



# Mid-level feature contributions to category-specific gaze guidance

Claudia Damiano<sup>1</sup> · John Wilder<sup>1</sup> · Dirk B. Walther<sup>1</sup>

Published online: 6 September 2018  
© The Psychonomic Society, Inc. 2018

## Abstract

Our research has previously shown that scene categories can be predicted from observers' eye movements when they view photographs of real-world scenes. The time course of category predictions reveals the differential influences of bottom-up and top-down information. Here we used these known differences to determine to what extent image features at different representational levels contribute toward guiding gaze in a category-specific manner. Participants viewed grayscale photographs and line drawings of real-world scenes while their gaze was tracked. Scene categories could be predicted from fixation density at all times over a 2-s time course in both photographs and line drawings. We replicated the shape of the prediction curve found previously, with an initial steep decrease in prediction accuracy from 300 to 500 ms, representing the contribution of bottom-up information, followed by a steady increase, representing top-down knowledge of category-specific information. We then computed the low-level features (luminance contrasts and orientation statistics), mid-level features (local symmetry and contour junctions), and Deep Gaze II output from the images, and used that information as a reference in our category predictions in order to assess their respective contributions to category-specific guidance of gaze. We observed that, as expected, low-level salience contributes mostly to the initial bottom-up peak of gaze guidance. Conversely, the mid-level features that describe scene structure (i.e., local symmetry and junctions) split their contributions between bottom-up and top-down attentional guidance, with symmetry contributing to both bottom-up and top-down guidance, while junctions play a more prominent role in the top-down guidance of gaze.

**Keywords** Eye movements · Visual attention · Scene perception

Humans have the remarkable ability to form simple and coherent representations of the world despite being bombarded with new information continuously. Selective attention plays a prominent role in filtering and serializing the information for further processing. Eye movements are an overt, easily observable form of deploying spatial attention, allowing for the processing of only the important pieces of a complex environment at any given time (Henderson, 2003).

Human gaze patterns have been the topic of attention research for decades (e.g., Greene, Liu, & Wolfe, 2012; Kowler, Anderson, Doshier, & Blaser, 1995; Liversedge & Findlay, 2000; Rayner, 2009; Yarbus, 1967). Many factors are believed to contribute to the control of human gaze, from the low-level physical features of a stimulus (Borji & Itti, 2013; Bruce & Tsotsos, 2009; Itti, Koch, & Niebur, 1998; Koch & Ullman,

1985; Tatler, Baddeley, & Gilchrist, 2005) to higher-level influences, such as task context or reward (Henderson, 2003; Hollingworth & Henderson, 2000; Navalpakkam & Itti, 2005; Torralba, Oliva, Castelhano, & Henderson, 2006; Wu, Wick, & Pomplun, 2014). However, it is so far unclear what role mid-level vision plays in guiding spatial attention and gaze. *Mid-level* vision is thought to underlie the perceptual organization of low-level features for the subsequent processing by higher-level recognition processes (Koffka, 1935; Wagemans et al., 2012; Wertheimer, 1938). Thus, it should be expected that mid-level vision also contributes to directing attention. Here we measured the contributions of *mid-level* features to content-specific gaze behavior using images of real-world settings. Specifically, we measured the effects of contour junctions and of local-part symmetry on fixation behavior, which allow for the prediction of scene categories. Both of these features are thought to underlie the detection of object parts in complex three-dimensional arrangements of surfaces and objects (Biederman, 1987; Wagemans, 1993).

Contrasts in low-level image properties, such as orientation, intensity, and color, contribute to the salience of images, which drives attentional orientation from the bottom up (Itti

---

✉ Claudia Damiano  
claudia.damiano@mail.utoronto.ca

<sup>1</sup> Department of Psychology, University of Toronto, Toronto, Ontario, Canada

et al., 1998; Parkhurst, Law, & Niebur, 2002). A model of computing salience based on these low-level features, first proposed by Koch and Ullman (1985), was implemented in a computational algorithm by Itti et al. (1998). This model was able to account, in part, for the gaze patterns made by humans when viewing complex real-world scenes (Itti & Koch, 2000; Peters, Iyer, Itti, & Koch, 2005). Fixated regions also tend to have higher spatial frequencies (Tatler et al., 2005) and contrast (Reinagel & Zador, 1999) than do locations that are not fixated. Thus, salience is certainly important for driving initial eye movements, but it can be quickly overridden by task goals, instructions, or the semantic scene context (Castelhano, Mack, & Henderson, 2009; Wu et al., 2014). Since the development of the initial salience map implementations, many advances have been made on gaze-predicting algorithms by incorporating task-related goals into the predictions (Torralba et al., 2006), as well as semantic knowledge about the object and scene content (Henderson & Hayes, 2017). Both low- and high-level forms of information are indicative of scene categories. In other words, feature distributions, such as the distribution of spatial frequency information (Berman, Golomb, & Walther, 2017; Oliva & Torralba, 2006) or objects (Greene, 2013; Oliva & Torralba, 2007), differ across categories and remain stable within a category. It is therefore not surprising that such features guide gaze in a category-specific manner. Indeed, our group has recently shown that scene categories are predictable from human fixations when observers viewed color photographs (O'Connell & Walther, 2015), indicating that the factors that guide overt attention are category-specific.

