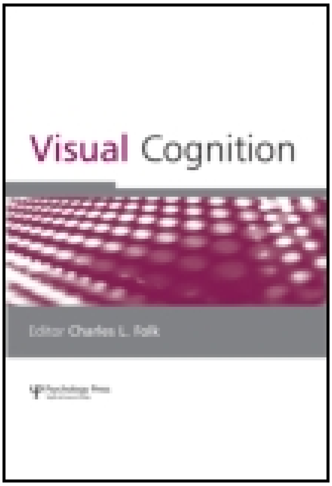

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pvis20

Are summary statistics enough? Evidence for the importance of shape in guiding visual search

Robert G. Alexander[a], Joseph Schmidt[ab] & Gregory J. Zelinsky[ac]

[a] Department of Psychology, Stony Brook University, Stony Brook, NY, USA

[b] Institute for Mind and Brain, University of South Carolina, Columbia, SC, USA

[c] Department of Computer Science, Stony Brook University, Stony Brook, NY, USA

Published online: 06 Mar 2014.

To cite this article: Robert G. Alexander, Joseph Schmidt & Gregory J. Zelinsky (2014) Are summary statistics enough? Evidence for the importance of shape in guiding visual search, Visual Cognition, 22:3-4, 595-609, DOI: 10.1080/13506285.2014.890989

To link to this article: http://dx.doi.org/10.1080/13506285.2014.890989

Visual Cognition, 2014
Vol. 22, Nos. 3–4, 595–609, http://dx.doi.org/10.1080/13506285.2014.890989

Routledge
Taylor & Francis Group

Are summary statistics enough? Evidence for the importance of shape in guiding visual search

Robert G. Alexander[1], Joseph Schmidt[1,2], and Gregory J. Zelinsky[1,3]

[1]Department of Psychology, Stony Brook University, Stony Brook, NY, USA
[2]Institute for Mind and Brain, University of South Carolina, Columbia, SC, USA
[3]Department of Computer Science, Stony Brook University, Stony Brook, NY, USA

(*Received 12 September 2013; accepted 30 January 2014*)

Peripheral vision outside the focus of attention may rely on summary statistics. We used a gaze-contingent paradigm to directly test this assumption by asking whether search performance differed between targets and statistically-matched visualizations of the same targets. Four-object search displays included one statistically-matched object that was replaced by an unaltered version of the object during the first eye movement. Targets were designated by previews, which were never altered. Two types of statistically-matched objects were tested: One that maintained global shape and one that did not. Differences in guidance were found between targets and statistically-matched objects when shape was not preserved, suggesting that they were not informationally equivalent. Responses were also slower after target fixation when shape was not preserved, suggesting an extrafoveal processing of the target that again used shape information. We conclude that summary statistics must include some global shape information to approximate the peripheral information used during search.

Keywords: Summary statistics; Visual search guidance; Gaze contingent; Eye movements; Extrafoveal processing; Shape.

Please address all correspondence to Gregory Zelinsky, Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, USA. E-mail: Gregory.Zelinsky@stonybrook.edu

We thank Thitapa Shinaprayoon, Harneet Sahni, and Samuel Levy for their help with data collection, and Ruth Rosenholtz, Krista Ehinger, Shuang Song, and all the members of the Stony Brook Eye Cog lab for invaluable discussions. This work was supported by NSF grants [IIS-1111047] and [IIS-1161876 to G.J.Z], and NIH Grant [R01-MH063748] to G.J.Z.

INTRODUCTION

Peripheral vision is qualitatively different from foveal vision (e.g., To, Gilchrist, Troscianko, & Tolhurst, 2011). This is perhaps most evident in the phenomenon of crowding, in which a peripherally-presented object becomes difficult to identify when it is near other objects (Pelli, 2008; Whitney & Levi, 2011). A plausible explanation for crowding – as well as other phenomena of peripheral vision – is that the visual system may represent information in the periphery as summary statistics (e.g., Balas, Nakano, & Rosenholtz, 2009; Freeman & Simoncelli, 2011; Greenwood, Bex, & Dakin, 2009; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). When objects are too close, multiple objects may be included in the same statistical representation, resulting in an intermingling of information between central and flanking objects and lessened discrimination ability. This lack of precision in peripheral vision has been implicated in scene recognition and gist perception (Oliva & Torralba, 2006), and likely includes averages of brightness, orientation, size, skew and kurtosis, and the emotion and gender of faces (see Alvarez, 2011; Haberman & Whitney, 2012).

