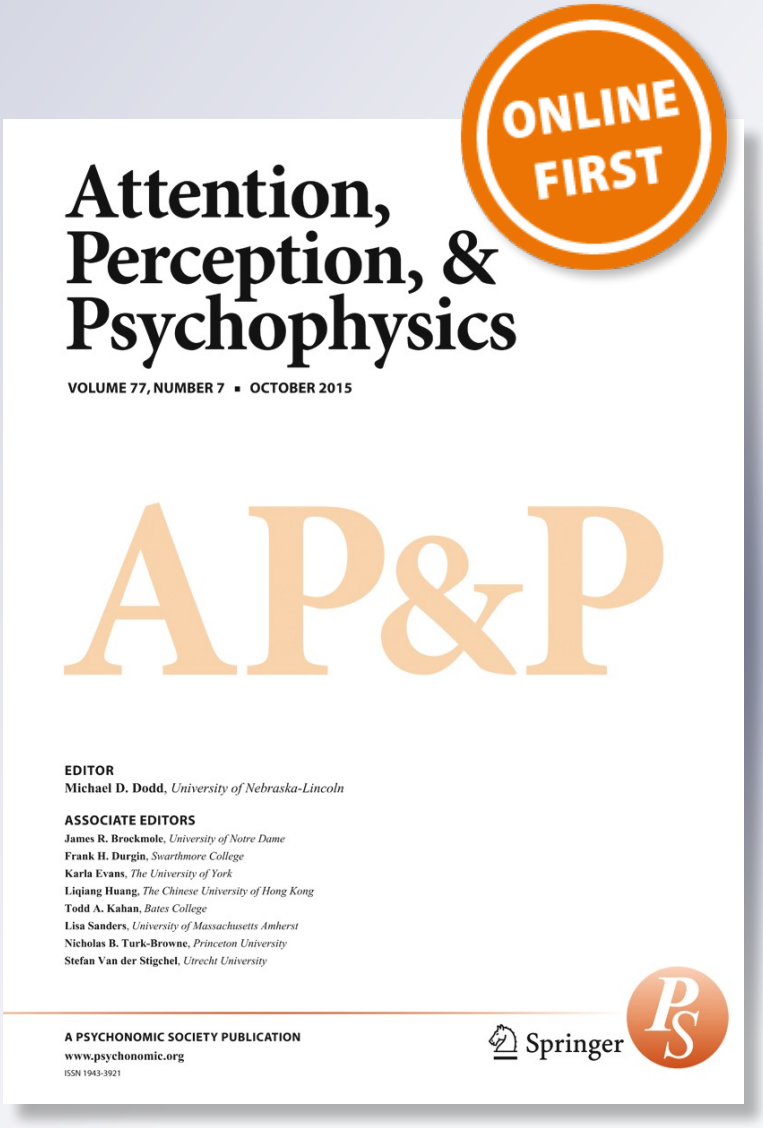
Using multidimensional scaling to quantify similarity in visual search and beyond

Michael C. Hout, Hayward J. Godwin, Gemma Fitzsimmons, Arryn Robbins, Tamaryn Menneer & Stephen D. Goldinger

Attention, Perception, & Psychophysics

ISSN 1943-3921

Atten Percept Psychophys
DOI 10.3758/s13414-015-1010-6



Springer

Atten Percept Psychophys
DOI 10.3758/s13414-015-1010-6

Using multidimensional scaling to quantify similarity in visual search and beyond

Michael C. Hout[1] · **Hayward J. Godwin**[2] · **Gemma Fitzsimmons**[2] · **Arryn Robbins**[1] · **Tamaryn Menneer**[2] · **Stephen D. Goldinger**[3]

Abstract Visual search is one of the most widely studied topics in vision science, both as an independent topic of interest, and as a tool for studying attention and visual cognition. A wide literature exists that seeks to understand how people find things under varying conditions of difficulty and complexity, and in situations ranging from the mundane (e.g., looking for one's keys) to those with significant societal importance (e.g., baggage or medical screening). A primary determinant of the ease and probability of success during search are the similarity relationships that exist in the search environment, such as the similarity between the background and the target, or the likeness of the non-targets to one another. A sense of similarity is often intuitive, but it is seldom quantified directly. This presents a problem in that similarity relationships are imprecisely specified, limiting the capacity of the researcher to examine adequately their influence. In this article, we present a novel approach to overcoming this problem that combines multidimensional scaling (MDS) analyses with behavioral and eye-tracking measurements. We propose a method whereby MDS can be repurposed to successfully quantify the similarity of experimental stimuli, thereby opening up theoretical questions in visual search and attention that cannot currently be addressed. These quantifications, in conjunction with behavioral and oculomotor measures, allow for critical observations about how similarity affects performance, information selection, and information processing. We provide a demonstration and tutorial of the approach, identify documented examples of its use, discuss how complementary computer vision methods could also be adopted, and close with a discussion of potential avenues for future application of this technique.

Keywords Methods · Similarity · Multi-dimensional scaling · Visual search · Eye-movements

During a typical visual search task, people are asked to detect a target embedded within a set of non-target (distractor) items. Visual search has been extensively studied as an independent topic of interest (Chan & Hayward, 2013; Palmer, Verghese, & Pavel, 2000; Wolfe, 1994, 1998, 2010), and remains one of the most widely-used methodologies for studying attention and visual cognition (Davis & Palmer, 2004; Evans et al., 2011; Treisman & Gormican, 1988; Wolfe & Horowitz, 2004). An enormous literature exists that seeks to understand how people find things, ranging from low-level shape search (e.g., finding rotated *T*s among rotated *L*s; Chun & Jiang, 1999; Dowd & Mitroff, 2013), to high-level scene perception, such as finding your favorite cereal on the shelves of a cluttered grocery store, or a familiar face in a crowd of people (Henderson, 2003; Henderson & Hollingworth, 1999; Hollingworth, Williams, & Henderson, 2001). Search is sometimes mundane, as when people perform laboratory tasks or look for their keys at home. But many searches are performed by professionals, with significant societal importance, such as the screenings performed by radiologists or airport security agents (Biggs & Mitroff, 2014; Godwin et al., 2010a, 2010b; Helbren et al., 2014; Menneer et al., 2009, 2012; Wolfe et al., 2007).

A crucial determinant of the ease and likelihood of success during search is the similarity between the background and the target that a person is trying to find (Duncan & Humphreys,

✉ Michael C. Hout
mhout@nmsu.edu

[1] Department of Psychology, New Mexico State University, P.O. Box 30001 / MSC 3452, Las Cruces, NM 88003, USA

[2] University of Southampton, Southampton, UK

[3] Arizona State University, Tempe, AZ, USA

Published online: 22 October 2015

1989). A bird-watcher, for example, will easily spot a cardinal (a bright red bird) on a lifeless tree in the dead of winter, but would have a much harder time spotting a brown thrasher on the same tree, as it will blend into the similar colors of its surroundings. Although this notion is widely appreciated, and although a sense of "similarity" might seem intuitive, it is seldom quantified directly. The problem for modern researchers is that visual similarity is an important but hard-to-quantify concept. For some stimulus characteristics (e.g., color, orientation), it is reasonably straightforward to quantify or manipulate similarity—for instance, by precisely varying the orientation of two stimuli—and doing so can foster insights into behavior. However, with more complex stimuli (e.g., real-world objects), the direct measurement and manipulation of features is not universally straightforward, and so quantification often relies on the consensus view of the researchers as to what constitutes a similarity dichotomy (similar vs. dissimilar).

In this article, we present a novel approach to overcoming this problem that combines multidimensional scaling (MDS) with behavioral measurements and eye-tracking. MDS is a statistical technique that—when applied to overt or indirect similarity judgments—can be used to uncover the dimensions by which people perceive similarity. Its application is widespread (Hout, Goldinger, & Ferguson, 2013, for a discussion), but in the psychological literature, it is most often used when researchers want to explore or confirm the features that make two things appear alike or different. We propose that MDS can be repurposed to quantify similarity successfully for many types of stimuli, and thereby open up theoretical questions in visual search and attention that cannot currently be addressed. More specifically, by conducting an MDS analysis on the stimuli that will be used in a visual search task, researchers can objectively quantify – in a continuous, rather than dichotomous fashion—the extent to which stimuli are (or are not) perceived as resembling one another. These quantifications can then be used in conjunction with simple accuracy or reaction time measures, as well as more temporally and spatially fine-grained eye-tracking metrics (e.g., fixation duration, scan-path ratio) to make critical observations about how similarity affects performance, information selection and information processing, and thereby provide novel insights into visual search. This technique is particularly useful when the stimuli of interest are complex, or high-dimensional, and therefore may be comprised of features that are unspecified or unknown *a priori*. Here, we provide a brief demonstration of this approach, identify documented examples of its application and previous successes, discuss the potential for computer vision methods to complement the use of MDS, and close with a brief discussion of potential avenues for the future use of this approach. Taken together, we will show that MDS enables not only careful stimulus control, but also facilitates the discovery of new insights into visual cognition.

Visual search and similarity

We have known for decades that similarity—the degree to which two objects resemble one another—plays a key role in modulating visual search behavior. Search is easy when the target is completely unlike all the distractors (so-called "*pop-out*" *or* "*feature search*"), and gets more difficult when the distractors share similarity with the target (so-called "*conjunction search*"; Treisman & Gelade, 1980). For example, in Fig. 1, the top panels represent easy searches because the targets are unique; there is only one blue item in the top-left panel, and only one sofa in the top-right. By contrast, the bottom panels demonstrate comparatively more difficult searches wherein the target is now similar to the distractors in some way; in the bottom-left panel, many *L*s share the color of the target, and on the bottom-right panel, all the distractors share an identity with the target, and must therefore be more closely scrutinized during search.