The past several decades in this field of research have revealed many factors that influence human gaze in one way or another. However, although the influences of low- and high-level properties on overt attention are well-documented, where does mid-level vision fit into the story? From their initial conception, Gestalt grouping principles such as parallelism, symmetry, similarity, proximity, and good continuation have been instrumental in informing our ideas about perceptual organization (Koffka, 1935; Wertheimer, 1938; for an extensive review, see Wagemans et al., 2012). Perceptual organization using Gestalt grouping rules allows for the detection of likely object locations and figure-ground segmentation. Both concepts depend on the idea of nonaccidental properties, such as symmetry (Wagemans, 1993) and contour junctions (Biederman, 1987), which serve as cues for edges, depth, and surface boundaries, giving rise to object and scene recognition.

Of particular relevance to the present study, both junctions and local symmetry distributions are also indicative of scene categories, such that distributions of junctions and local symmetry differ across category boundaries (Walther & Shen, 2014; Wilder, Rezanejad, Dickinson, Jepson, et al., 2017b; Wilder et al., 2018, Local symmetry facilitates scene processing, under review).

For example, all beach scenes contain a similar layout, and thus a similar spatial distribution of junctions, which differs from the junction distribution in, say, forest scenes. In fact, the decoding of scene categories from brain activity in scene-selective cortex is compromised severely when junctions are perturbed (Choo & Walther, 2016). Similarly, when the most symmetric parts of a scene image are removed, human categorization accuracy decreases substantially more than when the most asymmetric parts are removed (Wilder et al., 2018, Local symmetry facilitates scene processing, under review). It is unclear whether these mid-level features also guide human gaze in a category-specific manner, but given their importance for scene perception and that they are the perceptual cues for objects and scenes, we hypothesized that they do indeed play a role in category-specific gaze guidance. We intended to discover that role by placing it within the context of bottom-up and top-down influences on human gaze exploration of natural scenes.

As we mentioned above, our group was able to predict scene categories on the basis of human fixations when viewing color photographs (O'Connell & Walther, 2015). O'Connell and Walther found that category-specific gaze guidance involved both bottom-up and top-down attention. They showed that the shape of the time course of prediction accuracy over the course of a trial is diagnostic for the differential contributions of bottom-up and top-down factors. The initial, early contribution of bottom-up attention to category prediction accuracy quickly falls off within the first few hundred milliseconds. This sharp decrease in prediction accuracy is followed by a slow but steady increase, which is dominated by top-down influences (O'Connell & Walther, 2015).

Using the same time-resolved method of measuring the differential contributions to attentional guidance, we here demonstrate the influence of low-level, mid-level, and high-level components of visual information on gaze behavior. Considering how quickly scene perception occurs, it is possible that category-specific mid-level features attract overt attention rapidly and contribute more to the bottom-up influence. Conversely, other salient features, such as color and contrast, are well-known to capture bottom-up attention (Itti et al., 1998; Reinagel & Zador, 1999), and perhaps they mask the role of mid-level features at early time points. To discriminate these two possibilities, our paradigm makes use of both photographs and line drawings from six different scene categories (i.e., beaches, forests, mountains, city streets, highways, and offices; see Fig. 1). To establish a baseline behavioral measure, we recorded the eye movements of participants viewing the grayscale photographs of scenes. Then, to minimize the influence of low-level features, we asked observers to view line drawings of the scenes. Line drawings are valuable because they allow us to retain the mid-level features but eliminate the more salient low-level features, while still being easily perceived and rapidly



**Fig. 1** Example scene stimuli of beaches, forests, mountains, city streets, highways, and offices (from left to right). The top row shows grayscale photographs, and the bottom row shows line drawings of the same scenes

categorized (Walther & Shen, 2014) and eliciting the same category-specific neural patterns as photographs (Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011).

We used two measures to determine the role of symmetry and junctions in category-specific attentional guidance. First, the level of accuracy in predicting scene categories from gaze data serves as a measure of how diagnostic the respective features are for category knowledge. We show that low-level features contribute the least to category-specific gaze guidance, and that having full category knowledge contributes the most. Mid-level features, appropriately, rank between purely low-level and high-level knowledge. Second, to evaluate a feature's contributions to bottom-up versus top-down influences on attention, we measured its involvement in the initial sharp decrease in prediction accuracy versus the slow recovery. Note here that our use of the terms “bottom-up” and “top-down” is distinct from our use of the terms “low-level” and “high-level.” The idea that bottom-up guidance refers only to low-level, stimulus-driven attention is outdated and does not stand up to scrutiny (Anderson, 2013; Awh, Belopolsky, & Theeuwes, 2012). It has been shown that high-level knowledge, such as memory of past reward associations (Anderson, 2013; Awh et al., 2012; Marchner & Preuschhof, 2018), also drives attention automatically in a “bottom-up” manner. Thus, here we use the term “bottom-up” to refer to this feedforward automatic mode of attentional guidance, and we use the term “top-down” to refer to a more explicit mode of exploration. We use the term “low-level” to refer to simple image-based features, and “high-level” to refer to semantic content, such as category.