Recently it has been suggested that a summary representation of peripheral information might also explain aspects of visual search behaviour, a view best exemplified by the Texture Tiling Model (Balas et al., 2009; see also Rosenholtz, Huang, & Ehinger, 2012; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012). Models of visual search have long proposed that peripherally viewed patterns, when they have not yet been attended, are represented in terms of unbound simple visual features that are "free floating" over space (Treisman, 1988), and that these features are pre-attentively available and can be used to guide attention and eye movements to likely target locations (Wolfe, 1994). The Texture Tiling Model builds on this idea by using a texture synthesis algorithm (Portilla & Simoncelli, 2000) that inputs an unaltered original image (depicting arbitrary objects or scenes) and a seed image (typically a patch of white noise), then iteratively alters the seed image to match the summary statistics of the original – the new synthesized image is therefore equated to the original with respect to the feature statistics, but the spatial relations of these features to each other are broken.[1]

[1] Methods of computing summary statistics differ in the degree to which they break the spatial relationships between features; this breakage is pronounced in the single pooling region version of the Texture Tiling Model but far less so in more recent methods that employ multiple pooling regions (Freeman & Simoncelli, 2011; Rosenholtz, Huang, & Ehinger, 2012). Moreover, to the extent that higher-order statistics are used to compute summary representations, this breakage will never be entirely complete. However, while it is true that most methods do preserve spatial statistics to varying degrees, it is also true that these methods must discard some spatial relationship information if they are to explain the core phenomena of interest; if spatial information was preserved completely, the swapping of nearby features believed to largely determine perception in the visual periphery should not occur.

Tests of this model have used correlational designs relating present/absent target detection performance in a time-unlimited foveal detection task to performance in a time-limited search task where targets appeared in the visual periphery. In the foveal detection task, patches containing the target and distractors, or just distractors, were synthesized and target detection accuracy was assessed and correlated with accuracy from a variety of search tasks using non-synthesized objects. If the intermingling of object features makes it harder to detect targets in synthetic images, then eye movements or shifts of attention, thought to increase precision and the use of local features, would be needed to avoid target detection errors in the search version of the task. This is precisely what was found: As performance in the search task improved, so too did the detection of targets in the foveal task (Rosenholtz, Huang, Raj, et al., 2012).

However, the existing evidence relating summary statistics to search is lacking in several respects. First, the correlational nature of this work raises an obvious concern – even if summary statistics are sufficient to search, search may use different features that happen to correlate with the summary statistics. Second, if the *only* information used in peripheral vision is summary statistics, task performance should be *identical* for peripherally-presented original and statistically-matched images, not just in accuracy, but in manual reaction time and oculomotor measures as well – concrete predictions that have never been tested. Finding that additional time is required to recognize peripherally-viewed synthetic images would suggest that these images are not only missing needed information, but that some mechanism is available to recover this information so as to ultimately perform the task. Third, models of summary statistics may make predictions that are inconsistent with findings in the search literature. The clearest example of this is efficient conjunction search. Targets defined by a conjunction of features can be found more quickly than what would be predicted by a random selection of search objects, and adding features to the conjunction target makes search guidance more efficient, not less (Wolfe, 1994). This would not be possible if the features of a target and distractors were scattered over space. More generally, to the extent that models are able to capture the spatial relationships between different features and use this information to predict efficient search guidance, they will not be able to also account for crowding and related phenomena in peripheral vision (and vice versa). This relationship, however, has not been addressed.

Is search guided by features in specific spatial relationships, or just a statistical summary of those features? We directly addressed this question by placing targets and their statistically-matched synthetic counterparts – generated using the single pooling region version of the Texture Tiling Model (Balas et al., 2009) – in visual search arrays and measuring how often these objects were first fixated during search. These immediate object fixations are a conservative, directly observable, and accepted measure of search guidance (Yang & Zelinsky, 2009; Alexander & Zelinsky, 2011; Maxfield & Zelinsky, 2012), one that avoids the need to infer guidance efficiency from search slopes (Zelinsky & Sheinberg, 1997). Moreover, this measure is perfectly suited to the question at hand, as evidence for the preferential selection and fixation of an object must be based on a pre-attentive analysis of that object in the visual periphery. This is important because attending to a peripherally-viewed object might change the summary statistics that are computed for that object (for a recent discussion, see Haberman & Whitney, 2012).