Accordingly, similarity is an important concept appreciated by several major theories of visual search. In Treisman's *Feature Integration Theory* (FIT; Treisman & Gelade, 1980), the difference between these efficient (feature) and inefficient (conjunction) searches is explained by the number of features shared by the target and distractors; i.e., how similar the target is to other items. Wolfe's *Guided Search* model (GS; Wolfe, Cave, & Franzel, 1989; Wolfe, 2001) extends upon FIT, by suggesting that focal attention (which is required during inefficient search) is judiciously guided. In GS, guidance is achieved via a master map of attentional allocation, and the prioritization of attention is determined in a top-down fashion. More specifically, attention is influenced by the similarity between features in the scene and the mental representation of what is being searched for (see also Zelinsky, 2008; Hout & Goldinger, 2015). *Attentional Engagement Theory* (AET; also referred to as "similarity theory"; Duncan & Humphreys, 1989) also relies strongly on the concept of similarity. AET posits that representations of objects in a scene compete with one another for entrance into visual short-term memory, and that this competition is biased in favor of objects that are similar to the target (see also Hwang, Higgins, & Pomplun, 2009). In both GS and AET, search is slower when targets are more similar to the distractors (because resources are wasted on the examination of non-targets), or when the distractors are less similar to one another. The AET (Duncan & Humphreys, 1989, 1992) is perhaps the theory that most explicitly relies on a concept of similarity, but empirical evidence abounds that similarity has an important role in multiple aspects of search behavior, such as guidance (the ability to direct attention to the location of the target) and object identification (the ability to appreciate that an inspected item is the sought-after target).

For example, Neider and Zelinsky (2006) investigated search behavior under conditions of target-background similarity. They asked participants to search for pictures of real-world objects in

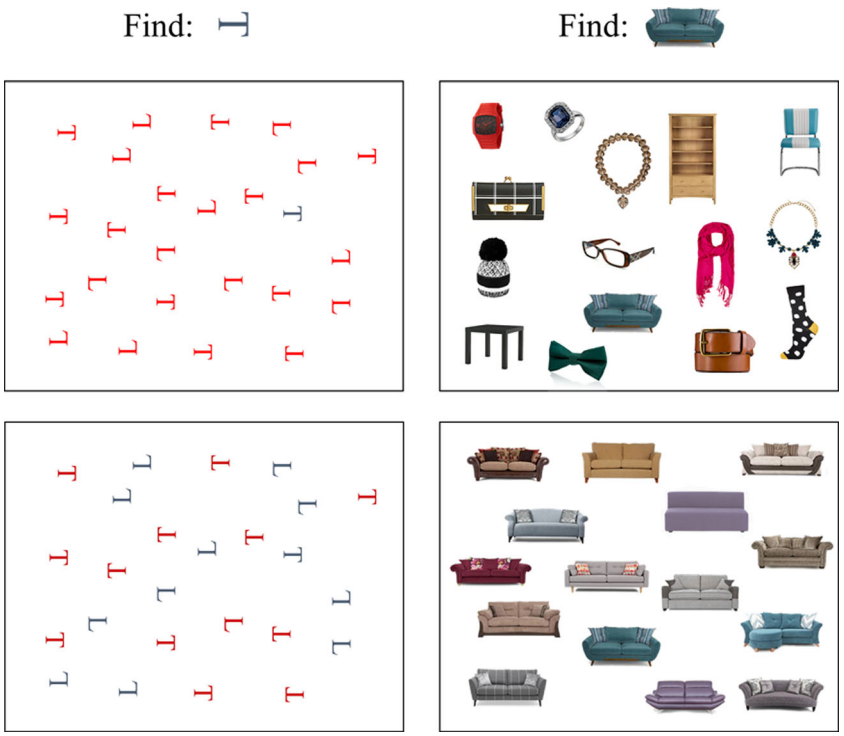


Fig. 1 Example visual search tasks. Top panels display easy searches for unique targets defined by color (left) or identity (right). Bottom panels show harder searches for targets that are more similar to the distractors

displays that camouflaged the targets. Specifically, backgrounds were created by taking a square region from the center of the target image and tiling the display with that visual information. Neider and Zelinsky (2006) tracked participants' eye-movements, and found that increasing the target-background similarity slowed search and increased error rates, while eye movement analyses showed that fixations tended to more often fall on discrete distractor items than on target-similar regions of the background. These data suggest that guidance processes are hindered by the similarity of the target to the background, but that pattern segmentation abilities leave saccadic targeting relatively unaffected. By contrast, Becker (2011) investigated the factors that determine how long the eyes remain fixed on distractors during search, asking if dwell times reflect target-distractor similarity or perceptual discriminability of the target-defining feature within the target. Her participants looked for Landolt Cs of varying line widths while having their eye movements recorded. If perceptual discriminability of the gap itself was the primary determinant of fixation duration, then the line-width of the distractors should have affected how long they were looked at (i.e., thin Landolt Cs should have elicited longer dwell times because their gap was harder to perceive), irrespective of whether or not they matched the line-width of the target. However, Becker (2011) found that distractors matching the line-width of the target were fixated longer, across all conditions of target-line width. This suggested that perceptual discriminability of the gap was not the predominant factor in determining fixation durations, but rather that target-distractor similarity reflected the time needed to reject an item as a non-target.

Moving beyond data from human observers, converging evidence for the importance of similarity in visual search comes both from simpler and from more complex systems; namely, from animal behavior and from computational approaches (e.g., priority maps derived via computational modeling). For example, Blough (1988) trained pigeons to peck at a unique target presented among identical distractors; the similarity between the target and distractor forms was quantified using multidimensional scaling (we elaborate on this technique below). The data took the form of an exponential relationship between the pigeon's ability to detect a target, and that target's similarity to the distractors that surrounded it, such that increased target-distractor similarity resulted in slower target pecking behavior. In addition, Avraham, Yeshurun, and Lindenbaum (2008) predicted human search performance from distractor homogeneity and target-distractor similarity using computational modeling.

Turning to computational and quantitative approaches, *priority maps*—mechanisms by which attentional allocation is prioritized over space (Zelinsky & Bisley, 2015)—have been constructed for comparison to (and the prediction of) human search behavior. Broadly, these take two forms: Bottom-up approaches that determine the perceptually salient feature contrasts that are present in a scene (so-called *saliency maps*; Borji, Sihite, & Itti, 2013; Koch & Ullman, 1985), and top-down, goal-driven approaches that quantify the extent to which areas of a scene match the observer's to-be-located target (so-called *target maps*; Rutishauser & Koch, 2007; Zelinsky, 2012). The output of both saliency and target maps

are conceptually alike, in that they quantify the similarity relationships in the scene, and use that information to predict behavior. In the case of saliency maps, this output often estimates the similarity of one location in a scene, relative to the local scene context (where "local" is a spatially defined parameter). In the case of target maps, the output often quantifies the similarity between one location in a scene and the searcher's information regarding the target. For instance, Zelinsky's *Target Acquisition Model* (TAM; Zelinsky, 2008) relies on visual similarity to drive its behavior. Specifically, TAM computes the similarity between the search target and visual information in the search display (using image processing techniques that represent scenes in a biologically plausible way) to determine where the model "looks." Similarity can be computed with respect to a particular target exemplar (Zelinsky, 2008) or in reference to a target category (Zelinsky, Adeli, Peng, & Samaras, 2013). TAM has been shown to successfully capture the eye-movement behavior of human observers across a range of manipulations (e.g., differences in target-distractor similarity) and ranges of complexity (e.g., simple alphabetic letter search, complex real-world scenes).

Recently, the concept of similarity has also guided theorizing about representations in visual working memory during search. Often, we must find several items (e.g., collecting one's keys, wallet, and bag before departing from home), and we tend to do so by looking for all the items at once, rather than by performing a series of consecutive, single-target searches (Hout & Goldinger, 2010, 2012; Menneer et al., 2012). When people look for multiple targets, the similarity among those items can affect search guidance, for example, by restricting attention to features that the targets have in common. Stroud, Menneer, Cave, and Donnelly (2012) had people look for two rotated Ts among rotated Ls, and they systematically varied the similarity in color of the two potential targets, by manipulating how far apart they were located in color space (Luria & Strauss, 1975). They calculated the probability with which participants fixated each distractor color, and used it as a measure of search guidance and color selectivity. They found a greater cost in guidance as targets became less similar to each other (Menneer et al., 2009, 2010; Stroud et al., 2011). Furthermore, it has been shown that the similarity of a searcher's mental representation to the exact appearance of the to-be-located target affects both the time taken to direct attention to the location of the target, as well as the time necessary to recognize it (Hout & Goldinger, 2015; Schmidt & Zelinsky, 2009).

Quantifying similarity using MDS

Despite the widespread acknowledgment in the literature that visual similarity affects search behavior, it has been surprisingly difficult to rigorously define how similar different stimuli are to one another. In experimental psychology, it has been noted that manipulating or measuring the similarity of stimulus items can be a challenging task (Hout et al., 2013; Hout, Papesh, & Goldinger, 2012, for reviews). One approach is to employ simplistic stimuli and vary a single feature dimension of each item, such as the color or orientation of a rectangular bar (Treisman, 1991). This approach provides rigorous control and can be useful for basic theoretical research, but often researchers wish to examine, control, or manipulate similarity with more complex stimuli, for increased ecological validity and generalization to real-world applications. Real-world objects are usually complex and comprise many features in multiple feature dimensions, which makes similarity among objects difficult to define, measure, and manipulate. A unique approach was taken by Alexander and Zelinsky (2011); they collected visual similarity rankings for two target categories. Participants were shown pictures of five objects at a time (e.g., lamppost, table, medicine tablet) and were asked to rank order the objects according to their similarity to the category "teddy bears" or the category "butterflies" (comparisons were made to one category at a time, and pictures of bears and butterflies were not shown). These rankings were later used to create visual search displays that contained distractors with "low," "medium," or "high" similarity to the selected target category.

Alexander and Zelinsky (2011) made an important step forward by attempting to quantify the similarity relationships among their stimuli. However, the rating system only provided the ordinal ranked similarity of a selected image relative to four others and likely would be less useful when employed with a large number of stimuli or when more precise, graded measurements of similarity are required (although it should be noted that they were able to predict these behavioral similarity ratings using computer vision techniques; we elaborate more on this study below). With this in mind, we propose a sophisticated approach to quantifying the similarity of both artificial and real-world objects by applying a multidimensional scaling analysis to experimental stimuli. MDS is clearly not a new technique (Shepard, 1980; Torgerson, 1958), but we suggest it could be used more broadly in vision science, adding greater precision to inferences drawn from experimental data. Its utility, at present, seems to have been largely overlooked by vision scientists.

MDS is a statistical tool that can be used to measure empirically *psychological* similarity (Hout, Papesh, & Goldinger, 2012; Shepard 1962a, 1962b). That is, MDS directly quantifies the extent to which someone (or a group of people) perceive items to be like (or unlike) one another. This stands in contrast to a computational approach that might, for instance, attempt to quantify the similarity between the physical (e.g., pixel) characteristics among a set of images. As input, MDS procedures take raw similarity estimates that are

provided for a selected set of items. These similarity estimates can be obtained in direct fashion (e.g., by asking people to indicate the level of similarity for each pair of items using a Likert scale) or in an indirect manner (e.g., by examining the time taken to determine that two items are not identical; Jaworska & Chupetlovska-Anastasova, 2009). These data are then subjected to data reduction procedures that are conceptually similar to *principal components analysis* and *exploratory factor analysis* (Shiffman, Reynolds, & Takane, 1981) to minimize the complexity of the ratings matrix.