Ultimately, we replicated our previous results concerning the differential contributions of bottom-up and top-down components to gaze guidance (O'Connell & Walther, 2015). Additionally, we showed that low-level features contribute mostly to bottom-up attentional guidance, that high-level knowledge contributes to both bottom-up and top-down guidance, and that mid-level features contribute somewhat to bottom-up but mostly to top-down guidance.

## Method

### Participants

A total of 77 undergraduate students from Ohio State University participated in the experiment. All of the participants had normal or corrected-to-normal vision, provided written informed consent to participate, and received partial course credit as compensation. The experiment was approved by the Institutional Review Board of Ohio State University.

### Stimuli and apparatus

In all, 432 grayscale photographs ( $800 \times 600$  pixels) and their corresponding 432 line drawings were used. The images depicted six scene categories (beaches, forests, mountains, city streets, highways, and offices; 72 images per category) and were the best exemplars of their respective categories out of a set of almost 4,000 images downloaded from the Internet (Torralbo et al., 2013). The line drawings were generated by having artists trace the most salient contours in the photographs (Walther et al., 2011). See Fig. 1 for example images. Each participant viewed a total of 432 different images, which were randomly assigned to be viewed as grayscale photograph (GS) or line drawing (LD), with the constraint that there were an equal numbers of images from each of the six categories in each group. Thus, each participant viewed all available images, but half of these were grayscale photographs and half were line drawings (216 GS and 216 LD).

The stimuli were presented using SR Research's Experiment Builder software, on a CRT monitor with a resolution of  $800 \times 600$  pixels. Participants were seated 50 cm from the monitor, so that the images subtended  $44 \times 33$  degrees of visual angle. Eye movements were recorded from the participants' dominant eye using an EyeLink 1000 eyetracker with a tower-mounted setup, with participants' heads being stabilized on chin-and-forehead rests.

## Procedure

The study consisted of eight blocks (four with grayscale photographs and four with line drawings). Each block consisted of 48 study trials, followed by 12 “new–old” recognition test trials using six images that had been seen in the study phase and six new images that had not yet been studied. The recognition memory task was used to ensure that the participants had explored the images during the study phase. The eyetracker was calibrated before each block. Each study trial began with drift correction, followed by an image of a scene presented for 3 s (see Fig. 2). Each test trial also began with drift correction. Then the image was displayed for 3 s, followed by a response screen, on which participants were asked to indicate whether the image was new or old by fixating the word “New” or “Old” (see Fig. 2) and pressing a button on a game controller to confirm their selection. The position of the response words on the screen (left vs. right of the initial fixation point) was randomized for each trial. Only data from the study phase were analyzed in this study. Test phase data were analyzed separately and published elsewhere (Damiano & Walther, 2018, Distinct roles of eye movements during memory encoding and retrieval, under review).

## Analysis

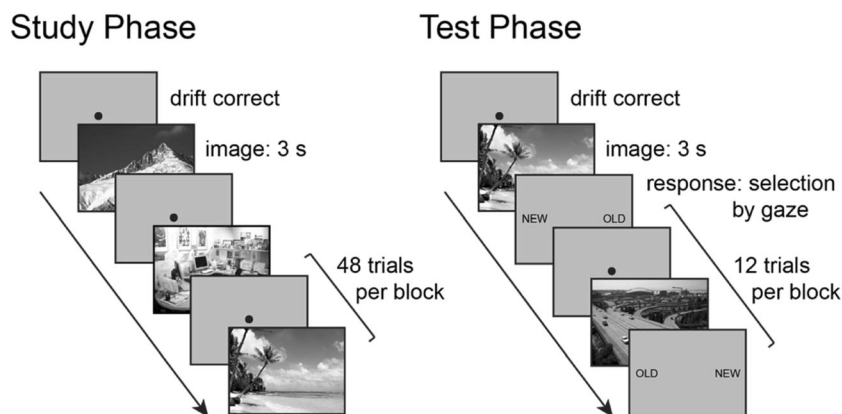
To predict an image’s scene category, we used a classification technique that computed category predictions by comparing each participant’s fixation density on a trial-by-trial basis to the average of the fixation density for each scene category of all other participants (O’Connell & Walther, 2015). This procedure was repeated for each participant in a leave-one-subject-out (LOSO) cross-validation.

**Fixation density maps** To train the classifier, one participant’s gaze data were held out while *fixation density maps* (FDMs) were calculated from all other participants’ fixation data. The first fixation of each trial was removed from the data before

calculating the FDMs, since its location and duration simply reflect carryover effects from the pretrial drift correction. FDMs are the sums of the remaining fixations as a function of their location ( $x$ - and  $y$ -coordinates), weighted by their duration in milliseconds. No outlier duration cutoffs were applied to the data. Category-specific FDMs were created for each of the six scene categories by averaging over all trials of a given category for all participants in the training set (i.e., all participants except the one held out).

