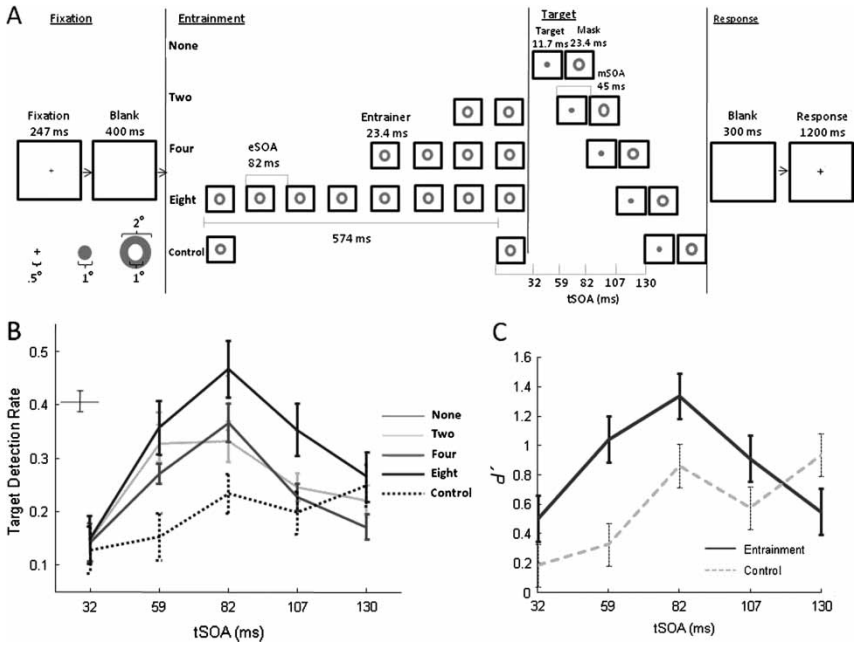

Figure 1. (A) Stimulus dimensions and trial timeline. (B) Target detection rate as a function of the time after final entrainer onset (tSOA) for each entrainment condition. Apparent is a peak in detection in phase with the preceding entrainment. (C) d' collapsed across entrainment condition compared to the control condition. All error bars represent within-subject SE.

$p = .0002$, and eight, $t(15) = 5.5$, $p < .0001$, entrainer conditions, but not for the control, or no-entrainment condition, $t(15) < 1$, *ns.* Furthermore, this effect scaled with the number of entrainers ($M_b = 0.22$), $t(15) = 6.53$, $p < .0001$. Last, collapsed across the 2-, 4-, and 8-entrainer conditions, d'also showed a significant quadratic trend, $t(15) = 4.3$, $p < .001$, revealing a peak in sensitivity elicited by the preceding entrainment that was not present for the control condition, $t(15) < 1$, *ns* (Figure 1c).

DISCUSSION

Here we showed for the first time that 12 Hz oscillations in the visual system entrained by rapid exogenous presentation of visual stimuli can cause oscillations in the ongoing stream of consciousness. When targets are presented in phase with preceding entrainers, sensitivity increases, and detection is more likely.

Since both "steady state" activity elicited by the entraining stimuli (Pastor, Artieda, Arbizu, Valencia, & Masdeu, 2003), and endogenous

oscillatory cortical activity (Mathewson et al., 2009) elicit similar oscillations in visual awareness, the possibility arises that our entrainment effect represents similar fluctuations in cortical excitability as the endogenous oscillations, providing a mechanism for temporal attention. This is supported by intracortical animal recordings showing that attending to trains of either 7 Hz tones or flashes entrains slow cortical rhythms in the attended modality only, leading to enhanced processing of the temporally attended modality (Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008).

Entrained excitability cycles may begin to explain many seemingly unrelated phenomena (Ward, 2003). For instance, in a rapid serial visual presentation (RSVP) paradigm, distractors and rarely occurring targets are presented serially at 10 Hz. Target detection increases as a function of the number of preceding distractors, an effect that has recently been referred to as *attentional awakening* (Ambinder & Lleras, in press; Ariga & Yokosawa, 2008), and is supported by an MEG study showing evidence of cortical synchronization to RSVP stimulation (Gross et al., 2004). Other recent work suggests that entrainment of neural oscillations may play a part in speech perception (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008).

Presented was a novel demonstration that, in the presence of rapid visual events, visual sensitivity is modulated to peak precisely at the time when a stimulus is expected, highlighting the inherently predictive function of brain processing (Enns & Lleras, 2008). Last, our results also reveal a new and powerful technique to experimentally control what does and does not reach conscious awareness.

REFERENCES

Ambinder, M. S., & Lleras, A. (in press). Temporal tuning and attentional gating: Two distinct attentional mechanisms on the perception of rapid serial visual events. *Attention, Perception and Psychophysics*.

Ariga, A., & Yokosawa, K. (2008). Attentional awakening: Gradual modulation of temporal attention in rapid serial visual presentation. *Psychological Research*, *72*(2), 192–202.

Enns, J. T., & Lleras, A. (2008). What's next? New evidence for prediction in human vision. *Trends in Cognitive Sciences*, *12*, 327–333.

Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K. L., Hommel, B., & Schnitzler, A. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proceedings of the National Academy of Sciences*, *101*, 13050–13055.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*(4), 313–319.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*, 110–113.

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To see or not to see: Prestimulus alpha phase predicts visual awareness. *Journal of Neuroscience*, *29*(9), 2725–2732.

Pastor, M. A., Artieda, J., Arbizu, J., Valencia, M., & Masdeu, J. C. (2003). Human cerebral activation during steady-state visual-evoked responses. *Journal of Neuroscience*, *23*(37), 11621–11627.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106–113.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*(12), 553–559.