The output from the MDS process is a similarity "map" that quantifies the perceived relationships among the stimuli. During the analysis, algorithms (of which there are a variety) move each item in the space (in an iterative fashion) to a location relative to the other items that, as best as possible, respects the raw similarity ratings provided by the participants. MDS is inherently spatial; thus, items that were rated as being highly similar to one another should be close to one another in the final output. To the degree that any two items were rated as dissimilar, the distance between them should grow. These "distances" are measured in *k*-dimensional Euclidean space; hereafter, for simplicity, we simply refer to them as "distances" in MDS space. Conceptually, this is much like Newton's "color wheel," which located like colors (e.g., orange and yellow) close to one another on the wheel, and unlike colors relatively more distant from one another (e.g., orange and purple; Newton, 1704). Model fit, or the degree to which the output space conflicts with the raw ratings, is reflected in a measure known as *stress*. When the algorithms have completed their iterations (optimization criteria vary, and usually rely on changes in stress), dimensional coordinates are provided for each item that can be used to create a visual representation of the outcome, as well as to quantify similarity as the distances between each pair of items in the space.

The dimensionality of the analysis is under control of the researcher, and there are various techniques for determining the appropriate complexity of the space (Kruskal & Wish, 1978; Lee, 2001; Oh, 2011). Often, the output is sufficiently simple to permit a visual representation of the underlying relational structures and factors that impacted the initial similarity ratings. When analysis is purely exploratory, simpler, low-dimensional solutions may in fact be preferred, because they permit visual inspection, but when visualization is unimportant, higher dimensionality spaces can be of use. Interpretation of the solution is subjective: Researchers examine the organization of the space and attempt to make inferences regarding the factors that influenced the similarity ratings, or attempt to verify *a priori* hypotheses regarding the underlying structure. To provide an example, Godwin, Menneer, & Hout (2014) conducted an MDS analysis on the numbers 0-9. The MDS analysis revealed a simple, two-dimensional space, the organization of which suggested that people appreciated the roundness and straightness of the lines (e.g., 0 was more similar to 6 than it was to 1), and the extent to which the numbers created open vs. closed spaces (e.g., 0 was more similar to 9 than it was to 3). This resulting MDS space was highly similar to one derived by Shepard, Kilpatric, and Cunningham (1975), demonstrating that, despite variations in fonts, stimulus size, and presentation methods, the underlying similarity space generated was consistent over several decades. This suggests that MDS taps into an underlying perceptual appreciation of similarity. It is important to emphasize, however, that the stimuli being rated need not be so basic, nor does the outcome of the analysis need to be composed of so few featural dimensions in order for a space to be of use to vision researchers.

For example, consider again the top-right panel of Fig. 1. An attention researcher may wish to know the extent to which a person searching for a blue couch is distracted by the presence of things that share the target's color (e.g., the blue bowtie), relative to those that share categorical similarity (e.g., other furniture, such as the chair). Here, the similarity of the stimulus set in question is likely to be governed by more complex relationships, such as color, shape, texture, and so on. One of the most appealing aspects of using MDS is that it is agnostic with respect to the underlying psychological structure that influenced participants' ratings. *A priori* hypotheses regarding how similarity estimates were given are not necessary. By examining the spatial output, the analyst can intuit and quantify the dimensions underlying people's ratings, but again, this process is not necessary for quantifying similarity between items. Even if an MDS solution is founded on six or seven underlying dimensions, which would be far too complex to visualize or experimentally control, the computational algorithm will provide psychological distance estimates that should (theoretically) predict search behavior.

It should also be noted that there are other theoretical accounts of similarity that are non-spatial in nature. For instance, feature-based accounts of similarity (Tversky, 1977) assume that the basic representational units of similarity are not continuous (as MDS techniques assume), but rather, are binary (i.e., they represent the presence or absence of a feature). Although feature-based statistical techniques like *additive clustering* (Shepard & Arabie, 1979; Shepard, 1980) have some mathematical advantages over spatial models such as MDS (e.g., by accounting for potential violations of the "triangle inequality assumption"; Tversky & Gati, 1982), they are disadvantageous in some circumstances because the analyst must be able to make predictions regarding the features of which their stimuli are comprised. This is particularly problematic when the stimuli are real-world objects (like couches and chairs), which are likely to be comprised of many simultaneous and non-equivalent features (e.g., color is likely a more salient feature than texture). Unlike featural approaches, however, spatial techniques (like MDS) do not require the analyst to know ahead of time what features will be

appreciated by their participants, which features are more or less salient, and so on.

To be clear, MDS is not a panacea, and is not necessarily preferable in every circumstance in which one may wish to quantify similarity. Although the subjective nature of MDS may be advantageous when exploring or establishing featural dimensions that might exist for given stimuli, this subjective judgment can become problematic in the context of extremely complex stimuli, wherein the distinctions between feature dimensions may be harder to distinguish and therefore outputs are less reliable. While non-spatial models often require firm predictions about relevant dimensions, MDS can allow a more exploratory approach. However, such exploration does not alleviate the responsibility of firm predictions when MDS data are used to underpin hypotheses for future testing. MDS is a particularly useful tool for quantifying similarity, but researchers may require other approaches for their specific materials or procedures.

Returning to Fig. 1, by looking at the computed distances from an MDS analysis (i.e., the Euclidean distance in *k*-dimensional space between each pair of points), an analyst may conclude that the blue couch is more similar to the blue bowtie than it is to the chair, or that the table and bookshelf are perceived to be approximately as similar to each other as are the sofa and the chair. The units of distance provided by an MDS output are arbitrary (and vary across scaling algorithms), but what is universally important are the *relative* distances between pairs of objects (e.g., the couch-to-bowtie distance relative to the couch-to-chair distance). These simple computations of Euclidean distances thus allow the researcher to empirically quantify the similarity of items in a stimulus set.

Outline of the approach

Broadly, our suggested approach can be accomplished in three simple steps: 1) Collect similarity ratings for all the stimuli to be used in the visual search experiment; 2) Apply MDS analysis to the similarity ratings; 3) Use the output to quantify the relationships among items for stimulus selection and/or control, and then examine search data as a function of the MDS-derived similarity. To provide a concrete demonstration, we later briefly review the applications of this approach in Godwin, Hout, and Menneer (2014), and Hout and Goldinger (2015).

The first step is to collect all the stimuli that are to be used in the visual search experiment, and have participants provide similarity ratings for the items. Ordinarily, this step is completed using some form of direct ratings procedure, whereby the participants knowingly rate or classify the items. In the simplest case, participants may be shown two items at a time and indicate for each pair how similar they are to one another, using a Likert scale or a slide bar (Faye et al., 2004; Rosenberg, Nelson, & Vivekananthan, 1968). Alternatively, participants may categorize the items according to some criteria, or sort the items into piles based on their relatedness (Borg & Groenen, 1997).

There are many data collection techniques that can be used, the suitability of which may vary according to the characteristics of the stimuli (Jaworska & Chupetlovska-Anastasova, 2009, for discussion). Recently, spatial techniques have been advanced (Goldstone, 1994b; Hout, Goldinger, & Ferguson, 2013; Kriegeskorte & Marieke, 2012) that offer an intuitive interface for collecting ratings, and simultaneously cut down on the time needed to rate all the items. By contrast, indirect methods may be used that capture a measure of similarity using secondary empirical measurements (i.e., stimulus confusability; Shepard, 1963). Generally, this process involves the use of a perceptual discrimination task, whereby participants briefly see pairs of stimuli and indicate whether they are the same or different. The rating data that are collected from these tasks are either the speed of accurate responses (Papesh & Goldinger, 2010), or the likelihoods that items within pairs will be mistaken for one another (Gilmore, Hersh, Caramazza, & Griffin, 1979). The logic is that if two items are very similar (e.g., pictures of a Basset Hound and a Blood Hound), they will be hard to differentiate, and therefore discrimination reaction times (RTs) will be long, and mistakes will be likely. When two items are dissimilar (e.g., pictures of a Basset Hound and a Chihuahua), discriminating between them will be easier, and RTs will be shorter and mistakes less likely.

In conducting this step of our approach, it is important that all of the stimuli to be used in the search task are included. Similarity is dynamic, and subject to variation when the backdrop for comparison is changed (Goldstone, Medin, & Gentner, 1991; Goldstone, Medin, & Halberstadt, 1997; Spencer-Smith & Goldstone, 1997). Therefore, adding stimuli to the search experiment after the similarity ratings and analysis have been completed can potentially alter the conclusions made regarding the relationships among the items, and subsequently their effect on search behavior.