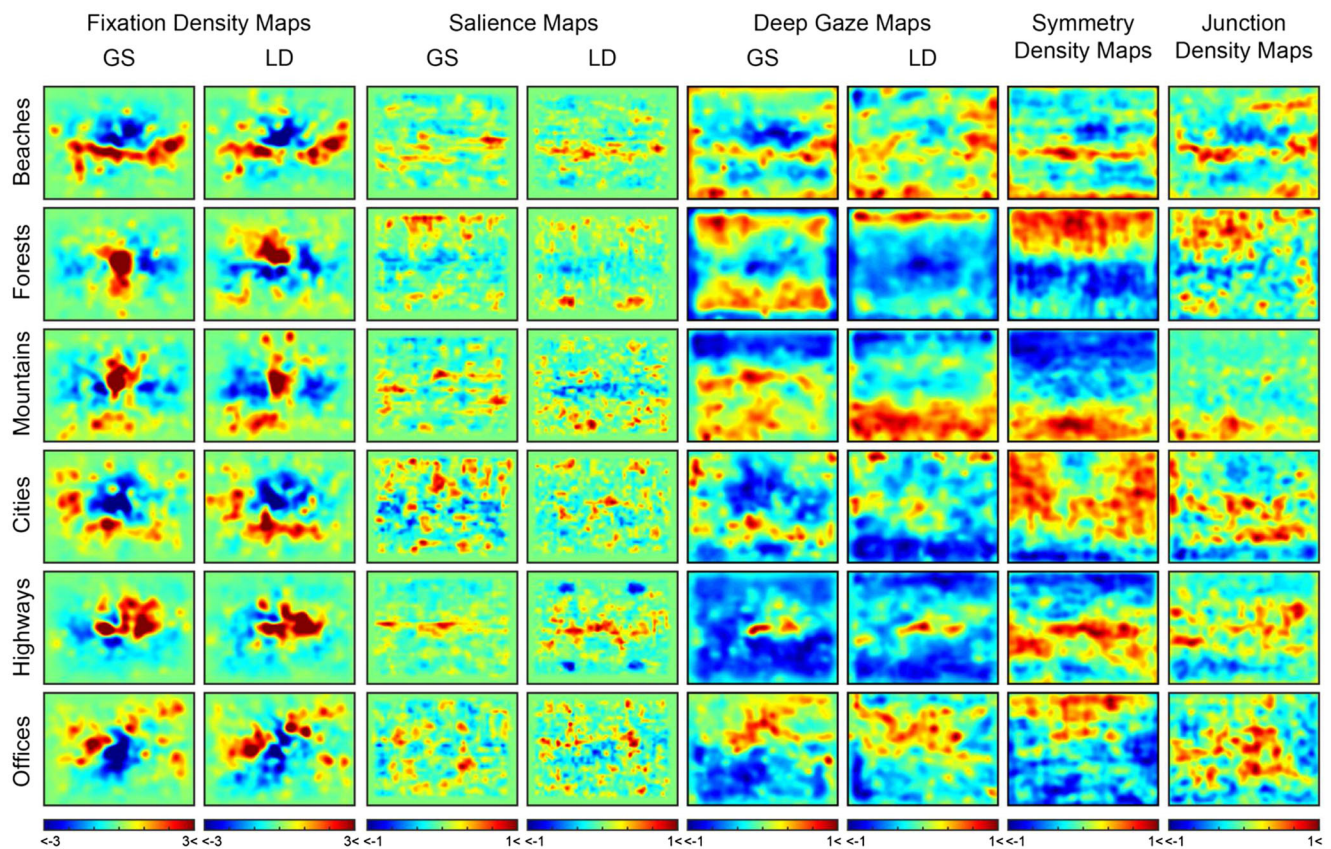
The maps were then smoothed with a two-dimensional Gaussian kernel ( $\sigma = 15$  pixels, equivalent to 0.8 degrees of visual angle, reflecting the eyetracker accuracy) and  $z$ -scored so that positive scores reflected locations that were fixated more often than average across all images from a given category, and negative scores reflected locations fixated less often than average across all images from that category. Marginal FDMs were created by subtracting the grand mean of all FDMs from each category-specific FDM (see Fig. 3, columns 1–2).

**Deep Gaze maps** We inputted each of our scene images into the leading model of gaze prediction, Deep Gaze II (Kümmerer, Wallis, & Bethge, 2016), and obtained a matrix for each image indicating the predicted fixation locations (see Fig. 4a). We used the Deep Gaze II model without center bias, since we had removed center bias from our FDMs and other feature maps, as well. These matrices were then averaged over all images of a given category. The Deep Gaze output maps were already smoothed; thus, the average maps were simply normalized in order to obtain Deep Gaze density maps (DGDMs), wherein positive scores reflected locations that were predicted to be fixated more often than average across all images from a given category, and negative scores reflected locations predicted to be fixated less often than average across all images from that category. The DGDMs were derived separately for photographs and line drawings (see Fig. 3, columns 5–6).