The interaction of surface feature and spatiotemporal continuity in object-based inhibition of return

A. Caglar Tas

University of Iowa, Iowa City, IA, USA

Michael D. Dodd

University of Nebraska-Lincoln, Lincoln, NE, USA

Andrew Hollingworth

University of Iowa, Iowa City, IA, USA

An important topic in the study of dynamic object perception is the problem of object correspondence: How does the visual system maintain stable object representations across frequent perceptual changes (such as motion) and disruptions (such as occlusion)? The computation of object correspondence could depend on the continuity of an object's spatiotemporal features or the continuity of an object's surface features. In the former case, an object would be perceived as a single, persisting entity if its position over time was consistent with that interpretation. In the latter case, an object would be perceived as a single, persisting entity if its surface features (colour, shape, etc.) were consistent across disruption or change. According to object file theory, correspondence operations consult only spatiotemporal information (Kahneman, Treisman, & Gibbs, 1992; Mitroff & Alvarez, 2007). However, recent findings suggest that surface feature information can also contribute to object correspondence across motion (Moore, Mordkoff, & Enns, 2007), occlusion (Hollingworth & Franconeri, in press), and saccades (Richard, Luck, & Hollingworth, 2008). The aim of the present study was to investigate the effects of spatiotemporal and

Please address all correspondence to A. Caglar Tas, Department of Psychology, 11 Seashore Hall E, University of Iowa, Iowa City, IA 52242-1407, USA. E-mail: caglar-tas@uiowa.edu

surface feature continuity on the phenomenon of object-based inhibition of return (IOR). Object-based IOR is manifested as impaired perceptual processing of and delayed orienting to a previously attended object, even when that object has moved to a new location (Tipper, Weaver, Jerreat, & Burak, 1994). Object-based IOR provides a direct measure of correspondence, as object-based IOR depends on the visual system treating an object as a continuation of a previously attended object. By manipulating the spatiotemporal and surface feature properties of objects, we sought to determine the types of information functional in defining a persisting object for the purpose of object inhibition.

In Experiment 1, the aim was to investigate whether a salient change in surface features can disrupt object correspondence (see Moore et al., 2007) and alter the magnitude of object-based IOR. To test this, we presented two differently coloured disks on the circumference of a virtual circle, separated by 180°. One disk was cued, and participants executed a saccade to that object. A second cue brought gaze back to the centre. Then, the two objects moved 90° clockwise around the virtual circle. A small target was presented in one of the two objects, and participants executed a saccade to that target as quickly as possible. Saccade latency was the dependent measure. On half of the trials, the two objects retained their originals colours throughout the trial, and we expected to observe a significant object-based IOR effect (increased saccade latency when the target appeared in the previously fixated object). On the other half, the two objects abruptly swapped colours during their motion. If object correspondence is established solely on the basis of spatiotemporal continuity, then we would expect an equivalent IOR effect in this colour-swap condition, because there was no discontinuity in spatiotemporal information. However, if surface features also play a role in establishing correspondence during motion, then a disruption of surface feature continuity should significantly reduce the IOR effect. The results of Experiment 1 favoured the second account. A significant IOR effect was found in the no-colour-swap condition, but this effect was eliminated in the colour-swap condition, demonstrating that a discontinuity in surface features blocked the interpretation of object correspondence.

In Experiment 2 we asked a question complementary to that asked in Experiment 1. Would a salient discontinuity in spatiotemporal information disrupt object correspondence on the basis of a surface feature match? Instead of moving smoothly to new locations, the two objects moved to the new locations in a single step across a blank ISI, creating salient discontinuity in spatiotemporal information. The only information available to distinguish the cued and uncued object was colour. In this experiment, a colour match was insufficient to produce object-based IOR: Saccade latency did not significantly increase when the target appeared in the object that

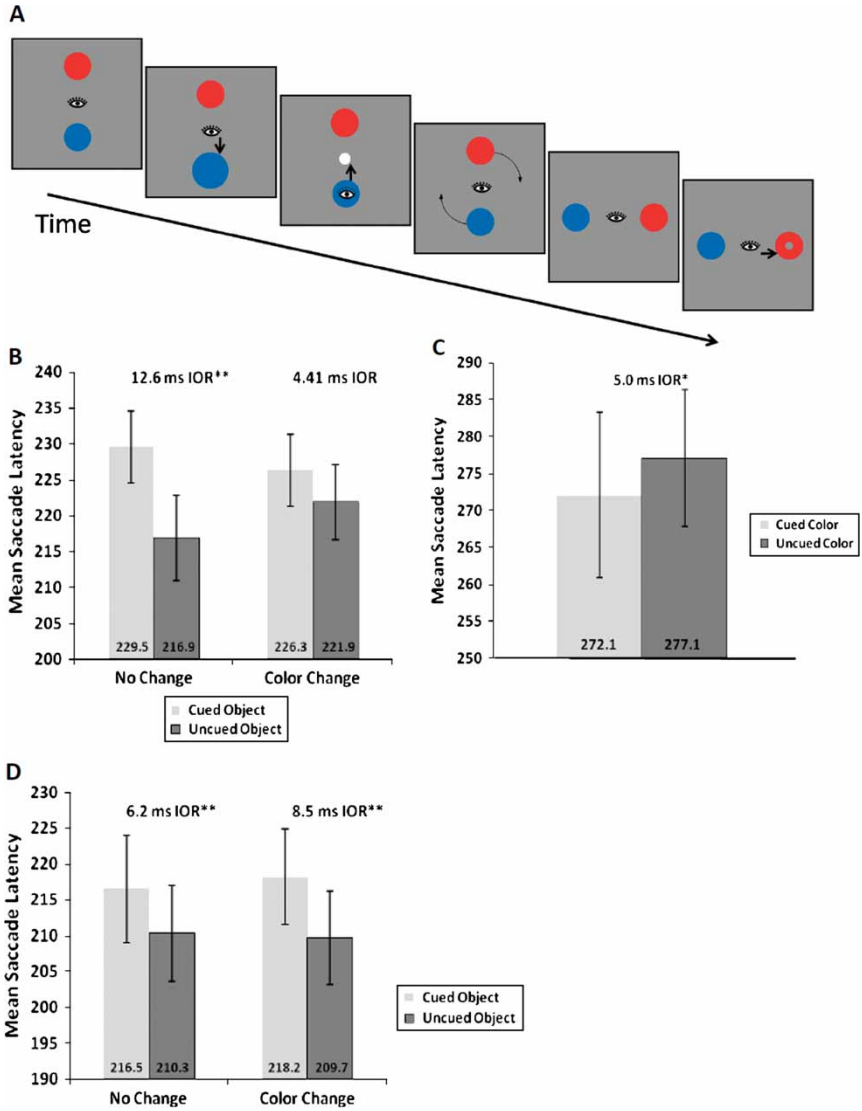

Figure 1. (A) Sequence of events in a no-colour-change/cued object trial for Experiment 1. In Experiment 2, the sequence was the same except there was a blank screen with the fixation dot instead of the motion in the fourth panel. In Experiment 3, there were two occluders in the path of the motion. (B–D) Mean saccade latencies for Experiments 1, 2, and 3, respectively. Error bars represent ± 1 SEM. $^{*}p < .5$, $^{**}p < .01$. To view this figure in colour, please see the online issue of the Journal.

matched the colour of the previously fixated object. These results indicate that a discontinuity in spatiotemporal features also blocks the computation of object correspondence.

