

Visual similarity is stronger than semantic similarity in guiding visual search for numbers

Hayward J. Godwin · Michael C. Hout ·
Tamaryn Menneer

© Psychonomic Society, Inc. 2013

Abstract Using a visual search task, we explored how behavior is influenced by both visual and semantic information. We recorded participants' eye movements as they searched for a single target number in a search array of single-digit numbers (0–9). We examined the probability of fixating the various distractors as a function of two key dimensions: the visual similarity between the target and each distractor, and the semantic similarity (i.e., the numerical distance) between the target and each distractor. Visual similarity estimates were obtained using multidimensional scaling based on the independent observer similarity ratings. A linear mixed-effects model demonstrated that both visual and semantic similarity influenced the probability that distractors would be fixated. However, the visual similarity effect was substantially larger than the semantic similarity effect. We close by discussing the potential value of using this novel methodological approach and the implications for both simple and complex visual search displays.

Keywords Eye movements · Visual search

During visual search, we attempt to detect a target object in the environment, such as when looking for our mobile phone on a cluttered office desk. One of the most important questions regarding how search is performed pertains to how visual

attention is guided to examine items that resemble the target (Wolfe, Cave, & Franzel, 1989). Classic models of search focused on how this *guidance* process operates on the basis of the visual features of the stimuli, such as color, shape, and orientation (for a review, see Wolfe & Horowitz, 2004). More recently, researchers have shown considerable interest in exploring the extent to which *semantic* information might also guide search behavior. For example, when searching for a kettle, we tend to be more rapid at detection when it is placed on a kitchen counter, relative to when it is placed on the floor, demonstrating that high-level knowledge (e.g., regarding where kettles are likely to appear) may be able to guide search (for a review, see Oliva & Torralba, 2007). Accordingly, models of search have begun to be modified to incorporate routes by which semantic information can guide search behavior (Wolfe, Võ, Evans, & Greene, 2011).

Given the importance of understanding the role of high-level factors in search guidance, in the present study we examined the extent to which search is guided by two stimulus properties: the visual properties and the semantic properties. Is search guided to distractors that are visually similar to the target, or to those semantically similar to the target? To investigate this question, we employed a number search task, wherein people looked for a target number displayed among distractors. Semantic similarity was quantified by the *numerical distance* between the target and distractors. The numerical information conveyed by the visual stimulus is explicit and unambiguous semantic information about that item. If a digit is visually recognized as conveying numerical information, its semantic meaning has been processed. Numerical digits therefore provide a controllable stimulus space in which to manipulate semantic similarity and, as such, serve as an ideal stimulus set to explore the contributions of semantic similarity versus visual similarity in guiding search. In a recent study, Schwarz and Eiselt (2012) asked participants to search for the number 5 while controlling the other digits present in the

Electronic supplementary material The online version of this article (doi:10.3758/s13423-013-0547-4) contains supplementary material, which is available to authorized users.

H. J. Godwin (✉) · T. Menneer
School of Psychology, University of Southampton, Highfield,
Southampton, Hampshire SO17 1BJ, UK
e-mail: hayward.godwin@soton.ac.uk

M. C. Hout
New Mexico State University, Las Cruces, NM, USA

displays. They found that when the distractor digits were numerically close to the target, reaction times (RTs) were increased as compared to when the digits were numerically distant, suggesting that search is guided to distractors that are semantically similar to the target to a greater degree than to those that are semantically dissimilar to the target. Schwarz and Eiselt (2012) also conducted an additional experiment in which participants searched for the letter *S* (which is highly similar in appearance to the 5 but semantically unrelated). They found that presenting participants with displays containing distractor digits that were numerically close to the number 5 failed to slow search for an *S*, suggesting that visual similarity could not explain their results.