**Fig. 2** Design of the experiment. Participants studied grayscale photographs (GS) and line drawings (LD) of scenes from six categories (beaches, city streets, forests, highways, mountains, and offices,

presented randomly), followed by a recognition memory test. Each block contained either GSs or LDs, but not both



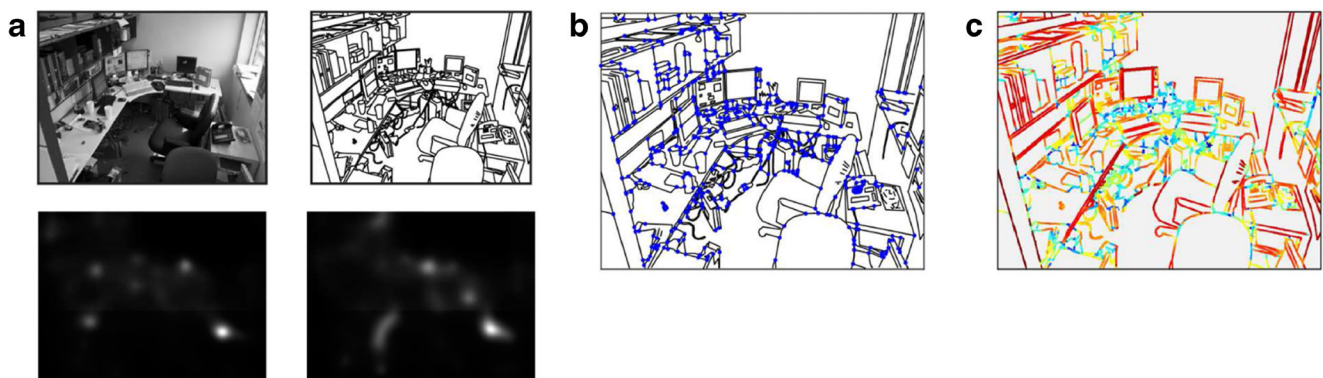
**Fig. 3** Columns 1–2: Average marginal fixation density maps (FDMs) for each category based on participants' fixations when viewing grayscale photographs (GS FDMs) and line drawings (LD FDMs). Marginal FDMs are created by subtracting the grand mean of all FDMs from each category-specific FDM. Columns 3–4: Average marginal saliency maps

(SMs) for grayscale photographs (GS SMs) and line drawings (LD SMs). Columns 5–6: Average marginal Deep Gaze density maps (DGDMs) for grayscale photographs (GS DGDMs) and line drawings (LD DGDMs). Column 7: Average marginal symmetry density maps. Column 8: Average marginal junction density maps

**Saliency maps** The features used by Deep Gaze are a mixture of low-level, mid-level, and high-level features. To investigate low-level features in isolation, we generated simpler saliency maps (SMs) using the Saliency Toolbox (Walther & Koch, 2006). SMs were generated for each image on the basis of luminance and orientation properties and averaged across all images belonging to the same

scene category. The SMs for each category were derived separately for the grayscale photographs and line drawings, since these image types have differences in low-level image properties (see Fig. 3, column 3–4).

**Symmetry density maps** The use of line drawings allowed us to compute the local symmetry of pairs of contours within an



**Fig. 4** (a) Deep Gaze II input and output; white reflects predicted gaze locations. (b) Line drawing with the junctions highlighted in blue for illustration; participants did not see the junctions highlighted. (c) Line

drawings with the most symmetric contours in red and the least symmetric contours in blue; participants did not see the color range during the experiment

image. Symmetry scores are obtained along the medial axis of each pair of contours within an image by calculating the degree of local parallelism between the contours on each side of this medial point. Each pixel on a contour therefore has two possible symmetry scores associated with it, and receives the higher of the two (please see Wilder, Rezanejad, Dickinson, Siddiqi, et al., 2017a, and Wilder et al., 2018, Local symmetry facilitates scene processing, under review, for more details on symmetry score calculations; see Fig. 4c for an example drawing with symmetry scores on each contour pixel)

Each symmetry density map (SymDM) was a sum of all symmetry scores as a function of their location and value. Again, the maps were smoothed and normalized so that positive scores reflected locations that contained more local symmetry than average across all images from a given category, and negative scores reflected locations that contained less symmetry than average across all images from that category (see Fig. 3, column 7).

**Junction density maps** Finally, the line drawings also allowed us to compute the junction locations within each image. Junction locations occurred wherever at least two line segments from separate contours intersected within certain angle parameters, and junctions that occurred within a distance of three pixels were combined into one single junction (for more details on junction computations, see Walther & Shen, 2014; see Fig. 4b for an example drawing with junctions highlighted).

Category junction density maps (JDMs) were computed similarly to the FDMs and other feature maps; each JDM was a sum of all junctions as a function of their location ( $x$ - and  $y$ -coordinates) and value (1 if the pixel contained a junction, otherwise 0), averaged over all images from a given category. The maps were then smoothed and normalized, as above, so that positive scores reflected locations that contained more junctions than average across all images from a given category, and negative scores reflected locations that contained fewer junctions than average across all images from that category (see Fig. 3, column 8).