To determine whether differences in oculomotor behaviour exist between targets and their synthesized versions we adopted a gaze-contingent display change paradigm. This highly influential paradigm was introduced by McConkie and Rayner (McConkie & Rayner, 1975, 1976; Rayner, 1975) to evaluate the information from parafoveal and peripheral vision that is used during reading, but has since been adopted by researchers to study related questions in search and scene perception (e.g., Nuthmann, 2013; see also Rayner, 1978, 1998, and 2009, for reviews). Regardless of the context, the basic experimental logic is the same: If altering information in the visual periphery (e.g., replacing the letters of text with all Xs) results in no change in oculomotor behaviour compared to an unaltered control, then one can conclude that the manipulated information was unimportant to the task (e.g., reading). The experimental logic used in the present study is essentially identical: If the information from the visual periphery used to guide eye movements during search can be characterized by a statistical summary, then search guidance should be unaffected by whether the target is synthetic or not, as the two would be informationally equivalent. If, however, unaltered (non-synthetic) targets are preferentially fixated relative to synthetic targets, this would demonstrate that the guidance process uses information that is not included in a statistical summary of the object's features, at least for the synthetic images used in this study. This would also be an indirect test of the single pooling region version of the Texture Tiling Model (Balas et al., 2009), as it was this model and the methods that it employs that was used to generate the synthetic objects used in this study.

METHOD

Participants

Twenty Stony Brook University students with normal or corrected-to-normal vision participated for course credit.

Stimuli and apparatus

Search displays consisted of one target image and three distractors, placed ~7.5° from central fixation in a square formation yielding one object per quadrant. One distractor was a "lure" from the same category as the target. Targets and lures were teddy bear images, as described in Alexander and Zelinsky (2012). Non-lure distractors were random category non-bear objects selected from the

Hemera® Photo-objects collection. Individual objects were resized to ~1.8° of visual angle and converted to greyscale so as to accommodate the computational method used to generate synthetic counterparts (described below).

The Portilla-Simoncelli texture synthesis method, a component of the Texture Tiling Model (Balas et al., 2009), was used to generate the feature-matched objects used in this study. The original source should be consulted for details (Portilla & Simoncelli, 2000), but briefly this method uses a steerable pyramid to take multi-scale linear filter decompositions of an image at many orientations, then computes local autocorrelation statistics, relative phase statistics, and the co-occurrence of wavelet responses across nearby pairs of positions, orientations, and scales. Combining these with the mean, variance, range, skew, and kurtosis of the pixel distribution results in approximately 700 summary statistics. A synthetic version of the original image is created by iteratively projecting these statistics onto a seed image, which results in a new image having the same summary statistics as the original. This method is completely deterministic, although different synthetic images can be obtained from the same original image by projecting the summary statistics onto different seed images.

We tested two varieties of seed images, both of which have been used in previous work (e.g., Balas et al., 2009). One seed consisted of a canvas filled with white noise (a “noise-seed”). Noise-seeds often result in wrap-around (see Balas et al., 2009), an artefact of the synthesis algorithm (owing to the confinement of synthesized features to a torus) resulting in the spreading of synthetic patterns beyond the bounds of the canvas and continuing on the opposite side of the image. To minimize wrap-around, and any task performance differences that might accompany it, the canvas size was expanded to 128 × 128 pixels, over 125% the size of the original teddy bear images. Several different noise-seeds were also used to generate slightly different synthetic versions of each teddy bear to allow for the selection of stimuli that were roughly centred and had no obvious wrap-around. Note that these precautions should only serve to improve search guidance to noise-seed synthetic targets and should work against us finding any differences between these targets and the originals. The second type of seed consisted of a Gaussian-filtered version of the original image (a “shape-preserving” seed), to which Gaussian white noise was added. Shape-preserving seeds tend to naturally minimize wrap-around by causing the pixels of the bear (as opposed to the white pixels surrounding the bear) to spatially cluster (Balas et al., 2009). The most salient difference resulting from the use of noise-seeds and shape-preserving seeds is that shape-preserving seeds tend to produce synthetic images that coarsely approximate the global shape of the original objects. Examples of both seed types and the resulting synthetic images are shown in Figure 1.

Gaze position was recorded using an EyeLink® 1000 eyetracker sampling at 2000 Hz using a 9-sample velocity/acceleration model. Participants sat 68.7 cm from a CRT monitor (1024 × 768 screen resolution operating at 120 Hz), and registered their manual responses using a gamepad controller.