In Experiments 1 and 2, object-based IOR was eliminated when there was visible discontinuity in either surface features or spatiotemporal features, suggesting that both contribute to object correspondence operations. In Experiment 3, we examined whether the perception of an abrupt surface-feature change was necessary to disrupt object correspondence by masking the transient associated with the colour swap in Experiment 1. The method was the same as in Experiment 1, except there were two occluders (slightly larger than the objects themselves) that lay in the path of object motion. On colour-swap trials, the two objects swapped colours during the very brief period that both were occluded, masking the transient associated with the colour change. A significant object-based IOR effect was observed both for colour-swap and no-colour-swap conditions. Importantly, there was no significant difference in the magnitude of IOR effects between the two conditions, suggesting that a colour change does not disrupt object correspondence if the transient created by the change is masked.

In summary, the present study indicated that both spatiotemporal and surface feature continuity are used to establish object correspondence. A salient, visible discontinuity in either dimension is sufficient to disrupt the interpretation of an object as a single, persisting entity. These data argue against the claim that spatiotemporal cues dominate the computation of object correspondence. Instead, the visual system consults multiple sources of relevant information to establish continuity across change and perceptual disruption.

REFERENCES

Hollingworth, A., & Franconeri, S. L. (in press). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface feature cues. *Cognition*.

Kahneman, D., Triesman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin and Review*, *14*(6), 1199–1204.

Moore, C. M., Mordkoff, J. T., & Enns, J. T. (2007). The path of least persistence: Evidence of object-mediated visual updating. *Vision Research*, *47*, 1624–1630.

Richard, A. M., Luck, S. J., & Hollingworth, A. (2008). Establishing object correspondence across eye movements: Flexible use of spatiotemporal and surface feature information. *Cognition*, *109*, 66–88.

Tipper, S. P., Weaver, B., Jerreat, L. M., & Burak, A. L. (1994). Object-based and environment-based inhibition of return of visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(3), 478–499.

The perception of space is warped by objects

Timothy J. Vickery and Marvin M. Chun
Yale University, New Haven, CT, USA

The perception of space within an object seems veridical or at least equivalent to the space outside of an object's boundaries. However, we discovered that an object's interior space is systematically distorted. We observed this phenomenological distortion in a display in which two dots were placed within a rectangular object, and two dots with equivalent spacing were placed outside the object (Figure 1A). The dots inside the object clearly appeared to be further apart than the dots outside the object.

Empirical studies of the perceived distance between distinct, grouped elements have produced conflicting results. In one study (Coren & Girgus, 1980), observers showed a weak tendency to underestimate space between grouped items, relative to space between ungrouped items. However, in a study that attempted to replicate Coren and Girgus, space within grouped items was expanded relative to space within ungrouped items (Vickery & Jiang, 2009). To resolve this discrepancy, and to further reveal the basic nature of the distortion, we mapped spatial perception within well-defined objects.

Figure 1A provides a salient phenomenological demonstration that regions within a figure region are spatially warped. The two dots inside the object are the same distance apart as the dots that are placed on the ground region, yet the spacing of the dots in the object appears greater. However, Figure 1B demonstrates that this distortion does not extend to the edges of the object. Here, the dot pairs appear equally spaced. We employed a distance-matching paradigm to measure this effect. On each trial, two reference dots appeared in the monitor's upper left quadrant, and to prevent collinearity, two adjustment dots appeared in the lower right. Participants adjusted the adjustment-dot spacing to match their perception of the reference dots' spacing.

EXPERIMENT 1

In Experiment 1 ($N = 9$), reference dots appeared centred on a black object ($2^\circ \times 6^\circ$), just inside or outside the object's edge, or outside the object (equally distant from the edge as the centre of the object). Dots were aligned

Please address all correspondence to Timothy J. Vickery, Department of Psychology, Yale University, 2 Hillhouse Ave., New Haven, CT 06520, USA. E-mail: tim.vickery@gmail.com

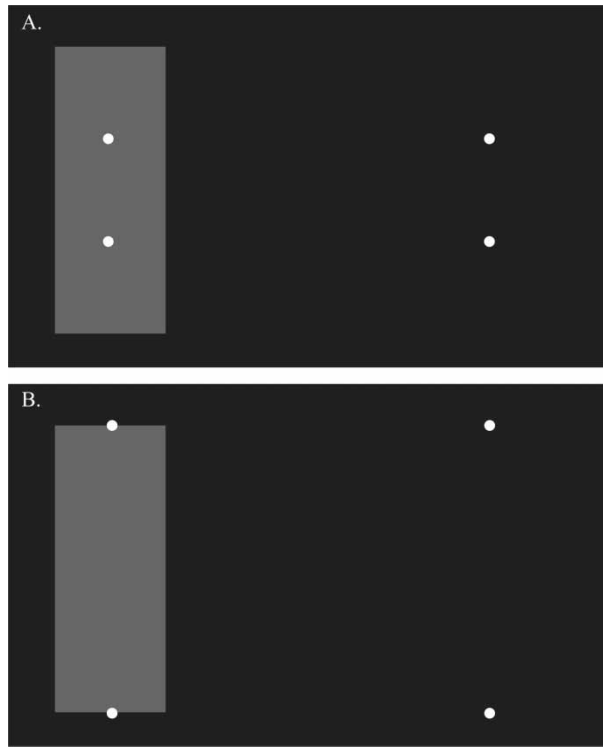

Figure 1. (A) Two dots within a rectangular object, and two dots with equivalent spacing outside the object. (B) Equally spaced dot pairs.

with the longer axis of the object, and presented at three spacings: 2°, 4°, or 6°. At the largest spacing the dots subtended the whole object. As a control, the dots sometimes appeared in empty space (grey) using the same three spacings and horizontal positions.