However, an outstanding question remains: What is the relationship between semantic guidance and guidance by visual properties? Given the overwhelming evidence showing that the visual characteristics of an object influence search (Wolfe & Horowitz, 2004), it is important to understand the interplay of visual and semantic features during search. In the present study, we went beyond Schwarz and Eiselt's (2012) work in two key ways. First, each digit (0 to 9) was employed as a target, to eliminate the possibility that the previous findings were the result of a peculiarity of the stimuli. Second, rather than attempt to equate the visual similarity of targets, we used multidimensional scaling (MDS) to obtain a psychologically tractable metric of the visual similarity among the numbers. Our approach enabled us to directly compare and contrast the relative influences of visual and semantic information in this task. In short, we sought to map out a more general picture of how both visual and semantic similarity influence search behavior.

We recorded the eye movements of participants as they searched for a single target digit among distractor digits (0 to 9 inclusive, excluding the target). Eye movements have been used in examining guidance processes in search in a number of prior studies. Specifically, participants tend to fixate objects that are visually similar to the target for a range of stimulus types (Becker, 2011; Luria & Strauss, 1975; Rayner & Fisher, 1987; Stroud, Menneer, Cave, & Donnelly, 2012; Williams, 1967), such as fixating blue and near-blue objects when searching for a blue target. In the present study, as we noted above, visual similarity was quantified using MDS, which is a tool for obtaining quantitative estimates of the similarity among groups of items (see Hout, Papesh, & Goldinger, 2012, for a review). MDS comprises a set of statistical techniques that take item-to-item similarity ratings and use data-reduction procedures to minimize the complexity of the similarity matrix. This permits a visual representation of the underlying relational structures that governed the similarity ratings. The output forms a similarity "map," within which the similarity between each pair of items was quantified. The appeal of this approach is that MDS is agnostic with respect to the underlying psychological structure that participants

have used to give their similarity ratings. For instance, when rating the visual similarity of numbers, people might appreciate the roundness or straightness of the lines, or the extent to which the numbers create open versus closed spaces. Even with no a priori hypotheses regarding the identity or weighting (e.g., perhaps "roundness" is more important than "open vs. closed spaces") of these featural dimensions, MDS has the ability to reveal any underlying structure in the output map. That is, by examining the spatial output, the analyst can intuit (and quantify) the dimensions by which participants provided their similarity estimates. By contrast, computational (i.e., nonpsychological) methods for measuring similarity may quantify this construct through a pixel-by-pixel analysis, or in some other fashion, that does not necessarily capture the way in which the human visual system assesses visual similarity.

In the present study, the MDS output for digits 0 to 9 was used to quantify the distance between each item pair in visual similarity space. Our prediction was that the probability of fixating a distractor would increase with its visual similarity to the target. Numerical distance between targets and distractors was used to quantify the semantic similarity. We also predicted that the probability of fixating distractors would increase with semantic similarity to the target. In addition, a key question was the relative strengths of guidance from these two sources of information. To address this question, we examined fixation probability data using a linear mixed-effects model.

Method

Participants

A group of 21 participants from the University of Southampton completed the MDS rating prestudy procedure. A separate group of 30 participants (25 females, five males) from the University of Southampton took part in the main eyetracking visual search study (mean age = 20.8 years, $SD = 3.5$).

Apparatus

We recorded eye movement behavior using an Eyelink 1000 running at 1000 Hz. Viewing was binocular, though only the right eye was recorded. We used a nine-point calibration that was accepted if the mean error was less than 0.5° of visual angle, with no error exceeding 1° of visual angle. Drift corrections were performed before each trial, and calibrations repeated when necessary. We used the recommended default settings to define fixations and saccades: Saccades were detected using a velocity threshold 30° per second or an acceleration that exceeded 8000° per second-squared.

Stimuli were presented on a 21-in. ViewSonic P227f CRT monitor with a 100-Hz refresh rate and a $1,024 \times 768$ pixel resolution. Participants sat 71 cm from the computer display,

and their head position was stabilized using a chinrest. Responses (“target present” or “target absent”) were made using a gamepad response box.

Stimuli

The stimuli consisted of the digits 0 to 9 written using a standard Verdana font (as used by Schwarz & Eiselt, 2012). They were $0.8^\circ \times 1.2^\circ$ of visual angle in size. On each trial, 12 stimuli were selected at random, placed upon a virtual 5×4 grid, and “jittered” by a random distance in a random direction within their respective grid cells. Across all trials, the stimulus selection process was controlled so that each participant was presented with the same number of instances of each distractor.