For instance, imagine similarity ratings are collected for an assortment of dog breeds. For simplicity, imagine there are only four dogs in which you are interested: A Basset Hound, a Beagle, a Chihuahua, and a Great Dane (Fig. 2). The Basset and the Beagle are likely to be close to one another in the MDS output, considering their similarity in appearance, and the pair of them are likely to fall somewhere in between the Chihuahua and the Great Dane, considering their medium-size stature. Now imagine that you are to consider a Blood Hound as well. The similarity in appearance between Basset and Blood Hounds is striking, but the Blood Hound approaches the Dane in size. Thus, both the Basset and the Great Dane are likely to be drawn towards the Blood Hound, warping the similarity space by making, for instance, the Basset now seem

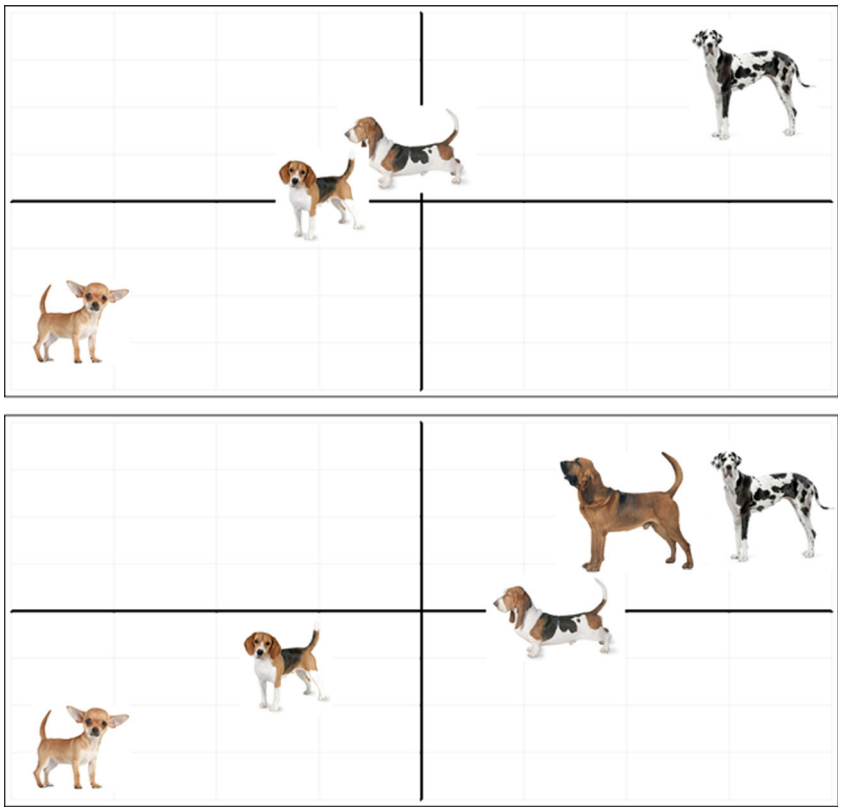


Fig. 2 Schematic MDS outputs for dog stimuli. In the top panel, the Beagle and Basset are indicated to be highly similar, relative to the other pairs of dogs. In the bottom panel, the change of context brought about by the addition of the Blood Hound draws the Basset away from the Beagle, due to its similarity in appearance to the new breed

slightly less like the Beagle than it was formerly. This cautionary note is not meant to suggest that similarity is too variable to be of utility. Quite the contrary is true, as perceptions of similarity have been shown to be very consistent over time when overall context remains constant (Godwin, Hout, & Menneer, 2014; Shepard, Kilpatric, & Cunningham, 1975). What we intend is simply to make clear the importance of, during the similarity ratings phase, considering all possible stimuli that are to be used in the later tasks.

In the second step of this approach, the similarity ratings are analyzed using MDS. A full and step-by-step guide is beyond the scope of the current article, but many resources are available for learning this statistical technique, including a tutorial by Hout, Papesh and Goldinger (2012; Kruskal & Wish, 1978; Giguere, 2006). To conduct the analysis, the similarity estimates must be gathered into a similarity matrix. This is a collection of similarity estimates between each pair of items in the set, whereby the rating for each pair is placed at the intersection of the appropriate row and column of the matrix. The data can then be analyzed in statistical software packages (e.g., SPSS, Matlab, or the open-source software, R), using one of several instantiations of scaling algorithms, such as *PROXSCAL* (Busing, et al., 1997) or *ALSCAL* (Young, Takane, & Lewyckyj, 1978). Choosing the right analytic model to analyze the data is an important step, because one must take into account both the manner in which the similarity estimates were obtained, as well as the analytic goals of the research team. The different models vary in the geometry they use to map the data, the algorithms used to maximize model fit, the ability to handle single vs. multiple similarity matrices, and so on. One must also take care to use metric scaling algorithms when similarity ratings are quantitative (i.e., interval or ratio level, such as perceptual discrimination RTs) and non-metric algorithms when similarity ratings are qualitative (i.e., ordinal level, such as Likert scale ratings).

Most germane to this approach is selection of an aggregate or individual differences approach to scaling. Typically, data are collected from multiple participants, and an aggregate similarity matrix is created, from which a single similarity map is produced. This method is useful when a researcher wants to quantify similarity according to average perceptions, much the same as when multiple participants contribute RT data to an experiment from which a single mean is acquired. However, it should be noted that our suggested approach is not limited to collective perceptions of similarity. There are options available, should the researcher wish to take into account between-participant differences. One scaling algorithm, called *INDSCAL* (Carroll & Chang, 1970), exists specifically for this purpose. INDSCAL will accommodate matrices from multiple participants, and in addition to the standard aggregate group plot, it will provide a secondary output that indexes the degree to which each participant agrees with the collective. An alternative to using INDSCAL would be simply to

conduct MDS analyses for the data from each individual participant, and use those data to examine participant-level performance during search as a function of personalized similarity ratings.

The third and final step of our approach is to examine the MDS output and prepare it for use in conjunction with behavioral and oculomotor observations. A key fact to remember is that MDS spaces portray information only about the *relationships* among the items, rather than about specific properties or absolute values. The direction of the dimensions, for instance, is unimportant. For example, in Fig. 2, one might infer that a dimension of "size" is present in the output, as small dogs are located on the left, medium-sized dogs in the middle, and large dogs on the right. If the orientation of the dogs were to be flipped, landing small on the right and large on the left, interpretation of the space would be unchanged, as the locations of the dogs would still convey information about their relative size, and the Euclidean distances would be unaffected. More importantly, it is crucial to keep in mind that the units provided by an MDS analysis are arbitrary. The Basset and the Beagle may be located 3 units away from one another, for instance, but those units are uninterpretable until the other distances between items are taken into consideration. By comparing the 3-unit Basset-Beagle distance to the 20-unit distance between the Basset and the Chihuahua, however, we are able to conclude that the first pair are more similar to one another by a specific order of magnitude.

How the researcher chooses to use these similarity values will be determined by the goals of the research question; that is, by the manner in which similarity is being explored or manipulated during search. There are various ways these similarity values can be used, depending on whether the researcher wishes to use them in an active (manipulative) or passive (exploratory) manner. By *active*, we mean controlling the makeup of the search task through the selection of images with pre-specified similarity relationships (e.g., the degree of similarity between the target and the distractors, or among the distractors themselves). By contrast, *passive* use of the similarity data would involve simply observing how the similarity relationships among items affect behavior. These methods are not mutually exclusive, of course, and could easily be employed simultaneously.

For example, imagine conducting a simple visual search experiment using letters of the alphabet as stimuli. One could easily collect similarity ratings for the letters A-Z, and analyze them using MDS. An active approach might involve creating stimulus displays with distractors that are highly visually similar to the target (e.g., search for the letter *M* among *N*s and *W*s) or displays that are highly visually dissimilar to it (e.g., search for the letter *M* among *D*s and *U*s), and then examining RT as a function of the similarity characteristics of the distractors. The makeup of these displays would therefore be empirically derived – with substantially more precision than an ordinal ranking system (e.g., as employed in Alexander and Zelinsky, 2011)—through the selection of letters with predetermined likeness to the target (e.g., by selecting distractor letters that have a minimum or maximum level of similarity to the target, or by choosing only the two most similar letters). In a passive approach, by contrast, one might display all the letters of the alphabet as distractors, and track the eye movements of participants. Empirical measurements here might include the probability that a given letter was fixated (or for how long) as a function of its similarity to the target. In the next section, we provide more tangible demonstrations of these tactics, as documented in two recent articles.

Documented examples of usage

Recently, we have used this MDS approach to further understand the role that semantic information plays in attentional guidance during search. When searching for things in the real-world, our attention is drawn to locations in which we are likely to spot our target (e.g., to countertops and tables if we are looking for a beer mug in a tavern), and to objects that are conceptually related to it (e.g., to a martini glass; Oliva & Torralba, 2007), suggesting that high-level knowledge is incorporated into the guidance of attention in a scene. There has been growing interest in understanding the role of semantics of late (Henderson, Malcolm, & Schandl 2009; Wolfe, Võ, Evans, & Greene, 2011), partly because until recently, it was believed that there was either zero or a very limited role for semantics in guiding search (Wolfe & Horowitz, 2004, review).

In Godwin, Hout, and Menneer (2014), we used MDS in combination with eye-tracking to tease apart the effects of visual and semantic similarity during search for numbers. Participants searched displays for a single-digit number, indicating the presence or absence of the target in each trial. We chose to use numbers as a starting point for investigating semantic influences on guidance because numbers represent a highly controllable stimulus space in which semantic similarity is inherent. Numerical meaning is extracted quickly from visual displays (Corbett, Oriet, & Rensink, 2006). Indeed, Schwarz and Eiselt (2012) found that the presence of numerically similar distractors (e.g., 4, 6) decreased speed and accuracy in search for the number *5* compared with dissimilar distractors (e.g., 1, 9), suggesting that attention was drawn to digits that were numerically close to the target value. They did not find this pattern of results when participants searched for the letter *S*, even though it is highly visually similar to the digit *5*, suggesting that the visual characteristics of the numbers were unlikely to be driving their semantic effects.

Godwin, Hout, and Menneer (2014) built upon the study by Schwarz and Eiselt (2012) by quantifying visual similarity using MDS ratings of the digits, and by tracking the

oculomotor behavior of our participants. One group of participants provided visual similarity ratings for the numbers 0-9; we then analyzed the data using MDS, and indexed visual similarity via the distance between each pair of numbers in MDS space. Semantic similarity (as in the Schwarz and Eiselt, 2012, study) was indexed by the numerical distance between the numbers. One potential concern with indexing visual similarity using MDS measures is that – despite our instructions to rate the stimuli purely based on their visual characteristics – it is possible that the semantic (numerical) similarity of the numbers also affected the ratings provided by our participants. To address this concern, we correlated the visual similarities (i.e., the MDS distances between each pair of numbers) with their corresponding numerical similarities, and found no hint of a correlation ($r = -0.06$). This suggests that our participants provided ratings that reflected solely the visual similarity among the numbers (or at least that they sufficiently minimized the influence of numerical similarity when providing their ratings).

Our next step was to examine participants' visual search performance as a function of both visual and numerical similarity (these participants were different from those that had provided the similarity ratings). We reasoned that if semantic information influences guidance during number search, then (when statistically controlling for visual similarity) we should see an elevated probability that participants will fixate numerically-similar distractors. However, we expected that the effects of semantic relatedness would be dwarfed by those of visual similarity. We used a linear mixed-effects model to analyze fixation probability, so that we could examine the effects of semantic relatedness while also accounting for the visual similarity between target-distractor pairs. Our results were in keeping with our hypotheses: Both factors guided attention, but visual similarity had an effect that was nine times greater in magnitude than that of semantic relatedness. By applying the approach outlined in this article, we were therefore able to directly quantify the role that both visual and semantic similarity play in visual search for numbers.