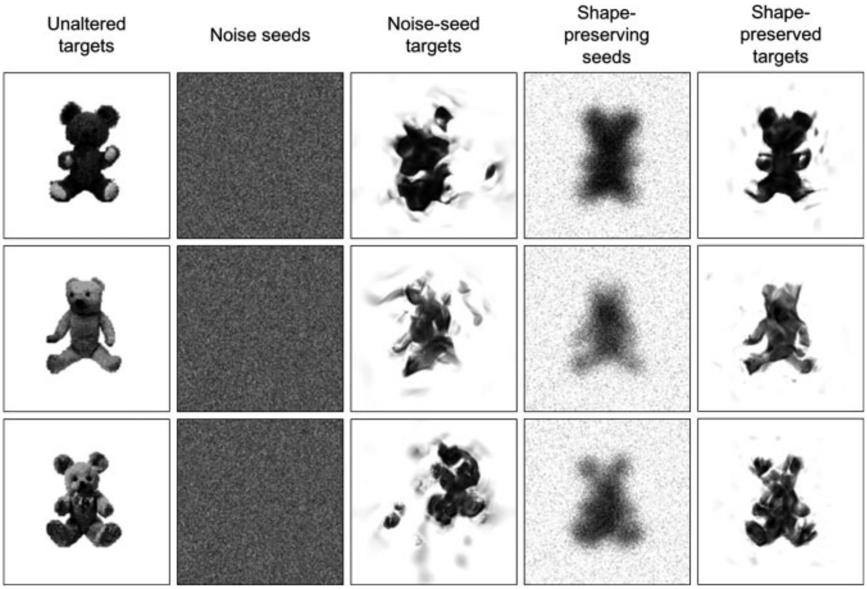

Figure 1. Examples of unaltered targets, seeds, and synthetic images generated from each seed. See text for additional details.

Procedure

There were five randomly interleaved within-participant conditions: An *unaltered* condition in which no synthetic images appeared, *noise-seed target* and *noise-seed lure* conditions in which a noise-seed was used to generate synthetic versions of the target or lure object, respectively, and *shape-preserved target* and *shape-preserved lure* conditions, identical to the noise-seed conditions except for the use of a shape-preserving seed. In all but the unaltered condition, the target or lure in the search display initially appeared in its synthetic version, but was replaced with the original version during the first saccade after search display onset. This gaze-contingent change was executed when eye velocity reached 42°/s, and was completed an average of 14 ms later while the eyes were still in motion. In the unaltered condition, the target exactly matched the preview throughout the trial.

Figure 2 summarizes the experimental procedure. Each trial was participant-initiated and began with a one second presentation of a target preview (always unaltered). This was followed by a central fixation dot that had to be fixated for at least 100 ms before the search display would appear. A target appeared in each search display, and the participant's task was to fixate it and press a button. There were 20 practice trials and 120 experimental trials, 24 per condition. After the experiment, a questionnaire was administered to assess whether participants were

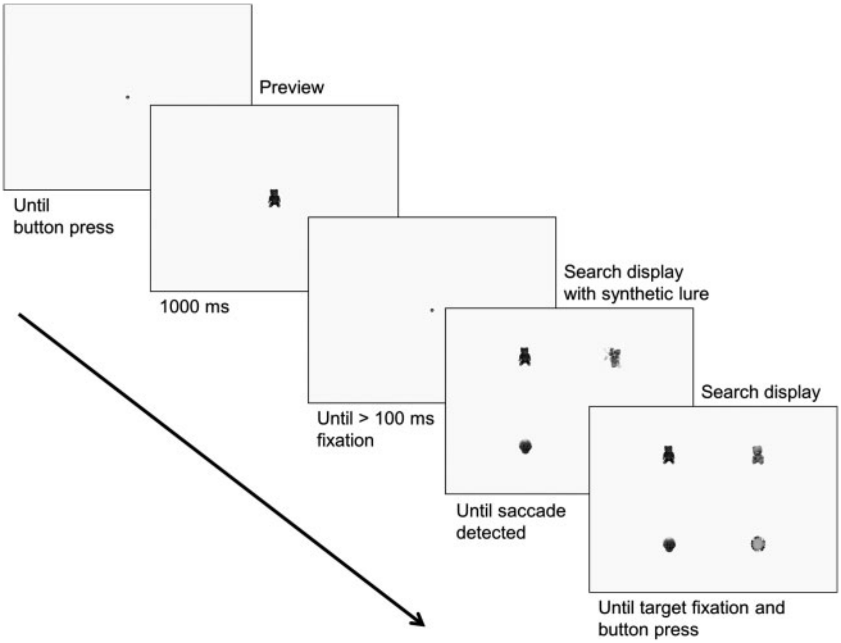

Figure 2. Procedure illustrating a trial from the noise-seed lure condition.

aware of the gaze-contingent display changes or the synthetic objects. Participants were told that there were two versions of the experiment, one in which the bear target shown at preview sometimes appeared distorted or weird for a brief moment during the search display and another in which this did not occur, and they were asked which version of the experiment they thought they had participated in. This was done to minimize under-reporting of awareness of the display changes, while not revealing the actual existence of these changes which would certainly have inflated the frequency of their report. These initial questions were followed by questions asking more explicitly about the display change manipulation, such as, “Did you notice ANY bears change?”. No participant reported noticing the gaze-contingent changes. Finally, participants were informed that gaze-contingent changes did occur and were shown examples of the synthesized targets. Here too, no participant reported seeing the synthetic objects.