**Category predictions** To predict the scene category, the FDMs from individual trials were compared to the category FDMs in order to generate a goodness-of-fit score for each category. The goodness-of-fit score was calculated separately for each category by summing the category's chosen map values over the fixated locations on that trial, weighted by fixation duration. This prediction was done with data from the entire trial, as well as with data from only a particular time interval within the trial. Each category prediction was determined by whichever map gave the best goodness-of-fit score. If the prediction matched the trial's true category label, the prediction was correct. This procedure was repeated for each participant. The prediction accuracy was computed as the fraction of trials with correct

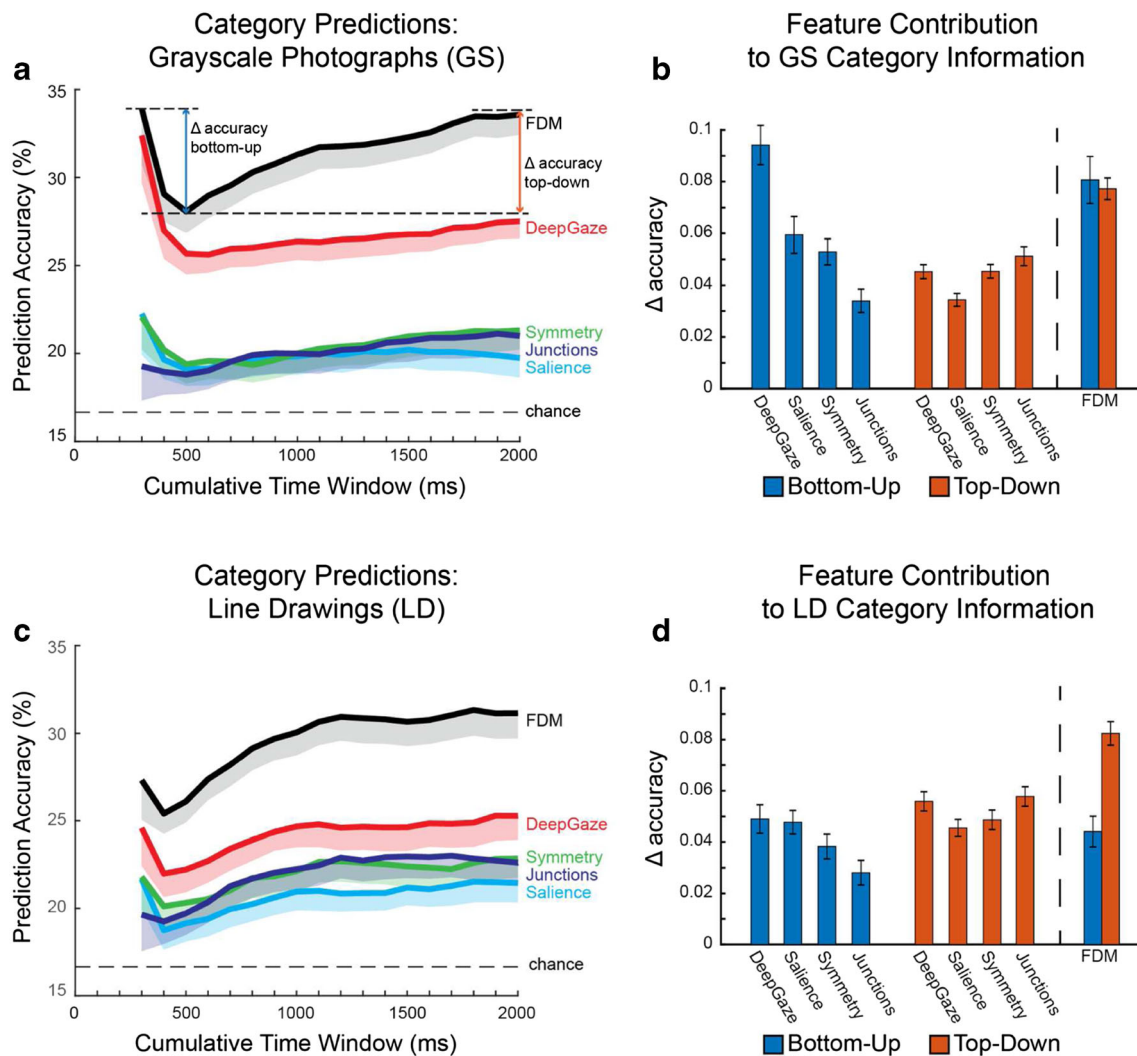
category predictions, and tested for significance using one-sided  $t$  tests comparing the prediction accuracy to chance (1/6). When multiple time intervals were tested, the significance levels were Bonferroni-corrected for multiple comparisons.

This process was also performed on each set of feature maps (i.e., DGDMs, JDMs, SymDMs, or SMs). When the prediction analysis was done using FDMs, each participant's fixations on a trial-by-trial basis were compared to the FDMs derived from all other participants' data. When the category predictions were made using the feature maps, each participants' fixations on a trial-by-trial basis were compared to the reference feature maps (see Fig. 3, columns 3–8).