MDS task procedure

Via a single trial of the *spatial arrangement method* of MDS (Goldstone, 1994; Hout, Goldinger, & Ferguson, 2013), participants were shown the digits 0 through 9, arranged in discrete rows, but with random item placement. Participants were instructed to drag and drop the images in order to organize the space, such that images that were closer in space denoted greater similarity.

Visual search task design and procedure

For the visual search task, the targets were the digits from 0 to 9, inclusive, which resulted in ten different targets in total. Each participant searched for the same target digit throughout all of their 288 trials, which were preceded by 20 practice trials. Equal numbers of participants searched for each of the ten different targets (i.e., three participants searched for each target). A single target was presented on 50 % of the trials. Trials began with a drift correction procedure, after which participants were presented with a reminder of the target at the center of the display, which they had to fixate for 500 ms in order for the trial to begin. Following an incorrect response, a tone sounded.

Results

MDS results

The MDS data were analyzed using the *PROXSCAL* scaling algorithm (Busing, Commandeur, Heiser, Bandilla, & Faulbaum, 1997), with 100 random starts. In order to choose the most appropriate number of dimensions, a *scree plot* was created, which displays stress (a measure of the fit between the estimated distances in space and the input proximity matrices) as a function of dimensionality (see Fig. 1). A useful heuristic is to find the “elbow” in the plot: the stress value at which added dimensions cease to substantively improve fit

(Jaworska & Chupetlovska-Anastasova, 2009). Our data show a clear elbow at Dimension 2; therefore, our MDS solution was plotted in two dimensions.

Figure 2 shows the results of the MDS analysis (also, in the supplementary materials, Table S1 provides the two-dimensional coordinates for each item, and Table S2 reports the distances in MDS space between each pair of items). No basic unit of measurement is present in MDS, so the interitem distance values are arbitrary and only meaningful relative to other item pairings from the space. One potential criticism of obtaining visual similarity measures via MDS is that observers may be unable to ignore semantic information. However, we found no significant correlation between visual and semantic similarity measures for the digits ($r = -.06, p = .35$), suggesting that semantic information did not influence the visual similarity ratings.

Behavioral analyses

Consistent with the simplicity of the task, response accuracy, measured as the proportions of correct responses, was high (target-present trials: $M = .96, SD = .04$; target-absent trials: $M = .99, SD = .01$). The target-absent median RTs (correct trials) were significantly longer than the target-present RTs [$M = 1,265$ ms, $SD = 375$, and $M = 783$ ms, $SD = 159$, respectively; $t(29) = 10.9, p < .0001$], which is to be expected in visual search tasks (Chun & Wolfe, 1996).

Examining the influences of visual similarity versus semantic similarity

In order to compare the influences of visual similarity to the target and semantic similarity to the target in determining the probability that objects would be fixated, we constructed a linear mixed-effects model (LME: Bates, Maechler, & Bolker, 2012). We adopted this approach because LMEs allow for variation in effects based on random factors (here, variation between different participants, different targets, and different distractors), and, more importantly, because LMEs are versatile when faced with examining data sets in which unequal numbers of observations are entered into different cells within the analyses, as was the case here.

We began with a basic LME model that lacked the main effects of either visual or semantic similarity. As a dependent variable, we coded whether each distractor was fixated. Since this was a binary variable (i.e., “fixated” vs. “not fixated”), we used a binomial model to analyze the data. Prior to the analysis, we removed the fixation data from incorrect-response trials, as well as any fixations that were shorter than 60 ms or longer than 1,200 ms in duration (~4 % of fixations were removed). After removals, the remaining data set comprised data from approximately 94,000 distractors regarding whether they were or were not fixated.