Results showed significant distortions of space in the presence of an object. At 2° and 4° spacings, when dots were placed inside the object, or on its inner or outer edges, participants reported distances significantly larger than reality and larger than reports in control conditions that included no object (all $ps < .01$; paired t-tests with corresponding no-object conditions unless noted). For 2° spacings, reports were 9.2% larger at the centre of the object than control (2.36° vs. 2.16°); for the 4° spacing the reports were an average 11.3% larger than control (4.56° vs. 4.09°). Participants still showed distortions of space when the dots were outside the object (2.17° vs. 2.11° and 4.23° vs. 4.06°; both $ps < .05$), but these distortions were significantly smaller than within the object (both $ps < .005$; t-tests between outside and inside positions, when the object was present). On the other hand, at the 6°

spacing, when the dot pair subtended the entire length of the object, no significant distortions of space were observed for dots centred on an object or near its edges, as suggested by Figure 1B.

EXPERIMENT 2

The distortions observed in Experiment 1 did not depend on the 3:1 aspect ratio of the object. In Experiment 2 ($N = 10$), we replicated the results when dots were inside a square. Within a $6^\circ \times 6^\circ$ square, dots separated by 1°, 2°, and 4° were reported to be an average of 1.15°, 2.32°, and 4.56° apart, respectively; corresponding baseline (no-object) conditions were estimated to be 1.07°, 2.05°, and 3.98° apart (all ps $< .005$). Expanded representations of space were also found for probes oriented along the shorter axis within a 3:1 rectangle, when the dots were contained within the object's boundaries (at 1° spacing, estimates were 1.19° with the object and 1.07° without the object, $p < .001$), but not when they spanned the object or crossed the object, at 2° and 4° spacings.

EXPERIMENT 3

In Experiment 3 ($N = 9$), we asked whether objecthood was critical to the spatial distortion effect. Two side-by-side objects were presented (each $2^\circ \times 6^\circ$, separated by 2° edge-to-edge). When two dots were presented inside the same object (at 4° spacing), their spacing was overestimated compared to when no objects were visible (4.61° vs. 4.11°, $p < .001$). Two dots at an equal distance, but inside two different objects, were also misjudged to be farther apart than reality (4.34° vs. 4.11°, $p < .005$), but the distortion was significantly smaller than in the same-object condition ($p < .05$). When dots appeared in the empty space between the objects, parallel to the objects, they were still perceived as farther apart than when no object was present (4.46° vs. 4.11, $p < .005$) , but the estimated separation was significantly smaller than when the dots appeared inside one object ($p < .01$). These results suggest that objecthood is important to the distortion effect, since the effect is not entirely explained by distortions that occur where local image properties are similar, but the dots did not appear inside the same object.

CONCLUSION

We demonstrated that the presence of an object distorts spatial perception within and near the object. Spatial distances appear larger within objects, except for distances that span the entire length of the object. Fellows (1968) observed distortions of line length analogous to Figure 1. A line framed by a

rectangle of greater length appears shorter than when framed by one with equal length. Shapes circumscribed by circles are also seen as larger than reality (the Delboeuf illusion; Jaeger, 2007). Our results imply that distortions are not restricted to circumscribed shapes, but reflect general spatial distortions that even extend beyond the borders of an object. Future experiments will be focused on mapping out warped space in detail, examining the distorting effects of different kinds of shapes, and identifying the source of this distortion.

REFERENCES

Coren, S., & Girgus, J. S. (1980). Principles of perceptual organization and spatial distortion: The Gestalt illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(3), 404–412.

Fellows, B. J. (1968). The reverse Müller-Lyer illusion and "enclosure". *British Journal of Psychology*, *59*(4), 369–372.

Jaeger, T. (2007). Circumscribed shapes are enlarged: Is this a variation of the Delboeuf illusion? *Perceptual and Motor Skills*, *104*(3, Pt. 2), 1116–1118.

Vickery, T. J., & Jiang, Y. V. (2009). Associative grouping: Perceptual grouping of shapes by association. *Attention, Perception, and Psychophysics*, *71*(4), 869–909.

The time course of initial scene processing

Melissa L.-H. Võ and John M. Henderson

Psychology Department, University of Edinburgh, Edinburgh, UK

As soon as we encounter a new scene, several processes will be triggered: From an initial sweep of global low-level feature computation (e.g., colour and low spatial frequencies) to high-level cognitive processes (e.g., activation of semantic knowledge; for a review see Oliva, 2005). Gist can be extracted very quickly on the basis of an initial scene analysis, but the programming of eye movements for the purpose of active scene inspection might involve more than just the identification of scene gist (for reviews, see Henderson, 2003, 2007). In this study, we investigated the time course of early scene processing for guiding action. We were specifically interested in the time course between the initial glimpse of a scene and the initiation of object search via eye movements.

Please address all correspondence to Melissa Le-Hoa Võ, Psychology Department, 7 George Square, S32, University of Edinburgh, Edinburgh EH8 9JZ, UK. E-mail: melissa.vo@ed.ac.uk

To date, studies investigating the time course of initial scene processing have focused mainly on the speed at which visual information can be processed to allow for rapid scene categorization or object identification (e.g., Greene & Oliva, 2009; Thorpe, Fize, & Marlot, 1996). These studies have provided compelling evidence that sophisticated scene analysis can be accomplished with scene presentation durations of 50 ms. However, the time course of early scene processing with regard to its influence on eye movement planning has largely been neglected.