Subsequently, we have applied this approach to the study of representations in visual working memory (VWM) during search, using it to predict eye movement behavior as a function of the accuracy of target templates (Hout & Goldinger, 2015). A target template is a mental representation of the thing (or things) a person is trying to locate. They are used to both guide attention, and to verify (or reject) incoming visual information as matching the target (Rao et al., 2002; Wolfe et al., 2004; Zelinsky, 2008). In laboratory search experiments, participants can often form a near-perfect target template, because they are typically shown, prior to search, a veridical representation of the target as it will appear in the search display. However, in the real world, targets are less well defined because specific details often are hard to predict (e.g., you are looking for a pepper, but do not know what kind), because things often change appearance relative to the last time they were encountered (e.g., when picking your friend up from the airport, you may be surprised to find he shaved his beard), and so on (Zelinsky et al., 2013a).

Using the approach outlined in this article, we examined the extent to which template accuracy impacts attentional guidance and decision-making during search. Inferring the accuracy in a mental representation is problematic for obvious reasons, so in Hout and Goldinger (2015), we chose instead to manipulate template accuracy in two ways: 1) By introducing inaccurate features aimed at contaminating the searcher's template, and 2) by adding extraneous features to the template that were unhelpful during search. We began by collecting similarity data for 240 different real-world object categories, and thereafter conducting an MDS analysis on each category (the pictures are available for free download from the *Massive Memory Database*, cvcl.mit.edu/MM/stimuli.html, and the MDS data are available as documented in Hout, Goldinger, and Brady, 2014). Our first group of experiments involved a search task wherein, in most trials, the template cue displayed the target exactly as it would appear in the search display (instructions told participants to "*Please search for this item or something very much like it*"). In a minority of trials, the eventual target deviated from the appearance of the cue in some way (but was always from the same unambiguous object category). To promote varying degrees of template accuracy, we sampled systematically from our MDS spaces. More specifically, we used the distance in MDS space, between the target cue and the actual target (that appeared in the display), as a proxy for template accuracy.

In the second group of experiments, we manipulated the "width" of the template feature space by having people look for two targets simultaneously, only one of which could appear in the display (unaltered, relative to its cue). Either object in the search template could be the target, so participants had to prepare to find either in the search display. We varied the distance in MDS space between these potential targets, predicting that, to the extent that these images were dissimilar, template accuracy would be reduced (Stroud et al., 2012). In both experiments, we found converging evidence for a dual-function theory of target templates. Specifically, we found that degraded template accuracy slowed the speed of accurate search (indexed by RT data), attributable to hindered attentional guidance and decision-making processes (indicated by scan-path ratios and decision-times, respectively).

Contrasting MDS with computational approaches for quantifying similarity

One possible reason that our proposed technique has not already been more widely adopted may have to do with the appeal of computer vision approaches that quantify similarity

on a physical basis (e.g., pixel comparisons), and thereby sidestep the need to collect similarity ratings from human participants. If automated (computationally derived) similarity estimates can predict human behavior on par with similarity derived from human ratings (e.g., MDS), then there is an incentive for researchers to adopt this powerful, and potentially less laborious approach. This is a nontrivial point; as will be discussed in the "General Discussion", one potential drawback to using human similarity ratings is that they are time-consuming to acquire. Recently, some work has been dedicated to examining new techniques that speed data collection (Hout, Goldinger, & Ferguson, 2013; Kriegeskorte & Marieke, 2012), but even so, collecting human similarity ratings on as many as a few hundred stimuli may be prohibitively time-consuming. By contrast, automated computational methods could theoretically produce estimates for millions of stimuli, or more. That being said, computer models do not typically agree universally with human raters, so it also can be informative to examine the situations in which ratings are out of alignment, and use that to inform future theorizing and modelling. Indeed, some work has already begun to combine and contrast behavior as predicted by human ratings and computer vision approaches.

A particularly good example of this is the Alexander and Zelinsky (2011) study discussed (briefly) earlier. This study exemplifies the idea that human ratings and computational methods can complement each other, and can be used to inform theories of search. In their first experiment, Alexander and Zelinsky (2011) had participants provide similarity ratings for 500 objects, indicating how similar each picture was to a cued target category (teddy bears or butterflies). These ratings were then used to classify the objects: Images that consistently received the lowest or highest similarity ratings were designated to the "low" and "high" similarity categories, respectively, and the images in between were labelled "medium." The authors found that these ratings later successfully predicted search behavior.

Search participants (Experiment 2) were asked to look for categorically defined targets (e.g., "look for a teddy bear"). There were three different types of trials that varied in how similar the distractors were to the target category (as classified via the ratings in Experiment 1): Low-similarity and high-similarity trials, wherein all the distractors had roughly equivalent similarity to the target, and mixed-trials, wherein two distractors were selected to have low, medium, and high similarity to the target, each. Across experiments, analyses were focused on target-absent trials, in order to focus exclusively on how the similarity of the distractors affected performance. Participants were faster to respond in the low-similarity trials, followed by mixed trials and high-similarity trials, and the eye-movement analysis revealed that initial fixation probabilities reflected target-distractor similarity (i.e., in the mixed condition, the first fixation in a trial was most likely to be directed at a high-similarity distractor, followed by medium- and low-similarity distractors).

In subsequent experiments, Alexander and Zelinsky (2011) derived similarity ratings from computer vision techniques. When human raters appreciate and rate stimuli, it is difficult to know whether or not semantic associations have contaminated the ratings, even if the instructions asked people to consider solely the visual characteristics of the pictures (Medin, Goldstone, & Gentner, 1993). By using computer vision to derive the ratings, however, Alexander and Zelinsky (2011) removed the potentially confounding influence of semantics, ensuring that the ratings reflected purely the visual features of the stimuli. The method they employed (described in more detail in Zhang, Samaras, & Zelinsky, 2008; and Zhang, Yu, Zelinsky, and Samaras, 2005) works by allowing multiple visual features (e.g., color, texture, global shape) to contribute independently to the classification of an image. Two image classifiers were trained – one that discriminated bear images from non-bears, and one that discriminated butterflies from non-butterflies – and were shown to successfully differentiate the target categories from other, random object categories. The distractor pictures were then rank-ordered, with respect to how well they fit the target classifiers; the top- and bottom-third of the pictures were then identified as high- and low-similarity distractors, and used in a subsequent search experiment (Experiment 3) analogous to the one that relied upon the human similarity ratings (Experiment 2). The critical difference was that in Experiment 3, the similarity ratings were derived from computational methods, rather than from people. Again, search performance reflected the similarity makeup of the distractors, with faster RTs among items that were less similar to the target category, and more first fixations falling upon items with higher similarity to the target category.

Although the two search experiments produced qualitatively identical results, there were pictures for which the human raters and computer model did not agree, leading the researchers to ask whether the human and computer ratings were based upon different determinants of similarity. For instance, differences might arise via the possibility of semantic influences on ratings given by people, differential weighting of features between methods, or the possibility that some features were appreciated by humans but were not included in the computational model. It should be noted that the model typically agreed with human raters: Of the low- and high-similarity classifications, there was roughly 38 % agreement, and only rarely did the human and computer ratings strongly conflict (i.e., less than 2 % of the pictures were rated as "low-similarity" by one method and "high-similarity" by the other, or vice versa). In their fourth (and final) experiment, Alexander and Zelinsky (2011) constructed distractor arrays with four objects. In the high-similarity target-absent trials, one object was chosen that was rated as highly similar by the human raters but not the model (human-only distractors),

one rated as highly similar by the model but not the human raters (model-only), one on which the humans and model agreed on a high similarity rating (human + model), and one on which both agreed on a medium-similarity rating (medium); the converse was true for low-similarity trials.

To examine which similarity measure best predicted human behavior, the authors examined which items were more (or less) likely to be fixated first; the logic being that the best predictor of high similarity should acquire the largest number of first fixations, and the best predictor of low similarity should acquire the fewest. In high-similarity trials, when people searched for butterflies, the human + model distractors were first fixated most frequently, followed by the human-only, and then the model-only and medium distractors (which were not statistically different from one another). When searching for teddy bears, all three high-similarity classes were fixated first more frequently than the medium distractors, but were not statistically different from one another. Turning to the low-similarity trials, when searching for butterflies, human + model distractors were fixated first the least frequently, followed by the human-only distractors, and then the model-only and medium distractors (which again were not statistically different from one another). Nearly identical results were obtained for the teddy bear search group, but there, the human + model distractors did not outperform the human-only pictures. Taken together, the results suggest that the best predictor of human behavior came from using distractors whose ratings were agreed upon by human raters and a computational model, and when the human and computational ratings were misaligned, the human ratings always performed as well as or better than the model.

The Alexander and Zelinsky (2011) study illustrates that similarity as rated by people may at times be a better predictor of human search behavior than a purely computational approach, but also, crucially, that the best predictions sometimes come from situations in which human and computational ratings agree. This is a strong argument in favor of combining computer vision techniques with human ratings (e.g., MDS). Although in the Alexander and Zelinsky (2011) study, a purely computational approach was unable to outperform human ratings, there is no reason to suspect that computational methods are universally lesser than human ratings, nor that human ratings will consistently outperform computer vision in the years to come. A full review of the computer vision literature is certainly beyond the scope of the current article, but it should be noted that considerable strides have been made in recent years to describe, classify, and categorize images, using computational techniques (de Campos, Csurka, & Perronnin, 2012; Krizhevsky, Sutskever, & Hinton, 2012; Yun, et al., 2013; Ristin, Gall, Guillaumin, & Van Gool, 2015; Zhou et al., 2015), and that these methods have proven useful in predicting human behavior.

For instance, Zelinsky, Peng, & Samaras (2013) showed that human participants could identify the target category that another searcher was looking for by examining what distractors that person examined, and that a machine vision decoder – specifically, a Support Vector Machine (SVM) classifier—performed on par with the human decoders. A recent study by Maxfield, Stalder, and Zelinsky (2014) found that the typicality of an image (i.e., how representative of a category an item is, as indicated by human participants) predicted search guidance and target verification; relative to images rated as being less typical, higher typicality items were fixated more quickly and were responded to faster once gaze fell upon them. Importantly, Maxfield and colleagues (2014) also trained an SVM classifier on the target categories and found that the computational confidence ratings—indexed via distance from the object classification boundary—mirrored the behaviorally obtained typicality ratings. Clearly, computational methods are already capable of predicting behaviorally meaningful results, and it seems that their capacity to do so will only grow over time (Zelinsky, Peng, Berg, & Samaras, 2013).