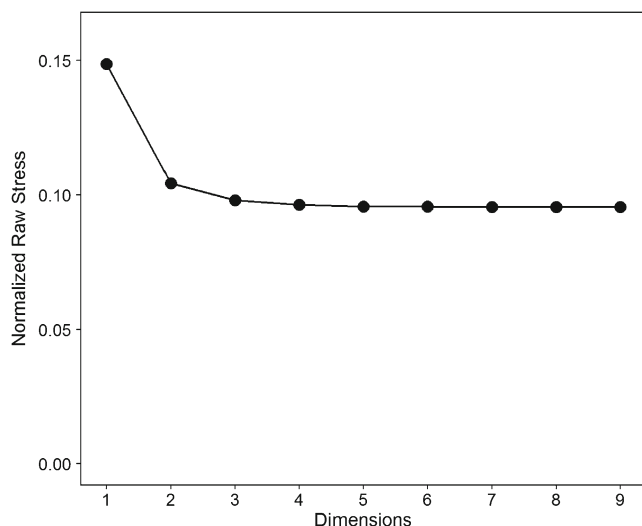


Fig. 1 Scree plot for the multidimensional scaling (MDS) data, showing stress plotted as a function of the dimensionality in which the MDS data were scaled

The random factors entered into the model were the participants, the different targets that the participants searched for, and the different distractors. The first fixed factor was Target Presence (i.e., target present or absent). The experimental factors of Visual Similarity to the Target and Semantic Similarity to the Target were added to this basic LME model in an iterative fashion, to determine the factors that improved the fit of the model to the data set. Visual similarity to the target was defined as the reverse coding of the MDS distance to the target (i.e., maximum MDS distance – the current MDS distance). In other words, increasing values on the visual similarity scale

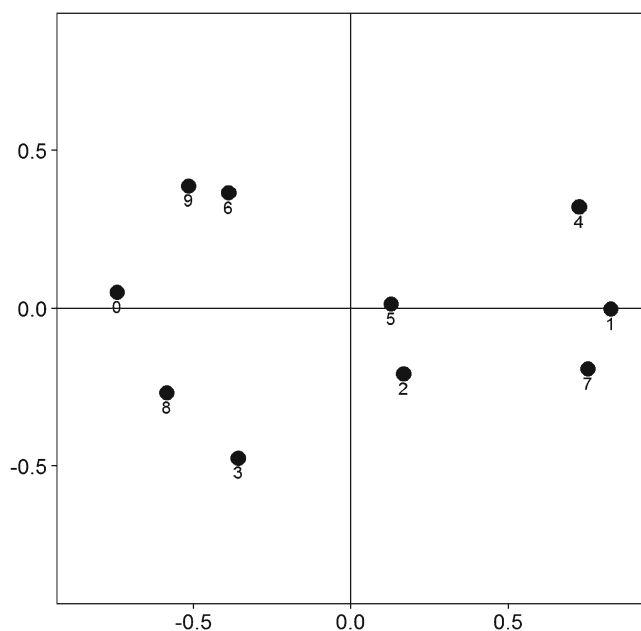


Fig. 2 Two-dimensional multidimensional scaling solution for the digit stimuli. The x- and y-axes represent the primary and secondary dimensions, respectively

meant that the distractors were increasingly similar to the target. Semantic similarity to the target was defined as the reverse coding of the numerical distance to the target (i.e., maximum numerical distance – the current numerical distance). This meant that increasing values on the Semantic Similarity scale indicated that distractors were increasingly similar to the target.

The final model reported in Table 1 is the model with the most optimal fit, and Fig. 3 presents the factors that reached significance in this final model. The p values were generated on the basis of posterior distributions for the model parameters obtained by Markov-chain Monte Carlo sampling.

The final LME model showed a number of interesting results. First, we found clear differences in the probabilities of fixating distractors in target-present versus target-absent trials, as evidenced by a significant effect of target presence. This result is not surprising: As we noted above, target-absent trials typically have longer RTs than target-present trials do (Chun & Wolfe, 1996), so, as a direct consequence, the probability that any object will be fixated is increased in target-absent versus target-present trials.