In a series of five experiments, we used the flash-preview moving-window paradigm (Castelhano & Henderson, 2007) to investigate initial scene analysis for eye movement planning. In this paradigm, a scene is briefly previewed prior to visual search for a target object. The search then takes place through a moving window that reveals only a small area of the scene tied to the current fixation. This paradigm allows isolating the effect of the initial scene glimpse on subsequent eye movements from the processing that takes place during later stages of scene viewing. Thus, we were able to test the minimum amount of scene presentation time needed to provide sufficient information to subsequently guide eye movements to probable target locations during search. Further, we investigated whether establishing an initial scene representation might depend on the time available to integrate the initially fleeting scene representation (e.g., Intraub, 1980; Potter, 1976). Our investigation of early scene processing therefore included manipulations of scene presentation duration (Experiments 1–3) and integration time (Experiments 4–5).

GENERAL METHODS

The stimulus material used in all experiments consisted of 45 full-colour images of real-world scenes. Eye movements were recorded with an EyeLink1000 tower system (SR Research, Canada) sampling at 1000 Hz. We used the flash-preview moving-window paradigm, which has been successfully applied to investigate the influence of the initial glimpse of a scene on subsequent eye movement control during search (Castelhano & Henderson, 2007; Võ & Schneider, 2009). In this paradigm, participants are first presented with a brief preview of the search scene excluding the target object, followed by the presentation of a target word indicating which object they will be looking for. The scene is then presented including the target for search, but participants are only able to view the scene through a 5° circular gaze-contingent window (see Figure 1A for an example trial sequence). We varied the duration of scene previews using 100 ms (Experiment 1), 75 ms (Experiment 2), and 50 ms (Experiment 3) to find the minimum preview duration required for efficient object search. We also

investigated whether providing additional integration time after scene preview but before initiation of search would increase any benefits of the preview (Experiments 4 and 5).

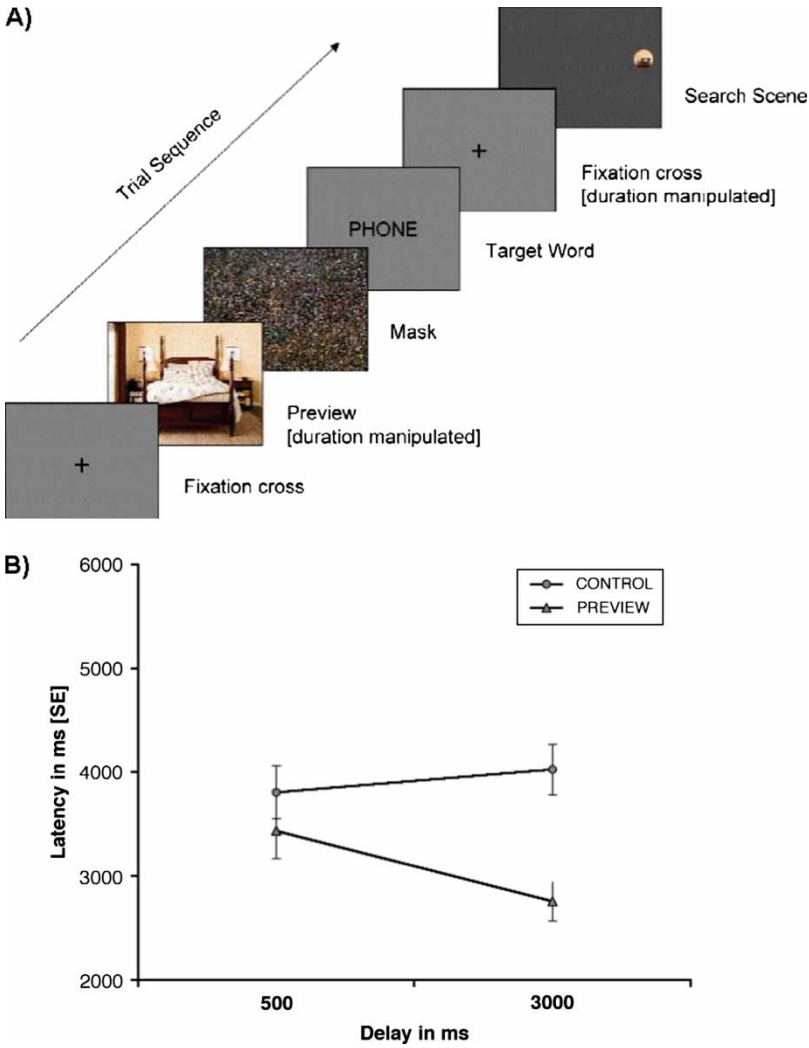

Figure 1. (A) Trial sequence of the flash-preview moving-window paradigm. (B) Mean latencies to first target fixation as a function of preview (scene preview vs. control) and delay (500 ms vs. 3000 ms) for Experiment 5 with a scene presentation duration of 50 ms. To view this figure in colour, please see the online issue of the Journal.

RESULTS

To investigate whether eye movement planning can benefit from information acquired from a very brief scene glimpse, we calculated a set of measures of viewers' search behaviour including response times, latency, and number of fixations to first target fixation. For the sake of brevity and since results were consistent across all measures, we only present latency measures. In the first three experiments, preview durations of 100 ms, 75 ms, and 50 ms were compared to both 0 ms and 250 ms preview conditions. We observed that 75 ms and 100 ms previews sufficed to produce significant search benefits compared to a 0 ms preview, $t(14) = 2.03$, $p < .05$, and $t(14) = 3.04$, $p < .01$, respectively, whereas a 50 ms preview failed to provide such search benefits, $t < 1$. Experiment 4 showed that inserting a delay following the target word and before the initiation of search led to an increase in search benefits when combined with a preview duration of 250 ms, $t(15) = 5.94$, $p < .01$. This result implies that an added delay supports search planning on the basis of which eye movements can be guided to probable target locations. Interestingly, the 50 ms preview that had failed to provide search benefits in Experiment 3 also led to increased search efficiency when additional integration time was provided in Experiment 5, $t(15) = 4.31$, $p < .01$ (see Figure 1B). Thus, preview durations as short as 50 ms can facilitate object search via eye movements when coupled with sufficient time to integrate visual information with target object knowledge.