Although human and computational methods are not mutually exclusive, there are several reasons that researchers may choose to adopt a solely human-based approach to obtaining similarity ratings. First, simplicity without the loss of fidelity: Asking human raters to provide similarity estimates is a comparatively straightforward task when one considers the vast range of machine vision techniques that one may choose to employ. Computer vision methodology is evolving and becoming more sophisticated, which is exciting, but that makes choosing the right technique (and implementing it) a potentially daunting task. By comparison, researchers have been asking human raters to provide simplistic, overt similarity ratings (or indirect ones, via tasks like perceptual discrimination) for decades, and literature abounds showing that people are generally able to accurately convey their perceptions of similarity. Indeed, spatial models of similarity (drawing upon human rating data) have been so useful that they have even given rise to one of the few "laws" of psychology, Shepard's (1987; 2004) "universal law of generalization," and have been incorporated into sophisticated, highly successful mathematical models of cognition, such as Nosofsky's *Generalized Context Model* (Nosofsky, 1986). So the simplicity of obtaining estimates from people, it can be argued, does not necessarily come at the cost of imprecision.

A second reason one may choose to use human ratings is that in some situations, it might be advantageous to adopt a metric of similarity that is not purely based on physical characteristics. In Godwin, Hout, and Menneer (2014), we aimed to isolate visual from semantic similarity, and in Alexander and Zelinsky (2011), a computer vision model was implemented specifically to eliminate the influence of semantics. However, return to Fig. 1 and consider the similarity of the

couch, relative to the bowtie and the chair. A purely visual approach to similarity may suggest that, due to the global shape, color, and texture of the couch, it is more similar to the bowtie than the chair. But if someone were asked to search for "a couch," that person may well seek out features that are related semantically (but not necessarily visually) to couches. Attention may therefore be directed more quickly to things that enable sitting; that is, objects that share categorical features, like having four legs, and a flat base on which to rest. Zelinsky et al. (2013a) recently advanced the TAM model (Zelinsky, 2008) showing that a computational model can in fact learn to discriminate a category (rather than an instance) of targets from non-targets, which suggests that computer vision approaches may soon be able to tackle this task more broadly. But simply by allowing human raters to consider all aspects of similarity when providing their estimates (i.e., not giving instructions that ask them to focus solely on visual characteristics), we may obtain similarity ratings that tap into these category-predictive features.

Third, data from human raters allows researchers to examine feature spaces without the necessity of explicating those features a priori. Simply put, MDS can be used in a purely exploratory fashion in order to arrive at hypotheses regarding the features people used to conduct their ratings, and the relative salience of those features. By contrast, computational models by their very nature require that object features be identified and represented, and usually require explicit descriptions of the manner in which similarity is determined or defined, as well as the weighting of the features relative to one another. Therefore computational approaches are sometimes less amenable to exploration than they are to confirmation, though they are certainly not exclusively used for confirmatory purposes. Computer vision approaches are advancing rapidly, and many techniques are progressed using computational experimentation, aimed at the exploration and identification of stimulus features.

With MDS, once subjective hypotheses have been arrived at, they can then be tested for confirmatory purposes using linear regression (Kruskal & Wish, 1978; Green, Camone, & Smith, 1989). For instance, suppose the MDS space of a collection of dogs tends to place short-haired dogs on one side of the space, and long-haired dogs on the other. An analyst could test this hypothesis (i.e., "dogs are differentiated based on the length of their coat") by asking a new group of participants to view the dogs, one at a time, and indicate (e.g., using a Likert scale), "how long is this dog's coat?" The ratings from this secondary task could then be regressed on the individual dog's locations in the MDS space to determine the degree to which this hypothesized feature (coat length) maps onto the similarity ratings. High regression weights would indicate that a particular dimension reflects the hypothesized construct. Regression weight (or effect size) comparisons could then be used to examine the degree to which a particular construct is represented on one of the MDS dimensions (after all, sometimes features may co-vary, as would be the case if coat length correlated with the size of the dog or the color of its fur), or the degree to which one construct is a better predictor of similarity, relative to another (e.g., perhaps color is a better predictor than is coat length). Because the order of the dimensions in an MDS analysis reflects their relative importance (i.e., the degree to which a particular dimension explains variance in the raw similarity ratings), an added benefit of this technique is to allow you to uncover not just what features have been appreciated, but also which features are most salient to human raters (e.g., perhaps coat length and color are both important, but color explains more variance, therefore suggesting it is a more salient feature than is coat length).

It may come to pass that the quality of data from human ratings is someday surpassed by those from computational methods. If and when this occurs, it certainly would be more advantageous to adopt a computer vision approach, in order to eliminate the necessity of additional data collection that may require time, effort and money, as well as to deal with the concern that human raters are prone to individual differences, lapses of attention, and so on. Currently, however, it seems that the most advisable approach would be to adopt a synergistic approach that combines human data with computational methods.

General Discussion

Clearly, similarity plays an important role in vision, but it should be appreciated that its utility is not limited to visual cognition or, for that matter, cognition more broadly construed (Hahn, 2014; Hout et al., 2013). The concept of "sameness" is important to understanding attention and perception (Nosofsky, 1986; Solan & Ruppin, 2001), as well as predictions from memory theories (Gillund & Shiffrin, 1984; Hintzman, 1986, 1988). Without being able to appreciate resemblance, we would be unable to successfully categorize new items as belonging to learned categories (Goldstone, 1994a; Goldstone & Steyvers, 2001; Nosofsky & Palmeri, 1997); indeed, Shepard's (Shepard, 1987, 2004) pivotal "universal law of generalization" hinges on the similarity between new stimuli and what has been experienced in the past. Lexical access (and production) is helped along by being able to recognize the similarity between an utterance and previous experiences with like-sounding expressions (Goldinger, 1998; Goldinger & Azuma, 2004). Additionally, in the classic other race effect, it is thought that faces of members of an outgroup are less easily perceived and remembered because they appear more similar to one another than do members of one's own group (Goldinger, He, & Papesh, 2009; Papesh & Goldinger, 2010). Similarity judgments have been used to uncover the visual features that relate to the perception of naturalness, and

have been shown to correlate strongly with the low-level visual features of images (as determined by computer algorithms; Berman et al., 2014). The concept of similarity even makes contact with the neuroscientific literature, as overt human similarity ratings have been shown to map onto activity in the human and non-human primate inferotermporal cortex (Mur et al., 2013). Moving outside of cognition, similarity has a part to play in computational approaches to pattern recognition and machine learning (Bishop, 2008), text retrieval (Salton, 1991), and aspects of artificial intelligence (Riesbeck & Schank, 1989), and MDS techniques have even been used to examine the structure of bacterial colonies associated with living organisms (McFrederick, Wcislo, Hout, & Mueller, 2014), to name a few examples.

Given its widespread applicability, it is striking that MDS has been applied to questions in vision science so infrequently. One of the primary barriers to using MDS techniques has historically been that collecting similarity data using traditional pairwise rating methods is an extremely time-consuming task (Hout, Goldinger, & Ferguson, 2013). The number of pairwise comparisons needed for any given group of items increases rapidly as the number of members in the set grows. Specifically, for n items, $n(n\text{-}1) / 2$ comparisons are needed (e.g., for a set of 5, that is 10 pairwise comparisons, but growing the set to a mere 30 items requires 435 comparisons). Thus, to conduct the approach outlined in this article with any more than around 30 items would require lengthy experimental protocols that may be fraught with problems, such as participant fatigue or disinterest, and evolving strategies during the ratings task. Clearly, if the similarity rating phase of our approach takes many times longer than the search phase itself, then this technique could become impractical and intractable.

Resources for conducting this approach

Fortunately, there are alternatives to the standard pairwise method of similarity data collection. In particular, spatial procedures for collecting ratings, such as the *spatial arrangement method* (SpAM; Hout et al., 2013) or *inverse MDS* (Kriegeskorte & Marieke, 2012) offer intuitive interfaces by which participants can depict their perceptions of the similarity between many items at once. SpAM and inverse MDS are different in some respects—for instance, inverse MDS requires repeatedly rating some items—but both involve the presentation of all (or many) of the objects in a set simultaneously. Instructions ask participants to move the items around in space (using the computer mouse), placing them at distances from one another that respect the perceived similarity of each pair. In essence, it is as if the rater is projecting her mental representation of the set onto a two-dimensional plane.

The most germane benefit of these approaches to collecting similarity ratings is that a single trial provides data on many item pairs simultaneously; ratings are simply the Euclidean distances between each pair of items on the screen, measured in pixels. Thus, data collection is greatly speeded, without suffering in quality (Hout et al., 2013). Whereas a 30-item set may take 25-30 minutes to rate using standard pairwise methods, a single trial of SpAM is sufficient to handle the task, often being completed in as little as 3-5 minutes. When object set sizes increase beyond 30 items, conducting a single-trial version of SpAM is unmanageable, due to the fact that the objects would have to be displayed in minute scale to be able to fit on the screen all at once. However, recent multi-trial versions of SpAM have proven useful in collecting similarity ratings for large sets of items across several trials. For example, in Berman et al., (2014), we collected similarity ratings for sets of 70 real-world scenes using controlled randomization procedures across only 29 SpAM trials.

It also should be noted that many resources are available that exist to aid researchers in the collection of such similarity data. Kriegeskorte and Marieke outline in detail their adaptive algorithm for inverse MDS (Kriegeskorte & Marieke, 2012), and on the first author's website (www.michaelhout.com) there are many freely available programs (written in E-Prime) for the collection of similarity data in a variety of techniques (including pairwise methods and SpAM), as well as Excel workbooks to aid in data organization and concatenation. Moreover, large-scale databases have recently been constructed that have used MDS to provide quantified similarity for sets of real-world object pictures (Hout et al., 2014; Horst & Hout, in press; Migo, Montaldi, & Mayes, 2013). Compared with simplistic stimuli, such databases allow researchers to explore broader issues under increased ecological validity while still maintaining control over visual similarity. With respect to the current manuscript, these databases even eliminate the necessity of conducting the first two steps of our approach.