Second, we found a significant influence of visual similarity to the target upon the probability of fixating distractors during search; see Fig. 3. This finding is in line with current models of search that emphasize the important role that distractor-to-target visual similarity plays in determining whether or not an object will be fixated (Becker, 2011; Luria & Strauss, 1975; Stroud et al., 2012). The tendency for decreased RTs in target-present trials was also reflected by the emergence of an interaction between visual similarity to the target and target presence. This interaction was caused by a reduction in the slope for visual similarity in target-present relative to target-absent trials (slope for target-present visual similarity to target = 0.39, $SE = 0.04$, $z = 9.34$, $p < .0001$; slope for target-absent visual similarity to target = 0.47, $SE = 0.03$, $z = 8.08$, $p < .0001$).

Finally, we found a significant effect of semantic similarity to the target that influenced the probability that objects would be fixated during search; see Fig. 3. This finding supports the study conducted by Schwarz and Eiselt (2012) and extends their result to the full range of digits from 0 to 9 inclusive, and also demonstrates that the effect of semantic similarity can modulate eye movement behavior directly.¹

¹ There is some debate as to the mental representation of numerical values—for example, as to whether the subjective distance between adjacent numbers follows a linear or a logarithmic function (see, e.g., Cohen & Blanc-Goldhammer, 2012). A logarithmic function would cause, for example, the semantic difference between 1 and 2 to be greater than that between 8 and 9. However, in the present data, the values used in the regression were averaged over all pairs of numbers a certain distance apart (i.e., the distance of 1 includes the data for 1 vs. 2 as well as the data for 8 vs. 9). Combining data over these number pairs for every distance value causes any subjective disparity between small and large numbers to be averaged out.

Table 1 Fixed-effect estimates and results of statistical tests for the mixed linear model, including the slope, the standard error of the slope, and the z value

Effect	Model Coefficients	SEM	z
Intercept	-1.62	0.14	-11.52**
Target Presence	-0.93	0.04	-23.38**
Visual Similarity to the Target	0.48	0.03	16.79**
Semantic Similarity to the Target	0.05	0.01	11.00**
Target Presence \times Visual Similarity to Target	-0.14	0.05	-3.05*

* $p < .01$, ** $p < .0001$

Discussion

Classic models of visual search focused on how the visual features of the targets and distractors in a search task

influenced behavior and search performance (Wolfe & Horowitz, 2004). Recently, research has shifted toward examining the role that semantic information plays in visual search, particularly in the context of scene perception and real-world tasks (Wolfe et al., 2011). As a result of these advances, a growing argument has supported models of search that include a route by which semantic information can guide attention. Numerical digits provide a stimulus domain that is controllable and quantifiable in terms of semantic similarity, and therefore they are an ideal testing ground for exploring the influence of semantic information upon search behavior. Using visual search for numbers, Schwarz and Eisele (2012) found that when participants searched for the number 5, their RTs were increased when the average numerical distance between the distractors and the target was reduced. This result suggests that participants were spending extra time examining these distractors because they were semantically similar to the target.