## Results

First, to confirm that scene categories were predictable from gaze when viewing scene images (both photographs and line drawings), we performed a 77-fold LOSO cross-validation, training and testing on participants' fixations when they viewed grayscale photographs and line drawings, separately. Prediction accuracy was compared to chance (1/6), using a one-tailed  $t$  test to determine whether the spatial distribution of fixations was specific to distinct scene categories. We were able to successfully predict the scene categories from gaze data when using all fixations within the entire trial (3 s), with an accuracy of 33.6% for grayscale photographs and 31.1% for line drawings, significantly above the chance level of 16.7% (1/6), with  $p < .0001$  for both image types (Cohen's  $d$  for grayscale photographs = 3.84; Cohen's  $d$  for line drawings = 3.44). The overall prediction accuracy in these FDM conditions reflect how similarly people behave when viewing different images from the same category. That is, we were able to predict category from gaze to the extent that (1) exemplar images are similar enough within a category and differ from the exemplars from other categories, and (2) viewing behavior is similar across individuals. Thus, our prediction accuracy acts as a ceiling score for the category predictability of fixations.

To examine the time course of category predictability from gaze, we performed a cumulative time bin analysis, with a start time of 0 ms and end times beginning at 300 ms and increasing in increments of 100 ms. Classification was significant at all time bins for both grayscale photographs and line drawings (see Fig. 5a and c). The prediction accuracies begin high and then quickly dip until 400 to 500 ms, followed by an accuracy increase until 2,000 ms, replicating O'Connell and Walther's (2015) previous findings that fixations while viewing images of real-world scenes reflect initial bottom-up attention capture, followed by top-down knowledge control. From 2,000 to 3,000 ms, the prediction accuracy did not increase or decrease, so we have limited the time window to 2,000 ms for the remainder of the analyses.



**Fig. 5** **a** Cumulative time bin category prediction analysis on grayscale photographs. Shaded regions represent the 99% confidence interval. **b** Individual feature contributions to bottom-up and top-down guidance on grayscale photographs. Error bars represent the standard errors of the

means. **c** Cumulative time bin category prediction analysis on line drawings. Shaded regions represent the 99% confidence interval. **d** Individual feature contributions to bottom-up and top-down guidance on line drawings. Error bars represent the standard errors of the means

We used the above-mentioned results from the grayscale photographs to establish a baseline of bottom-up and top-down influence on category-specific gaze guidance. Note that the shape of the prediction curve is of great importance for the interpretation of the results that follow. Namely, the initial drop in accuracy reflects the shift from the greatest degree of bottom-up influence to a decreased degree of that influence for the guidance of gaze. Thus, we calculated the absolute difference between the initial prediction accuracy and the minimum prediction accuracy in order to obtain a contribution score of bottom-up guidance. Conversely, the slow rise that follows reflects the influence of top-down gaze guidance. Again, we calculated the difference between the prediction accuracy at 2 s and the minimum accuracy in order to reveal the level of top-down guidance. In other words, we used these absolute differences to describe the influences of bottom-up and top-down attention on category-specific gaze guidance

(see Fig. 5a for an example of the calculation on the FDM curve; see the FDM column on the right of Fig. 5b for the corresponding difference values). Comparing the levels of influence from these two sources of information using a paired-sample *t* test, we found that, in grayscale photographs, eye movements are influenced equally by bottom-up (BU) and top-down (TD) guidance (BU = .081 vs. TD = .077,  $t(76) = 0.31$ ,  $p = .76$ , Cohen's  $d = 0.06$ ). To confirm that the use of line drawings served to decrease the influence of bottom-up attention, we calculated the same absolute differences as above and found that, whereas the top-down influence remained the same as in grayscale photographs (LD top-down = .082, GS top-down = .077),  $t(76) = 0.83$ ,  $p = .41$ , Cohen's  $d = 0.14$ , the bottom-up influence is much lower (LD bottom-up = .044, GS bottom-up = .081),  $t(76) = 3.95$ ,  $p < .0005$ , Cohen's  $d = 0.54$ ; see Fig. 5b and d, FDM columns. After establishing these baselines, we ran the same analyses using the various

generated feature maps (see the Method section) as references for category predictions. Thus, for each feature we obtained an overall contribution of that feature to gaze guidance (reflected in the prediction accuracy; Fig. 5a and c), as well as its relative contributions to bottom-up versus top-down attentional guidance (Fig. 5b and d).

To test which low-level and mid-level features were guiding gaze in a category-specific manner, we performed the LOSO cross-validation procedure four more times for each image type using our previously generated features maps (i.e., Deep Gaze, salience, symmetry, and junctions) as references. We were able to predict scene categories from fixations for all features at all times throughout the trial, meaning that each feature contributed some category-specific information (i.e., information that varied across categories but remained constant within a category). Please see Fig. 5a and c for the prediction accuracies across all time points, but here we note specific properties of the time course analysis results.