Future directions

In the visual search literature, there has been an extensive debate regarding the degree to which salience influences attentional control, particularly in terms of when and whether objects are fixated or inspected by searchers (Tatler, Hayhoe, Land, & Ballard, 2011). Although the role of bottom-up salience was initially rather popular, it has since been found that salience can be over-ridden by top-down information such as goals or target templates (Chen & Zelinsky, 2006; Kunar, Flusberg, & Wolfe, 2008). However, since that time, salience often has been used as a way of invoking careful controls of the visual characteristics of a given set of stimuli. One important difference between saliency approaches and the methods described in this article, it should be noted, are the

relationships that the two approaches try to quantify. Saliency models typically quantify how different a current location in a display is, relative to rest of that display. The logic is that points in space that are highly dissimilar from their surroundings are more salient; that is, they are more likely to attract attention. These models have been applied to real-world search and have been shown to successfully predict, for instance, the likelihood that a given location will be fixated by an observer (Itti, 2005; Itti & Koch, 2001).

By contrast, the approach laid out in this article quantifies the similarity relationships between sets of objects, and these objects need not be present in the same visual environment. That is, this approach is not limited to studying bottom-up influences on attention. Indeed, our technique could be used to study bottom-up saliency, for instance, by controlling how similar a set of distractors are to one another (thereby making distractor arrays that are more or less homogeneous; cf. Duncan & Humphreys, 1989; Avraham, Yeshurun, & Lindenbaum, 2008). Alternatively, it could be used to study top-down attention, as in the Hout & Goldinger (2015) study wherein we manipulated the visual similarity between the target image a searcher experienced and their mental representation of that item derived from a somewhat different looking cue (Zelinsky, 2008, for a computational approach to studying mental representations). Thus, one potential future avenue for the use of our MDS approach would be to use similarity-based metrics to serve as an additional, complementary tool for controlling experimental stimuli in a wide range of visual-cognitive tasks (Borji & Itti, 2013, and Borji, Sihite, & Itti, 2013 for more discussion of saliency modeling).

In reading (Rayner, 1998, 2009), examining eye movements has provided detailed insights into the moment-to-moment processing that takes place as words are read and integrated to form a coherent understanding of the sentence content (Liversedge & Findlay, 2000). Currently in reading research, the measure of visual similarity is quite coarse. A common measure of visual similarity is quantified by examining the size of a word's orthographic neighbourhood. An orthographic neighbor is defined typically as a word that differs from a target word by a single letter, such as *house* and *mouse* (Coltheart, Davelaar, Jonasson & Besner, 1977). The visual similarity of words during reading has shown to have an impact on reading behavior. Perea and Pollatsek (1998), for instance, found that when a target word had a high frequency (i.e., a word that appears regularly in English text) neighbor, the reader would sometimes mistakenly identify the target word as its higher-frequency neighbor and this would cause re-reading of the text to correct the mistake and make sense of the context. This finding suggests that similar words can be misidentified as visually similar words. One potential future use of our approach for reading (and word identification) studies therefore would be to control the visual similarity among word stimuli, by using the MDS ratings of individual letters. This method could provide a novel addition to existing word similarity measures and be useful for research in reading. By incorporating an MDS-based measure of similarity among letters, similarity between words can be more fine-grained than such a coarse measure currently allows.

Conclusions

The concept of similarity is a pivotal aspect of many psychological studies, particularly those pertaining to visual search (Duncan & Humphreys, 1989; Treisman & Gelade, 1980). MDS is a simple and robust way of quantifying similarity scores between stimuli on any dimension(s). Data collection methods with regard to MDS are easy to implement (Hout, Goldinger & Ferguson, 2013), and the analysis tools for MDS are already included in many popular statistical programs (e.g., SPSS, R, or the statistics toolbox for Matlab). Although combining MDS with standard behavioral measurements (e.g., RTs) is useful, we feel that the inclusion of eye-tracking is a particularly fruitful avenue for future research. Eye-tracking analyses have the ability to uncover more nuanced effects of similarity in visual search (Hout, Goldinger & Brady, 2014; Maxfield & Zelinsky, 2012; Stroud et al., 2012). For instance, by deconstructing the search RT into periods of scanning behavior (i.e., how efficiently are the eyes guided to the target) and decision-making behavior (i.e., how quickly targets identified following fixation, or how quickly distractors are rejected), researchers can better elucidate the manner in which similarity relationships affect one's ability to perform a search. Moreover, computational approaches to similarity also hold a great deal of promise; by combining human ratings with those derived from computer vision, researchers have the potential to greatly inform theories of visual search. Computational approaches may even someday prove to be better predictors of human behavior, and thereby supplant the need for human ratings. For now, more generally, we hope that the use of MDS in psychological studies will soon be commonplace, ensuring the ease and accuracy of quantifying similarity to better examine theories in visual search and beyond.

Finally, throughout the present review, we have focused on outlining how MDS ratings can be used to quantify the visual similarity between objects. It also is possible to use MDS to go beyond the visual modality and compute the similarity between other forms of stimuli, objects, concepts, and information. For example, in a recent study, Montez, Thompson, and Kello (2015) asked participants to recall as many animal names as possible within a given time period. The purpose in doing so was to examine memory recall. Participants arranged the animal names on a whiteboard and were then asked to categorize groups of animal names. Again,

although there was a visual component to this task, the core usage of MDS was that it enabled the researchers to quantify the *semantic* rather than *visual* similarity between items. This demonstrates that MDS is not just beneficial in the visual domain, but in other domains and modalities as well. Indeed, it may be the case that MDS will ultimately prove to be highly beneficial beyond the visual modality, enabling researchers to capture, for example, both the visual and semantic similarity between objects and items together.

Acknowledgments H. G. and T. M. were supported by funding from the Economic and Social Sciences Research Council (grant ref. ES/I032398/1). SDG was supported by NIH / NICHD grant R01 HD075800-02. The authors would like to thank Greg Zelinsky, Justin Maxfield, and Carrick Williams for their helpful comments on an earlier version of this manuscript.

References

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

Avraham, T., Yeshurun, Y., & Lindenbaum, M. (2008). Predicting visual search performance by quantifying stimuli similarities. *Journal of Vision, 8,* 1–22.

Becker, S. I. (2011). Determinants of dwell time in visual search: Similarity or perceptual difficulty? *PLoS One, 6*, e17740.

Berman, M. G., Hout, M. C., Kardan, O., Hunter, M. R., Yourganov, G., Henderson, J. M., … Jonides, J. (2014). The perception of naturalness correlates with low-level visual features of environmental scenes. *PLoS One, 9,* e114572. doi: 10.1371/journal.pone.0114572

Biggs, A. T., & Mitroff, S. R. (2014). Improving the efficacy of security screening tasks: A review of visual search challenges and ways to mitigate their adverse effects. *Applied Cognitive Psychology*. doi: 10.1002/acp.3083

Bishop, C. M. (2008). *Pattern recognition and machine learning. In Information Science and Statistics*. Berlin: Springer.

Blough, D. S. (1988). Quantitative relations between visual search speed and target-distractor similarity. *Perception & Psychophysics, 43,* 57–71.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer-Verlag.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35,* 185–207.

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Transactions on Image Processing, 22,* 55–69.

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.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika, 35,* 283–319. doi: 10.1007/BF02310791

Chan, L. K. H., & Hayward, W. G. (2013). Visual search. *WIREs Interdisciplinary Reviews: Cognitive Science, 4,* 415–429.

Chen, X., & Zelinsky, G. J. (2006). Real-world search is dominated by top-down guidance. *Vision Research, 46,* 4118–4133.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation.

Coltheart, M., Davelaar, E., Jonasson, J. F., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention & performance VI* (pp. 535–555). Hillsdale: Erlbaum.

Corbett, J. E., Oriet, C., & Rensink, R. A. (2006). The rapid extraction of numeric meaning. *Vision Research, 46,* 1559–1573.

Davis, E. T., & Palmer, J. (2004). Visual search and attention: an overview. *Spatial Vision, 17,* 249–255.

de Campos, T., Csurka, G., & Perronnin, F. (2012). Images as sets of locally weighted features. *Computer Vision and Image Understanding, 116,* 68–85.

Dowd, E. W., & Mitroff, S. R. (2013). Attentional guidance by working memory overrides saliency cues in visual search. *Journal of Experimental Psychology: Human Perception & Performance, 39,* 1786–1796. doi: 10.1037/a0032548

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458. doi: 10.1037/0033-295X.96.3.433

Duncan, J., & Humphreys, G. W. (1992). Beyond the search surface: visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 578–588. doi: 10.1037/0096-1523.18.2.578

Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., Pinto, Y., … Wolfe, J. M. (2011). Visual attention. *WIREs Interdisciplinary Reviews: Cognitive Science, 2,* 503-514.

Faye, P., Brémaud, D., Durand Daubin, M., Courcoux, P., Giboreau, A., & Nicod, H. (2004). Perceptive free sorting and verbalization tasks with naive subjects: an alternative to descriptive mappings. *Food Quality and Preference, 15,* 781–791. doi: 10.1016/j.foodqual.2004.04.009

Giguere, G. (2006). Collecting and analyzing data in multidimensional scaling experiments: a guide for psychologists using SPSS. *Tutorials in Quantitative Methods for Psychology, 2,* 26–37.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67. doi: 10.1037/0033-295X.91.1.1

Gilmore, G. C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics, 25,* 425–431. doi: 10.3758/BF03199852

Godwin, H., Hout, M. C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review, 21,* 689–695. doi: 10.3758/s13423-013-0547-4

Godwin, H. J., Menneer, T., Cave, K. R., Helman, S., Way, R. L., & Donnelly, N. (2010a). The impact of relative prevalence on the dual-target cost for threat items in airport X-ray baggage search. *Acta Psychologica, 134,* 79–84.

Godwin, H. J., Menneer, T., Cave, K. R., & Donnelly, N. (2010b). Searching for high and low prevalence X-ray threat targets. *Visual Cognition, 18,* 1439–1463.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279. doi: 10.1037/0033-295X.105.2.251

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review, 11*, 716–722. doi: 10.3758/BF03196625

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in cross-race face learning: insights from eye movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1105–1122. doi: 10.1037/a0016548

Goldstone, R. L. (1994a). The role of similarity in categorization: providing a groundwork. *Cognition, 52,* 125–157. doi: 10.1016/0010-0277(94)90065-5

Goldstone, R. L. (1994b). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers, 26,* 381–386. doi: 10.3758/BF03204653

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology, 23,* 222–262. doi: 10.1016/0010-0285(91)90010-L

Goldstone, R. L., Medin, D. L., & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition, 25,* 237–255. doi: 10.3758/BF03201115

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130,* 116–139. doi: 10.1037/0096-3445.130.1.116

Green, P. E., Camone, F. J., Jr., & Smith, S. M. (1989). *Multidimensional scaling: Concepts and applications*. Needham Heights: Allyn and Bacon.