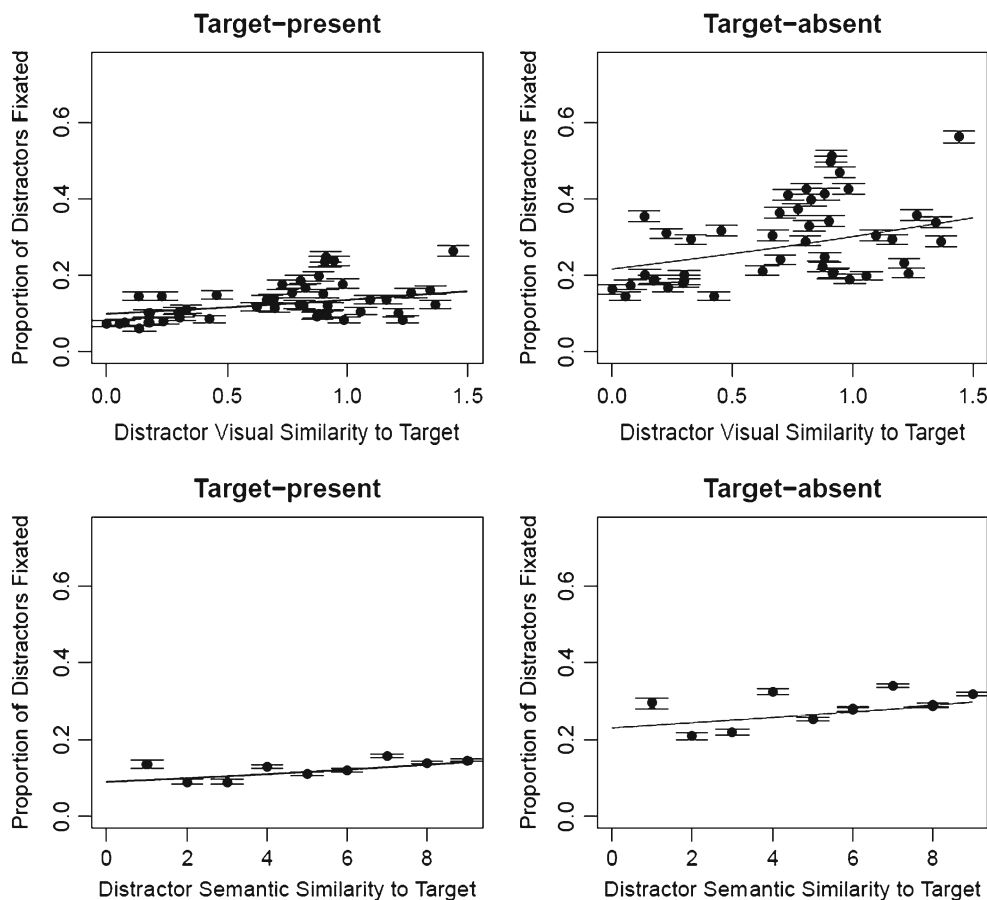


Fig. 3 Proportions of distractor objects fixated as a function of the visual similarity (upper row) and semantic similarity (lower row) to the target object, for both target-present (left column) and target-absent (right column) trials. For visual similarity, increases along the x -axis represent increased similarity between each object and the target. Each point on the graphs for visual similarity represents a single visual similarity value between a pair of digits. Given that we used ten different digits (0–9),

there are $(10 * 9)/2 = 45$ similarity values in total. For semantic similarity, increases along the x -axis likewise represent increased similarity between each object and the target. Each point on the graph represents a single semantic similarity value between a pair of digits. Given that numerical distance is standardized between pairs, data were collapsed across different pairs for a given similarity value (e.g., nine pairs had the maximum similarity of 9)

Here, we extended Schwarz and Eiselt's (2012) study as follows. First, we wanted to ensure that semantic information can guide search to other digits, aside from the number 5. This is important in terms of both replicating their results and establishing the fact that semantic information guides search more generally. To address this issue, we asked participants to search for a range of digits, not just the number 5. Second, given the evidence in favor of how visual features can guide search behavior, we also wanted to know the extent to which the visual similarity between digits may *also* influence search. This was an important novel development, because it enabled us to compare and contrast the roles of guidance by visual versus semantic information.

Participants were asked to search for the digits from 0 to 9, inclusive. The key dependent variable was the probability of fixating each distractor. If visual similarity guides search, participants should have been highly likely to fixate distractors that were visually similar to the target. Furthermore, if semantic information guides search, they should have been highly likely to fixate distractors that were a small numerical distance from the target (i.e., that were numerically or semantically similar to the target). Our analyses revealed that, in fact, *both* of these factors guided behavior, though the magnitudes of the two factors were quite different. This was evidenced by the difference in slopes for the two factors in the LMEs. The slope provides an indication of the extent to which the probability of fixating distractors varied with the independent variable. For visual similarity, the slope was approximately nine times larger than that for semantic similarity (0.48, as compared with 0.05).