RESULTS

Comparisons to a chance baseline were conducted using a one-sample *t*-test with a chance level of 0.25, reflecting the random direction of gaze to a search object. All other analyses used linear mixed effects modelling (LME, Baayen, Davidson &

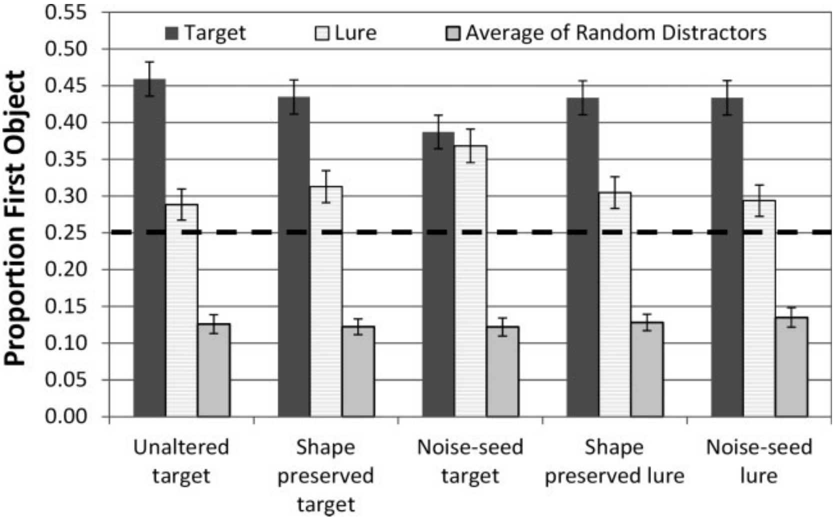

Figure 3. Proportion of trials in which the target (dark grey bars), lure (light grey bars), or non-lure distractor (medium grey bars) was the first object fixated in each of the five experimental conditions. The horizontal dashed line indicates the level of preferential fixation predicted by chance, and error bars indicate one standard error of the mean.

Bates, 2008) or logit mixed effects modelling (Jaeger, 2008) in R (R Development Core Team, 2012). These techniques were used because our measure of guidance – whether or not an object is fixated first – is binomial, and ANOVA is not appropriate for analysing binomial data (Agresti, 2002; Jaeger, 2008). Moreover, mixed effects modelling tends to be more powerful than ANOVA (see Luke & Christianson, 2011), and unlike ANOVA, LME skips the omnibus analysis and makes individual pairwise comparisons between all conditions and a designated baseline, removing the need for post-hoc *t*-tests. This has the added advantage of making the statistic immune to the inclusion of conditions that are not significantly different from one another, whereas including these conditions in an omnibus ANOVA could lead to a non-significant result. We fit the intercepts and slopes for the participant-by-condition random effects, and included in the final models those slopes that contributed to better fits, as indicated by likelihood ratio tests (Baayen, et al., 2008). For all non-binomial measures, *p*-values were obtained using Markov Chain Monte Carlo (MCMC) simulations.

Figure 3 plots proportions of immediate fixations by object type for each of the five experimental conditions. Error trials were fewer than 6% in all conditions and were removed from further analysis. In all conditions, the target was fixated first significantly more often than chance would predict, all $t(19) \geq 4.81$, all $p \leq .001$, suggesting that there is sufficient information even in noise-seed targets to guide search. More critically, the noise-seed target was less likely to be fixated first relative to unaltered targets (β = 0.28, *SE* = 0.14, $z = -2.09$, $p = .04$), but first object fixations did *not* differ between the unaltered and the shape preserved target conditions or either of the lure conditions ($\beta \geq 0.05$, $SE \geq 0.13$, $z \leq -0.39$, $p \geq .55$). This suggests that noise-seed targets are missing information that is useful for search guidance that is retained by shape-preserved targets, and more speculatively, that shape-preserved targets are informationally equivalent to unaltered targets.