DISCUSSION

The present findings demonstrate that information derived from as little as 50 ms of scene presentation can be used to facilitate control of subsequent eye movements during search in natural scenes. However, to use the information derived from such brief presentations, additional integration time following scene presentation and target specification is needed: Although a 50 ms scene preview did not result in search benefits in Experiment 3, Experiment 5 showed that a preview duration of 50 ms combined with extra integration time increased search performance. We therefore propose that a presentation time of 50 ms can be sufficient to establish scene representations that are able to facilitate eye guidance, but only when enough time is available to integrate the information provided from an initial glimpse with knowledge of the search target. The results confirm the great speed at which the visual-cognitive system is able to extract useful scene information. The present findings also show that fast scene processing is not limited to activating gist. Instead, scene representations generated from a brief scene glimpse can provide sufficient information to guide subsequent behaviour, as shown here by facilitated eye movement guidance in visual search.

REFERENCES

Castelhano, M., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *40*(4), 464–472.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*(4), 219–222.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(1), 1–12.

Oliva, A. (2005). Gist of the scene. In. L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *The encyclopedia of neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 509–522.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6), 520–522.

Võ, M. L.-H., & Schneider, W. X. (2009). A glimpse is not a glimpse: Differential processing of flashed scene previews leads to differential target search benefits. *Visual Cognition*. Advance online publication. doi:10.1080/13506280802547901

Modelling and quantifying tradeoffs in multiple object tracking

Edward Vul and Michael C. Frank

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

George A. Alvarez

Psychology Department, Harvard University, Cambridge, MA, USA

Josh B. Tenenbaum

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Multiple object tracking (MOT) has recently become a popular method to investigate the cognitive architecture of human visual attention (Pylyshyn &

Please address all correspondence to Edward Vul, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. E-mail: evul@mit.edu

Storm, 1988). MOT experiments have documented successes and failures in this task, but how do these experiments map onto cognitive architecture? Thus far, failures have been attributed to limits of attention, memory, an object system, a tracking module, motion perception, and other cognitive structures (Alvarez & Franconeri, 2007; Blaser, Pylyshyn & Holcombe, 2000; Fencsik, Urrea, Place, Wolfe, & Horowitz, 2006; Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008; Keane & Pylyshyn, 2006; Makovski & Jiang, 2009; Pylyshyn & Storm, 1988; Tripathy & Barrett, 2004). Here we propose to connect experimental findings to theories of cognitive architecture through a computational analysis of the task. To do so, we define an ideal Bayesian observer for multiple object tracking and compare human performance to that of the optimal model in a series of novel experiments that quantitatively measure tradeoffs between constraints on object tracking.

MOT is a well-posed problem for machine learning, with optimal models defined for variants of this problem (Streit & Luginbuhl, 1995). In the variant posed in human MOT experiments, observers are presented with a set of labelled observations and then a sequence of unlabeled observations as objects move around. At each point in time, an observer must determine which observation corresponds to which of the initial observed labels, given uncertainty from perceptual noise as well as stochastic dynamics of the objects. We use a Rao-Blackwellized particle filter (based on a Kalman filter) to define an ideal observer for tracking under these conditions.

Human observers find tracking easier when objects move slower (Alvarez & Franconeri, 2007), are further apart together (Franconeri et al., 2008), when there are fewer objects, when fewer objects have to be tracked (Pylyshyn & Storm, 1988), and when objects have additional nonspatial features (Fenscik et al., 2006; Makovski & Jiang, 2009). Moreover, people can track objects through nonspatial features (like orientation, spatial frequency, and colour; Blaser et al., 2000). For the ideal observer, most of these effects fall out of the available sources of information—but do people combine information as the ideal observer does? We measured *tradeoffs* between these different constraints: How does maximum object speed change as a function of spacing; how does minimum spacing change as a function of colour stability? We presented subjects with MOT trials generated with simple linear dynamics. These dynamics can be summarized by the standard deviation of their position (when this is larger, objects tend to be further apart) and the standard deviation of their velocity (when this is larger, objects tend to move faster). In addition, each object could be assigned nonspatial features (such as colour hue-angle), which also evolved stochastically. Given these object dynamics, we could directly compare the tradeoffs between speed, space, inertia, number, and colour for the ideal observer and for human observers. If these different measures trade off for humans as they do for the ideal

observer, this indicates that people combine these varied sources of information in one tracking process, as does the ideal observer.

EXPERIMENTS

First, to measure the tradeoff between speed and spacing, we asked 10 subjects to track three out of six objects. Their goal was to adjust the difficulty of the tracking task by manipulating spatial standard deviation while speed stayed constant so that they could track the objects for 5 s. We parametrically varied speed, and could thus obtain an *iso-difficulty* contour of spacing versus velocity. Figure 1 shows that when objects move faster, subjects adjust the spacing to be wider to achieve a comfortable level of difficulty. We elicited matching performance from our model by simulating 5 s trials where the model had to track three out of six objects. Figure 1C shows the model "settings" for a 0.85 threshold; the upper and lower error bounds represent the settings to achieve an accuracy of 0.8 and 0.9, respectively. The tradeoff between space and speed when people track objects matches the tradeoff seen in the ideal observer.

We tested the prediction that people combine space and feature (e.g., colour) information when tracking. We measured iso-difficulty contours while subjects viewed objects with slowly varying colour. When colours drift slower, the whole iso-difficulty contour shifts: People can track objects at a faster speed given a particular spacing. These effects match those seen in the ideal observer, thus, it seems that when tracking objects both in space and in feature-space, the two sources of information are combined and additional feature information can compensate for less spatial information.