Although searching for a number may be more “abstract” than searching real-world scenes, numbers do provide an environment in which semantic relationships can be controlled in an unambiguous fashion. The present research and the results obtained address important questions for visual search regarding guidance from semantic information. We have demonstrated under tightly controlled conditions (and using a novel combination of methodologies) that it is possible to define and contrast the contributions of visual versus semantic information in guiding search behavior. From this controlled environment, the results can be used to inform future investigation of more complex, scene-based tasks, in which multiple forms of high-level information converge to modulate search behavior (e.g., Wolfe et al., 2011). Indeed, there is growing evidence that, when searching real-world scenes, semantic information may play a much stronger role in guiding fixations than it did in the present study (see, e.g., Henderson, Malcolm, & Schandl, 2009). Our findings therefore add to this growing literature by mapping out the contributions of semantic and visual information when simple, controlled stimuli are utilized.

Finally, it should be noted that the present study serves as a proof of concept for a novel methodological approach:

namely, the use of MDS-derived similarity indexes in conjunction with eyetracking (see also Alexander & Zelinsky, 2011). It is worth noting that our MDS space is very much like one that was derived for numbers by Shepard, Kilpatrick, and Cunningham (1975) nearly 40 years ago, evidencing a stable mental representation for these stimuli, despite differences in the raters and methods of data collection (a comparison of the spaces is provided in the [supplemental materials](#)). Adopting our approach will likely benefit researchers wishing to study the role of semantic information in more complex visual search displays, as well as provide them with a stable and reliable metric to quantify visual similarity between more complex objects. Furthermore, this approach can not only help to quantify visual similarity, but also help to provide predictions regarding eye movement behavior during visual search. Future studies along these lines may also benefit from specifically controlling the visual similarity of objects—for example, by using different fonts and sizes.

Author note H. J. G. and T. M. were supported by funding from the Economic and Social Sciences Research Council (Grant No. ES/I032398/1). The authors thank Florence Greber and Rawzana Ali for their assistance with data collection.

References

- Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision*, *11*(8), 9. doi:10.1167/11.8.9. 1–15.
- Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using Eigen and Eigen++ [Software]. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity of perceptual difficulty? *PLoS One*, *6*, e17740. doi:10.1371/journal.pone.0017740
- Busing, F. M. T. A., Commandeur, J. J. F., Heiser, W. J., Bandilla, W., & Faulbaum, F. (1997). PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints. *Advances in Statistical Software*, *6*, 67–73.
- Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*, 39–78. doi:10.1006/cogp.1996.0002
- Cohen, D. J., & Blanc-Goldhammer, D. (2012). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, *18*, 331–338. doi:10.3758/s13423-011-0059-z
- Goldstone, R. L. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, *26*, 381–386. doi:10.3758/BF03204653
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, *16*, 850–856. doi:10.3758/PBR.16.5.850
- Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*, 256–281. doi:10.1037/a0028860
- Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2012). Multidimensional scaling. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*, 93–103. doi:10.1002/wcs.1203

- Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multi-dimensional scaling (MDS) and its utility in various psychological domains. *Tutorial in Quantitative Methods for Psychology*, 5, 1–10.
- Luria, S. M., & Strauss, M. S. (1975). Eye movements during search for coded and uncoded targets. *Perception & Psychophysics*, 17, 303–308. doi:10.3758/BF03203215
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11, 520–527. doi:10.1016/j.tics.2007.09.009
- Rayner, K., & Fisher, D. L. (1987). Letter processing during eye fixations in visual search. *Perception & Psychophysics*, 42, 87–100. doi:10.3758/BF03211517
- Schwarz, W., & Eiselt, A.-K. (2012). Numerical distance effects in visual search. *Attention, Perception, & Psychophysics*, 74, 1098–1103. doi:10.3758/s13414-012-0342-8
- Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, 7, 82–138. doi:10.1016/0010-0285(75)90006-7
- Stroud, M. J., Menneer, T., Cave, K. R., & Donnelly, N. (2012). Using the dual-target cost to explore the nature of search target representations. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 113–122. doi:10.1037/a0025887
- Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, 27, 355–360. doi:10.1016/0001-6918(67)90080-7
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419–433. doi:10.1037/0096-1523.15.3.419
- 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, 495–501. doi:10.1038/nrn1411
- Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, 15, 77–84. doi:10.1016/j.tics.2010.12.001