Comparing guidance to the lure with guidance to the target provides additional information about information quality in peripheral vision. If a synthetic target only weakly matches the target preview, gaze may be preferentially directed to the lure instead, which shares the target category and likely has target-similar visual features. Except for the noise-seed lure condition (which was numerically but not significantly above chance), the lure was fixated first more often than chance, all $t(19) \geq 2.08$, all $p \leq .05$; noise-seed lure, $t(19) = 1.97$, $p = .06$, suggesting that the lure was a reliable attracter of gaze. Yet, when the *target* was generated using a noise seed, the *lure* was fixated more often than lures in the unaltered condition (β = 0.37, *SE* = 0.14, $z = 2.59$ $p = .01$), and indeed in all conditions other than when the target shape was preserved ($\beta \geq 0.28$, $SE \leq 0.14$, $z = 2.01$, $p \leq .04$; shape-preserved target, β = 0.251, *SE* = 0.14, $z = -1.80$, $p = .07$). This raises the intriguing possibility that categorical guidance, indicated here by guidance to a lure, may be mediated by information approximated by a noise-seed synthetic target. Finally, to test whether the synthesis method was creating some target-dissimilar artefact that might cause eye movements not to be directed to the synthetic targets, we compared the first object fixation rates between lures in the unaltered condition and either noise-seed or shape-preserved lures. If such an artefact existed, immediate fixations on synthetic lures should be less frequent than those on unaltered lures because the synthetic lure would presumably also share the artefact, creating the mismatch to the guiding target representation. However, this analysis revealed no reliable differences ($\beta \geq 0.25$, $SE \geq .14$, $z \leq 0.17$, $p \geq .55$), suggesting that guidance patterns were not driven by the presence of some oddity introduced by the synthesis method (regardless of seed type).[2]

To explore the potential for later search guidance we analysed the time between search display onset and when the target was first fixated (time-to-target). Time-to-target was longer in the noise-seed target condition compared to unaltered targets (β = 52.00, *SE* = 11.08, $t = 4.69$, $p < .001$), providing

[2] In a related analysis we asked whether differences in first object fixations were due to a speed–accuracy trade-off. However, the time to fixate the first object was not reliably different between conditions, ($\beta \geq -3.50$, $SE \geq 4.51$, $t \leq 0.06$, $p \geq .43$), indicating that our evidence for guidance was not reflecting a trade-off between speed and accuracy.

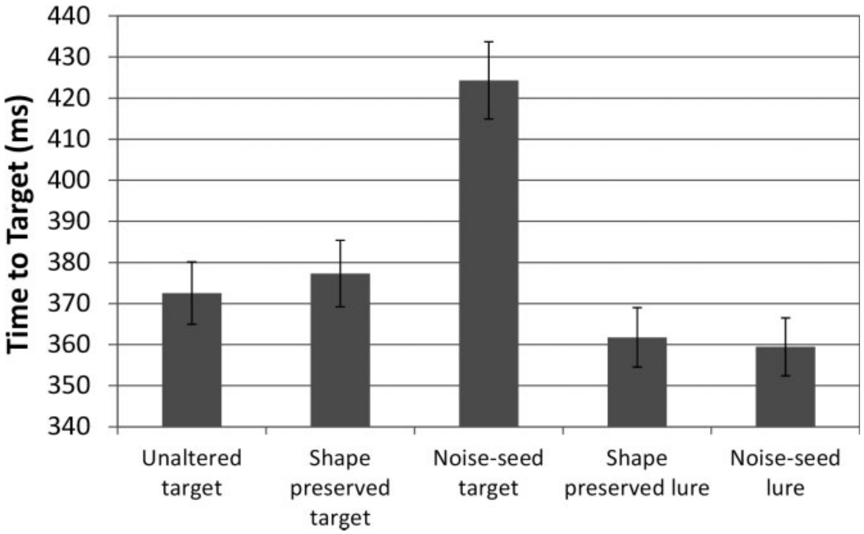

Figure 4. Time from search display onset to first fixation on the target for each of the five experimental conditions. Error bars indicate one standard error of the mean.

converging evidence that objects generated from noise-seeds are missing information used to guide search (Figure 4). Also consistent with the first-fixated analysis, time-to-target did not reliably differ between the unaltered target and the other conditions ($\beta \geq -15.67$, $SE \geq 11.09$, $t \geq -1.41$, $p \geq .15$), suggesting that the shape-preserved targets captured this missing information. Note that the noise-seed lure and shape-preserved lure conditions were included in this analysis, and appear in Figure 4, because it is possible that synthetic lures might have affected the time to fixate the target if lures were sufficiently non-bearlike and, consequently, no longer served as lures. This, however, proved not to be the case.

Does extrafoveally processing a synthesized search target lead to faster recognition of its unaltered counterpart after its fixation? To answer this question we analysed verification time – the time from first fixation on the target until the button response. Verification times were longer with noise-seed targets than unaltered targets ($\beta = 48.31$, $SE = 17.32$, $t = 2.79$, $p = .01$; Figure 5), despite the fact that these conditions differed only in terms of a peripherally-viewed synthetic target for ~153 ms, the average latency of the initial saccade following search display onset. Not only is this evidence for extrafoveal processing affecting later target verification, but it shows that this extrafoveal processing benefit is weaker in the case of a noise-seed target, presumably because it lacks information that might be useful in recognizing the unaltered target. This finding is consistent with predictions made by Rosenholtz, Huang, Raj, et al. (2012) and