Many features of MOT reflect the computational structure of the task and the limited information available for tracking. However, this does not

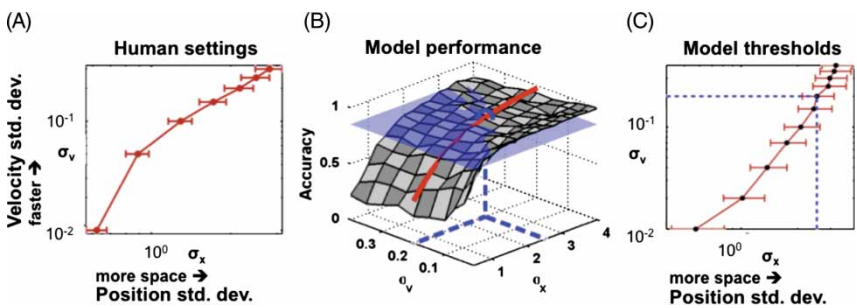

Figure 1. (A) Subjects set a "comfortable" spacing for tracking 3 of 6 objects at a particular speed. (B) Model accuracy for tracking 3 of 6 objects as a function of speed and spacing. (C) Inferred models "settings" — (85% accuracy) threshold spacing for a particular speed. To view this figure in colour, please see the online issue of the Journal.

account for limits to the numbers of objects tracked, or, more specifically: A tradeoff between the number of objects tracked and their speed limit (Alvarez & Franconeri, 2007; Pylyshyn & Storm, 1988). This limitation must be a consequence of uncertainty that may be modulated by task—a flexible resource. Within our model, the limited resource may be either visual attention, which improves the fidelity of measurements; or memory, which enables more or less noiseless propagation of state estimates through time. In both cases, when more objects are tracked, less of the resource is available for each object resulting in an increase of noise and uncertainty. By adding such a flexible noise term to our model we find the characteristic tradeoff between the number of targets and the speed with which they may be tracked is clearly evident. Thus, although many results in MOT are the consequence of the information available for the computational task, the speed–number tradeoff seems to be the result of a cognitive, task-modulated, source of uncertainty such as memory or attention.

DISCUSSION

We asked which limitations are responsible for particular failures in human multiple object performance by comparing the performance of an optimal tracking algorithm to human performance. We show that in novel behavioural experiments, the ideal observer mimics human performance with only perceptual uncertainty: Tracking is harder when objects move faster or are closer together; inertia information is available, but may not be used; and objects can be tracked in features as well as space. However, effects of the number of objects tracked do not arise from only perceptual uncertainty for the ideal observer: To account for the tradeoff between the number of objects tracked and their speed, a task-dependent resource must be introduced—we introduce this resource as a memory constraint.

REFERENCES

Alvarez, G., & Franconeri, S. (2007). How many objects can you attentively track? Evidence for a resource-limited tracking mechanism. *Journal of Vision*, *7*, 1–10.

Blaser, E., Pylyshyn, Z., & Holcombe, A. (2000). Tracking an object through feature space. *Nature*, *408*, 196–199.

Fencsik, D. E., Urrea, J., Place, S. S., Wolfe, J. M., & Horowitz, T. S. (2006). Velocity cues improve visual search and multiple object tracking. *Visual Cognition*, *14*, 92–95.

Franconeri, S., Lin, J., Pylyshyn, Z., Fisher, B., & Enns, J. (2008). Evidence against a speed limit in multiple object tracking. *Psychonomic Bulletin and Review*, *15*, 802–808.

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level non-predictive function. *Cognitive Psychology*, *52*, 346–368.

Makovski, T., & Jiang, Y. (2009). Feature binding in attentive tracking of distinct objects. *Visual Cognition*, *17*, 180–194.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.

Streit, R., & Luginbuhl, T. E. (1995). *Probabilistic multi-hypothesis tracking* (Tech. Rep. No. 10428). Newport, RI, USA: NUWC.

Tripathy, S. P., & Barrett, B. T. (2004). Severe loss of positional information when detecting deviations in multiple trajectories. *Journal of Vision*, *4*, 1020–1043.

The contralateral delay activity component of the event-related potential reflects the number of locations but not the number of objects in visual short-term memory

Lingling Wang, Steven B. Most, and James E. Hoffman

University of Delaware, Newark, DE, USA

Visual short-term memory (VSTM) is a limited capacity system that maintains visual information across short delays after visual input has ceased. The average storage capacity of VSTM has been demonstrated to be only 3–4 items at any given time (e.g., Luck & Vogel, 1997). Recent studies have used event-related potentials (ERPs; Vogel & Machizawa, 2004; McCollough, Machizawa, & Vogel, 2007) to establish a neurophysiological index of VSTM capacity: Participants were presented with a brief array containing objects in both the left and right hemifields, and they attempted to maintain items from a single hemifield in VSTM. During the retention interval, a negative deflection emerged at posterior electrode sites contralateral to the attended hemifield, which was sustained during the entire retention period. The amplitude of this contralateral delay activity (CDA) increased as the number of objects in the memory array increased, but it approached asymptote at each individual's behaviourally measured VSTM storage capacity.

Although the amplitude of the CDA is modulated by the number of representations in VSTM, it remains an open question as to whether it reflects the number of objects or the number of locations occupied by those objects. These two factors are usually confounded; however, the results of a

Please address all correspondence to Lingling Wang, Department of Psychology, University of Delaware, 108 Wolf Hall, Newark, DE 19716-2577, USA. E-mail: dangdang@psych.udel.edu

recent fMRI study distinguished neural activity reflecting the number of objects held in VSTM from that reflecting the number of locations (Xu & Chun, 2006). In that study, a memory array was presented simultaneously at different locations, sequentially at different locations, or sequentially at a single centre location. Activation in the superior intraparietal sulcus (IPS) and lateral occipital complex (LOC) increased with the number of objects presented, regardless of whether they were shown at the same or different locations, but activation in inferior IPS increased with set size only when the objects appeared at different spatial locations. The results suggested that the superior IPS and LOC may be sensitive to the number of objects held in VSTM, whereas the inferior IPS may be sensitive to the number of locations occupied by those objects.

In most previous CDA research (e.g., McCollough et al., 2007; Vogel & Machizawa, 2004), VSTM objects were presented simultaneously, at separate locations from each other, so the number of objects and the number of spatial locations occupied by the objects were always the same. Thus, it is possible that the CDA amplitude reflects the number of spatial locations occupied by the VSTM objects. The present study utilized a manipulation similar to that of Xu and Chun (2006) to determine whether the CDA component reflects the number of objects or the number of locations in VSTM.