Hahn, U. (2014). Similarity. *WIREs Interdisciplinary Reviews: Cognitive Science, 5,* 271–280. doi: 10.1002/wcs.1282

Helbren, E., Halligan, S., Phillips, P., Boone, D., Fanshawe, T. R., Taylor, S. A., Manning, D., Gale, A. G., Altman, D. G., & Mallett, S. (2014). Towards a framework for analysis of eye-tracking studies in the three dimensional environment: a study of visual search by experienced readers of endoluminal CT colonography. *The British Journal of Radiology, 87*, pp.20130614 1-20130614 6.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7,* 498–504. doi: 10.1016/j.tics.2003.09.006

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243–271. doi: 10.1146/annurev.psych.50.1.243

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

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428. doi: 10.1037/0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551. doi: 10.1037/0033-295X.95.4.528

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review, 8,* 761–768.

Horst, J. S., & Hout, M. C. (*in press*). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*.

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics, 72,* 1267–1282. doi: 10.3758/APP.72.5.1267

Hout, M. C., & Goldinger, S. D. (2012). Incidental learning speeds visual search by lowering response threshold, not by improving efficiency: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 90–112. doi: 10.1037/a0023894

Hout, M. C., & Goldinger, S. D. (2015). Target templates: the precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics, 77,* 128–149. doi: 10.3758/s13414-014-0764-6

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: a multidimensional scaling database with similarity ratings for 240 object categories from the Massive Memory picture database. *PLoS One, 9*, e112644. doi: 10.1371/journal.pone.0112644

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 (WIREs): Cognitive Science, 4, 93-103. doi: 10.1002/wcs.1203

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,* 1–18. doi: 10.1167/9.5.25

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye-movements in dynamic scenes. *Visual Cognition, 12,* 1093–1123.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203.

Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology, 5,* 1–10.

Koch, C., & Ulllman, S. (1985). Shifts in selective visual attention: Toward the underlying neural circuitry. In L. M. Vaina (ed.), *Matters of Intelligence,* 115-141. Reidel Publishing Company.

Kriegeskorte, N., & Marieke, M. (2012). Inverse MDS: inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology, 3,* 1–12.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 1–9*.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling. Sage university paper series on quantitative applications in the social sciences* (pp. 07–011). Beverly Hills: Sage Publications.

Kunar, M. A., Flusberg, S., & Wolfe, J. M. (2008). The role of memory and restricted context in repeated visual search. *Perception & Psychophysics, 70,* 314–328. doi: 10.3758/PP.70.2.314

Lee, M. D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modelling. *Journal of Mathematical Psychology, 45,* 149–166. doi: 10.1006/jmps.1999.1300

Liversedge, S., & Findlay, J. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4*(1), 6–14.

Luria, S. M., & Strauss, M. S. (1975). Eye movements during search for coded and uncoded targets. *Perception and Psychophysics, 17,* 303–308. doi: 10.3758/BF03203215

Maxfield, J. T., Stalder, W. D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision, 14,* 1–11.

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

McFrederick, Q. S., Wcislo, W. T., Hout, M. C., & Mueller, U. G. (2014). Host developmental stage, not host sociality, affects bacterial community structure in socially polymorphic bees. *FEMS Microbiology Ecology, 88,* 398–406. doi: 10.1111/1574-6941.12302

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278.

Menneer, T., Cave, K. R., & Donnelly, N. (2009). The cost of search for multiple targets: effects of practice and target similarity. *Journal of Experimental Psychology: Applied, 15,* 125–139. doi: 10.1037/a0015331

Menneer, T., Donnelly, N., Godwin, H. J., & Cave, K. R. (2010). High or low target prevalence increases the dual-target cost in visual search. *Journal of Experimental Psychology: Applied, 16,* 133–144.

Menneer, T., Stroud, M. J., Cave, K. R., Li, X., Godwin, H. J., Liversedge, S. P., & Donnelly, N. (2012). Search for two categories of target produces fewer fixations to target-color items. *Journal of Experimental Psychology: Applied, 18*(4), 404–418.

Migo, E. M., Montaldi, D., & Mayes, A. R. (2013). A visual object stimulus database with standardized similarity information. *Behavior Research Methods, 45,* 344–354.

Montez, P., Thompson, G., & Kello, C. T. (2015). The Role of Semantic Clustering in Optimal Memory Foraging. *Cognitive Science,* 1–29. doi: 10.1111/cogs.12249

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology, 4,* 1–22.

Neider, M. B., & Zelinsky, G. J. (2006). Searching for camouflaged targets: effects of target-background similarity on visual search. *Vision Research, 46,* 2217–2235.

Newton, I. (1704). *Opticks*. London: Smith and Walford.

Nosofsky, R. M. (1986). Attention, similiarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57. doi: 10.1037/0096-3445.115.1.39

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300. doi: 10.1037/0033-295X.104.2.266

Oh, M.-S. (2011). A simple and efficient Bayesian procedure for selecting dimensionality in multidimensional scaling. *Journal of Multivariate Analysis, 107,* 200–209. doi: 10.1016/j.jmva.2012.01.012

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

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40,* 1227–1268.

Papesh, M. H., & Goldinger, S. D. (2010). A multidimensional scaling analysis of own- and cross-race face spaces. *Cognition, 116,* 283–288. doi: 10.1016/j.cognition.2010.05.001

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 767–779.

Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42,* 1447–1463. doi: 10.1016/S0042-6989(02)00040-8

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi: 10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62,* 1457–1506. doi: 10.1080/17470210902816461

Riesbeck, C., & Schank, R. (1989). *Inside case-based reasoning*. NJ: Lawrence Erlbaum Associates.

Ristin, M., Gall, J., Guillaumin, M., & Van Gool, L. (2015). From categories to subcategories: large-scale image classification with partial class label refinement. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 231–239*.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional scaling approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9,* 283–294. doi: 10.1037/h0026086

Rutishauser, U., & Koch, C. (2007). Probabilistic modelling of eye movement data during conjunction search via feature-based attention. *Journal of Vision, 7,* 1–20.

Salton, G. (1991). Developments in automatic text retrieval. *Science, 253,* 974–980.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology, 62,* 1904–1914. doi: 10.1080/17470210902853530

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. (1962a). The analysis of proximities: multidimensional scaling with an unknown distance function. *Part I. Psychometrika, 27,* 125–140. doi: 10.1007/BF02289630

Shepard, R. N. (1962b). The analysis of proximities: multidimensional scaling with an unknown distance function. *Part II. Psychometrika, 27,* 125–140. doi: 10.1007/BF02289621

Shepard, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors, 5,* 33–48. doi: 10.1177/001872086300500104

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210,* 390-398. doi: 10/1126/science.210.4468.390.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323. doi: 10.1126/science.3629243

Shepard, R. N. (2004). How a cognitive psychologist came to seek universal laws. *Psychonomic Bulletin & Review, 11,* 1–23. doi: 10.3758/BF03206455

Shepard, R. N., & Arabie, P. (1979). Additive clustering: representation of similarities as combinations of discrete overlapping properties. *Psychological Review, 86,* 87–123. doi: 10.1037/0033-295X.86.2.87

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology, 7,* 82–138. doi: 10.1016/0010-0285(75)90006-7

Shiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling: Theory, methods, and applications*. New York: Academic Press.

Solan, Z., & Ruppin, E. (2001). Similarity in perception: a window to brain organization. *Journal of Cognitive Neuroscience, 13,* 18–30. doi: 10.1162/089892901564144

Spencer-Smith, J., & Goldstone, R. L. (1997). The dynamics of similarity. *Bulletin of the Japanese Cognitive Science Society, 4,* 38–56.

Stroud, M. J., Menneer, T., Cave, K. R., Donnelly, N. and Rayner, K. (2011). The reduction of color selectivity in search for multiple targets: Evidence from eye movements. *Applied Cognitive Psychology.*

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.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of Vision, 11,* 1–23. doi: 10.1167/11.5.5

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Treisman, A. (1991). Search, similarity, and the integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 252–276. doi: 10.1037/0096-1523.17.3.652

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A. M., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review, 95,* 15–48.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. doi: 10.1037/0033-295X.84.4.327

Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review, 89,* 123–154.

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

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention*. London: University College London Press.

Wolfe, J. M. (2001). Guided Search 4.0: A guided search model that does not require memory for rejected distractors. *Journal of Vision, 1,* 349. doi: 10.1167/1.3.349

Wolfe, J. M. (2010). Visual search. *Current Biology, 20,* 346–349.

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.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44,* 1411–1426. doi: 10.1016/j.visres.2003.11.024

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General, 136,* 623-638. doi: 10.1037/0096-3445.136.4.623

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

Young, F. W., Takane, Y., & Lewyckyj, R. (1978). ALSCAL: A nonmetric multidimensional scaling program with several individual-differences options. *Behavior Research Methods, 10,* 451–453. doi: 10.3758/BF03205177

Yun, K., Peng, Y., Samaras, D., Zelinsky, G. J., & Berg, T., L. (2013). Studying relationships between human gaze, description, and computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 739-746.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115,* 787–835. doi: 10.1037/a0013118

Zelinsky, G. J. (2012). TAM: Explaining off-object fixations and central fixation tendencies as effects of population averaging during search. *Visual Cognition, 20,* 515–545.

Zelinsky, G. J., Adeli, H., Peng, Y., & Samaras, D. (2013a). Modelling eye movements in a categorical search task. *Philosophical Transactions of the Royal Society B, 368,* 20130058.

Zelinsky, G. J., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the New York Academy of Sciences, 1339,* 154–164.

Zelinsky, G. J., Peng, Y., Berg, A. C., & Samaras, D. (2013b). Modeling guidance and recognition in categorical search: bridging human and computer object detection. *Journal of Vision, 13,* 1–20.

Zelinsky, G. J., Peng, Y., & Samaras, D. (2013c). Eye can read your mind: decoding gaze fixations to reveal categorical search targets. *Journal of Vision, 13,* 1–13.

Zhang, W., Samaras, D., & Zelinsky, G. J. (2008). Classifying objects based on their similarity to target categories. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 1856–1861). Austin: Cognitive Science Society.

Zhang, W., Yu, B., Zelinsky, G. J., & Samaras, D. (2005). Object class recognition using multiple layer boosting with heterogeneous features. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2,* 323-330.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2015). Object detectors emerge in deep scene CNNs. *Proceedings of the International Conference on Learning Representations (ICLR),* 1-12.