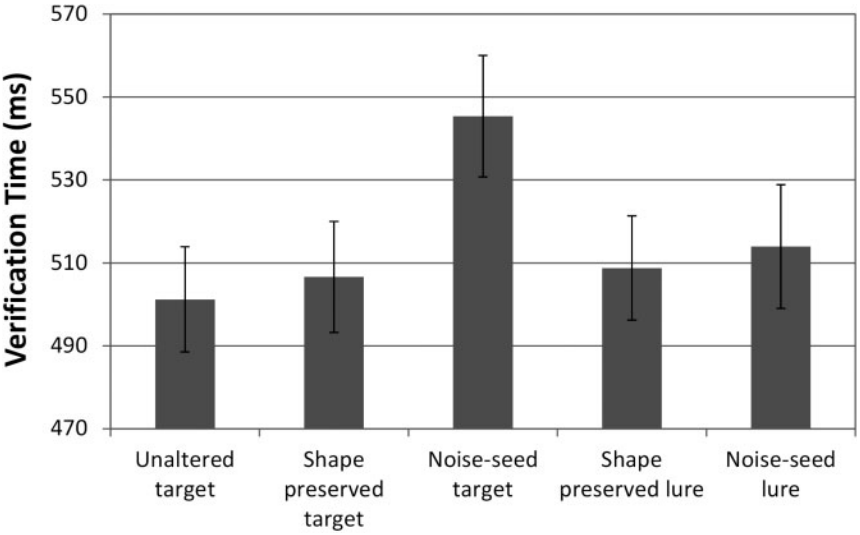

Figure 5. Time from first fixation on the target until the button press response for each of the five experimental conditions. Noise-seed and shape-preserved lure conditions were included here for consistency with the other figures, although no effect on target verification time was expected in these conditions. Error bars indicate one standard error of the mean.

their conjecture that objects are represented by summary statistics before attention is directed to a location, and not following the application of attention. Given that this extrafoveal processing benefit occurs after attention is directed to an object, one would therefore not expect the synthesized objects to match the unaltered objects, and indeed no such differences were found between unaltered targets and any of the other three conditions ($\beta \geq 5.11$, $SE \geq 17.30$, $t \leq 0.87$, $p \geq .38$), again suggesting rough informational equivalence between shape-preserved and unaltered objects. Note also that the noise-seed lure and shape-preserved lure conditions were again included in Figure 5, but this was done to maintain consistency with the other figures and no differences in verification time would be expected (and none were found).

DISCUSSION

Our results offer partial support for the use of summary statistics to guide search. To the extent that *only* summary statistics are available in the visual periphery, we should have found no guidance differences between unaltered and synthesized targets in our task. Whereas this proved to be the case for shape-preserved targets, we observed a significant decrease in guidance to noise-seed targets. Although it is not yet known whether this limitation of a noise-seed will generalize to free viewing tasks and scenes (but see Loschky, Hansen, Sethi, & Pydimari, 2010), our finding should serve as a cautionary note to studies assuming information equivalence between unaltered stimuli and stimuli synthesized from a noise-seed (e.g., Rosenholtz, Huang, Raj, et al., 2012). It might also be the case that summary statistics *are* adequate for describing search guidance, and that the difference reported here between unaltered and noise-seed targets reflects instead a failure of the current synthesis method to fully capture these statistics. However, while this cannot be ruled out, the fact that this method, when combined with shape, was largely successful in producing a synthetic target capable of strong search guidance argues against this possibility. Global shape, an arguably non-summary statistic, was probably responsible for the observed difference.

The fact that none of our participants reported seeing synthetic objects in our gaze-contingent paradigm, despite their presence on 4/5ths of the trials, is also telling, and suggests that the information available from synthetic images matches reasonably well the information available from peripheral vision. This is consistent with Freeman and Simoncelli's (2011) finding that observers could not discriminate unaltered scenes from synthetic scenes that were generated by a similar method using noise-seeds. However, caution should again be exercised when interpreting such demonstrations, as the features used to guide visual search may be different from those underlying conscious perception (e.g., Nagy & Sanchez, 1990; Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992; Wolfe et al., 2011). Our data extend these findings by showing that search guidance uses information from the visual periphery not captured by noise-seed targets, even though observers failed to report seeing strange, weird, or distorted objects upon explicit post-experiment questioning. As the code from newer texture synthesis methods (e.g., Freeman & Simoncelli, 2011) become publically available, claims about original and synthesized versions of images being metamers can be evaluated in terms of highly sensitive oculomotor measures using the same gaze-contingent paradigm described in the present study.