Each trial began with a 200 ms arrow cue indicating whether the right- or left-side of the display was relevant on that trial (see Figure 1A). The cue was followed by (1) a bilateral memory array, (2) a 1000 ms retention period, and (3) a test array. Each memory and test array contained either one or two coloured squares in each hemifield, and participants were asked to detect whether any item on the relevant side of the test array was different from the corresponding item in the memory array. In both hemifields of the memory and test arrays, each coloured square appeared for 100 ms; when the memory array contained two squares, they were separated by a 100 ms interstimulus interval. In the set size 2 condition, the squares were presented sequentially either at the same spatial location or at two different locations within a hemifield. On half of the trials, the test array was identical to the memory array; on the other half of trials, the colour of one square in the cued visual field of the test array was different from the colour of the corresponding item in the memory array. At the end of each trial, participants were required to indicate whether the memory and test arrays were the same or different.

Average accuracy for one item ($M = 98\%$, $SD = 0.8$) was significantly higher than for two items, regardless of whether the two items appeared at the same location ($M = 94\%$, $SD = 1.0$), $t(11) = 5.418$, $p < .001$, or at different locations ($M = 84\%$, $SD = 1.0$), $t(11) = 14.990$, $p < .001$. The latter

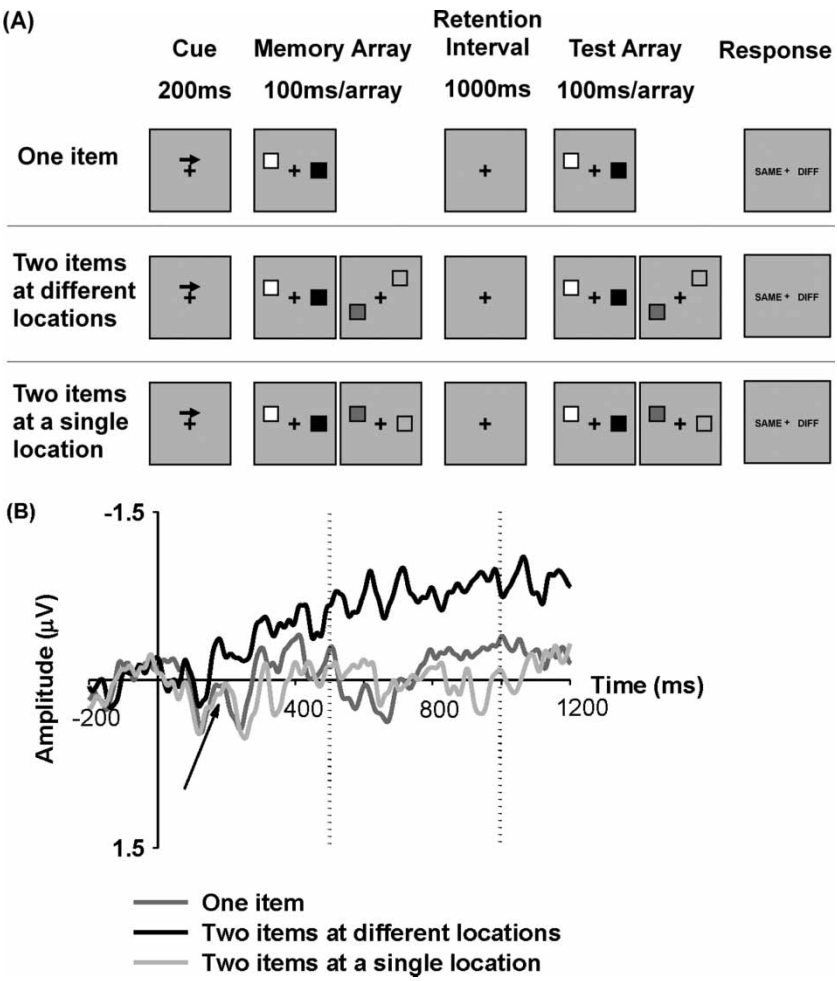

Figure 1. (A) Schematic diagram of one trial sequence. The varied grey scales of the squares in the memory and test arrays stand for different colours used in the actual experiment. The top panel is an example of a one-item no-change trial for the right hemifield. The middle panel is an example of a no-change trial for the right hemifield, with two items at different locations. The bottom panel is an example of a no-change trial for the right hemifield, with two items in a single location. (B) CDA waveforms collapsed across hemispheres. The arrow points to the onset of the second item (200 ms after the onset of the first one) when the memory array consisted of two items. The mean CDA amplitudes were calculated by averaging the waveforms over the interval between the two dotted lines (500–1000 ms after the onset of the memory array).

two conditions were also significantly different, $t(11) = 9.428$, $p < .001$: Performance was better when two items appeared at a single location than when they appeared at two different locations.

ERPs were time-locked to onset of the memory array and extended from –200 to 1200 ms. Data were filtered with a 40 Hz low-pass filter; artefacts detected in individual channels (fast average amplitude > 200 μV, different average amplitude > 100 μV, or zero variance) or segments (greater than 10 bad channels, eye movement or eye blink detected, eye threshold = 70 μV) were eliminated from subsequent analyses. Bad channels were replaced by data interpolated from surrounding electrode sites. Consistent with previous studies (e.g., Vogel & Machizawa, 2004), the CDA component at each electrode site was computed by subtracting ipsilateral activity from contralateral activity for each condition. Only correct trials were used for further analysis. CDA amplitude was measured as the mean voltage in a time window of 500–1000 ms after the onset of the first memory array. Activities from central, temporal and parietal electrode sites in each hemisphere were collapsed and averaged.

Amplitude measures (see Figure 1B) were submitted to a one-way (with three levels: One item, two items at a single location, two items at different locations) repeated measures ANOVA, which revealed a significant main effect, $F(1, 11) = 5.577$, $p = .038$. In subsequent pairwise t-tests, CDA amplitude in response to two items at two different locations was significantly larger than that for one item, $t(11) = 2.408$, $p = .035$, and for two items at a single location, $t(11) = 2.241$, $p = .047$. Moreover, the amplitudes in response to one item and to two items in a single location were not significantly different from each other.

In summary, the present results suggest that the amplitude of the CDA—previously suggested to reflect the number of objects held in VSTM—may instead be sensitive to the number of spatial locations attended in VSTM.

REFERENCES

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

McCollough, A. W., Machizawa, M. G., & Vogel, E. K. (2007). Electrophysiological measures of maintaining representations in visual working memory. *Cortex*, *43*, 77–94.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95.