Of equal theoretical importance is our finding of comparable search guidance between unaltered and shape-preserved synthetic targets. The features used to guide search are still largely unknown (Wolfe & Horowitz, 2004), and this is particularly true for real-world objects (Zelinsky, 2008). However, the fact that these two types of targets produced similar guidance suggests that a specification of these features may be within reach; a summary statistical representation of a target, combined with global shape information (as approximated by a shape-preserving seed), may capture the most relevant feature dimensions for guiding search. Yet again, however, caution must be exerted. Colour is important for guidance (e.g., Hwang, Higgins, & Pomplun, 2009), and this contribution was not evaluated in the present study; as methods for synthesizing colour objects are developed, these will need to be tested. In addition, the Portilla and Simoncelli (2000) texture synthesis algorithm is only one method for capturing and visualizing summary statistics, and the use of other algorithms (or modifications of this algorithm) may produce different results, as could the use of target classes other than teddy bears. Finally, our design allows for the possibility that a shape-preserved target contains *more* information than what would have been needed to match guidance to unaltered targets. Future work will need to better specify the additional features captured by a shape-preserved target so as to better understand the exact information used to guide search.

At issue here is whether shape itself might be considered a form of summary statistic. Very recent work using multiple pooling regions has suggested that this may be the case (Freeman & Simoncelli, 2011; Rosenholtz, Huang, & Ehinger, 2012). Rather than accumulating statistics over a single pooling region, akin to information pooling over a single receptive field, information about global shape may be preserved if it is accumulated over multiple and overlapping pooling regions. This possibility is reminiscent of a quandary faced by researchers studying the coding of saccade targets by the superior colliculus. Collicular movement fields, even those near the fovea, are too large to explain the high spatial precision of saccade targeting. However, if a system uses information from multiple overlapping movement fields, high precision would be predicted (Sejnowski, 1988; see also McIlwain, 1986, and Rousselet, Husk, Bennett, & Sekuler, 2005). A similar explanation may apply here; the integration of feature information over multiple overlapping pooling regions may allow for the recovery of spatial relationships between these features – and therefore, shape. Whether this form of shape coding should be considered a higher-order summary statistic is a topic that this research community will need to engage.

REFERENCES

Agresti, A. (2002). *Categorical data analysis* (Vol. 359). Chichester: John Wiley & Sons.

Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision*, *11*(8), 1–15.

Alexander, R. G., & Zelinsky, G. J. (2012). Effects of part-based similarity on visual search: The Frankenbear experiment. *Vision Research*, *54*, 20–30.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12):13, 1–18.

Chubb, C., Nam, J. H., Bindman, D. R., & Sperling, G. (2007). The three dimensions of human visual sensitivity to first-order contrast statistics. *Vision Research*, *47*(17), 2237–2248.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201.

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences*, *106*(31), 13130–13135.

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). Oxford: Oxford University Press.

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, *9*(5), 1–18.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimari, T. (2010). The role of higher-order image statistics in masking scene gist recognition. *Attention, Perception & Psychophysics*, *72*(2), 427–444.

Luke, S. G., & Christianson, K. (2011). Stem and whole-word frequency effects in the processing of inflected verbs in and out of a sentence context. *Language and Cognitive Processes*, *26*(8), 1173–1192.

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition*, *20*(10), 1153–1163.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586.

McConkie, G. W., & Rayner, K. (1976). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (pp. 137–162). Newark, DE: International Reading Association.

McIlwain, J. (1986). Point images in the visual system: New interest in an old idea. *Trends in Neuro Sciences*, *9*, 354–358.

Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. *Journal of the Optical Society of America A*, *7*(7), 1209–1217.

Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition*, *21*(7), 803–837

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744.

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*, *18*(4), 445–451.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*(1), 49–70.

R Development Core Team (2012). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria. From http://www.R-project.org/

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*(1), 65–81.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, *85*, 618–660.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search, *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology*, *3*, 1–15.

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12*(4), 1–17.

Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2005). Spatial scaling factors explain eccentricity effects on face ERPs. *Journal of Vision*, *5*(10), 755–763.

Sejnowski, T. J. (1988). Neural populations revealed. *Nature*, *332*(6162), 308.

To, M., Gilchrist, I., Troscianko, T., & Tolhurst, D. (2011). Discrimination of natural scenes in central and peripheral vision. *Vision Research*, *51*(14), 1686–1698.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, *40*(2), 201–237.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 34–49.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501.

Wolfe, J. M., Reijnen, E., Horowitz, T. S., Pedersini, R., Pinto, Y., & Hulleman, J. (2011). How does our search engine "see" the world? The case of amodal completion. *Attention, Perception, & Psychophysics*, *73*(4), 1054–1064.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*, 2095–2103.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*(4), 787–835.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel–serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(1), 244–262.