First, we performed a two-way analysis of variance (ANOVA) comparing the overall prediction accuracies (all fixations up to 2 s) for each feature (i.e., DGDMs, SMs, SymDMs, and JDMs) and across image types (i.e., GS and LD), to determine the features' contributions to category-specific gaze guidance. We found a main effect of image type [ $F(1, 608) = 6.14, p < .05, \eta^2 = .007$ ], such that the features contributed more to category-specific gaze guidance in LD (accuracy = 23.1%) than in GS (accuracy = 22.4%). There was also a main effect of feature [ $F(3, 608) = 89.55, p < .0001, \eta^2 = .29$ ]. Post-hoc tests showed that the accuracy was higher with DGDMs (26.4%) than with SymDMs (22.1%), JDMs (21.8%), and SMs (20.6%; all  $ps < .0001$ ). SM accuracy was lower than the accuracies for SymDM ( $p < .001$ ) and JDM ( $p < .01$ ). JDM and SymDM accuracy did not differ significantly ( $p = .89$ ).

The interaction between image type and feature was also significant [ $F(3, 608) = 12.91, p < .0001, \eta^2 = .04$ ]. This was due to the accuracy with DGDMs being higher for GSs (27.5%) than for LDs (25.3%), and the feature accuracy being higher in LDs (SMs = 21.5%; SymDMs = 22.9%; JDMs = 22.6%) than in GSs (SMs = 19.7%; SymDMs = 21.3%; JDMs = 21.0%).

Next, we performed a paired-sample  $t$  test between the FDM and Deep Gaze prediction accuracies at the *first* time point to determine whether they captured similar levels of bottom-up contribution, since Deep Gaze II is the current leading model of salience. We saw that the initial prediction accuracy was close to the baseline (FDM) prediction (Deep Gaze initial accuracy = 32.4%, FDM initial accuracy = 33.9%),  $t(76) = 1.74, p = .09$ , Cohen's  $d = 0.18$ , confirming that the output from Deep Gaze II is extremely similar to the human data at the earliest time point, at least for photographs. Notice that Deep Gaze does not fare as well as humans for LDs (Deep Gaze initial accuracy = 24.6%, LD FDM initial accuracy = 27.3%),  $t(76) = 3.24, p < .005$ , Cohen's  $d = 0.42$ .

Inspecting the overall shapes of the prediction curves, we find that the salience, symmetry, and junctions curves differ slightly from each other but are similar across LDs and photographs. Specifically, the salience curve has a steep decline followed by a relatively flat incline, reflecting that these features (i.e., luminance and orientation) contribute relatively strongly to bottom-up gaze guidance, but little to top-down guidance. The symmetry curve contains both a decline and an incline. Finally, the junctions curve has a relatively flat initial decline, followed by a steadily increasing incline (see Fig. 5a and c). To formally test whether these observable differences in the shapes of the curves were indeed numerically significant, we performed two separate two-way ANOVAs on the absolute accuracy differences (i.e., bottom-up and top-down contributions) with image type (GS and LD) and image feature (salience, symmetry, and junctions) as factors.

For the bottom-up influence, we found a main effect of image type, such that the features contributed more to bottom-up information in GSs ( $M = .049$ ) than in LDs ( $M = .038$ ) [ $F(1, 456) = 6.34, p < .05, \eta^2 = .01$ ]. We also found a main effect of feature [ $F(2, 456) = 9.59, p < .0001, \eta^2 = .04$ ]. The interaction was not significant [ $F(2, 456) = 0.36, p = .70, \eta^2 = .002$ ]. Post-hoc tests on the feature factor reveal that the bottom-up influence of salience ( $M = .054$ ) was similar to that of symmetry ( $M = .046, p = .28$ ). However, both salience and symmetry contributed to bottom-up gaze guidance more than did junctions ( $M = .031, ps < .0001$  and  $.05$ , respectively).

Regarding top-down influence, we also found a main effect of image type, such that the features contributed more to top-down information in LDs ( $M = .051$ ) than in GSs ( $M = .044$ ) [ $F(1, 456) = 6.60, p < .05, \eta^2 = .01$ ]. A main effect of feature was also found [ $F(2, 456) = 9.62, p < .0001, \eta^2 = .04$ ]. The interaction was not significant [ $F(2, 456) = 0.70, p = .50, \eta^2 = .003$ ]. Tests on feature revealed that salience ( $M = .040$ ) has less of a top-down influence than did junctions ( $M = .055, p < .0001$ ). Symmetry ( $M = .047$ ) did not differ from either salience or junctions ( $ps = .083$  and  $.063$ , respectively).

## Discussion

We were able to successfully predict scene categories from the location and duration of fixations obtained from human observers viewing grayscale photographs and line drawings of real-world scenes over a 2-s time period. Using the shapes of the prediction curves, we were able to isolate the degrees of bottom-up and top-down influence of low- and mid-level features on category-specific gaze guidance. Using the FDM results as an anchor, we found that observers are typically influenced by bottom-up and top-down attentional guidance in equal amounts when the relevant stimulus-driven information is available. In contrast, when low-level information is removed from the image (i.e., in line drawings), observers

presumably rely more heavily on their higher-level category knowledge than on low-level information to explore the scene appropriately. The low-level features, as expected, contribute more strongly to bottom-up guidance than do mid-level features in general. Of the mid-level features we tested (symmetry and junctions), symmetry contributes more strongly to bottom-up guidance than does junctions. The trend reverses for top-down guidance, whereby the low-level features have less top-down influence on gaze than do mid-level features. Junctions now have more influence on top-down guidance than symmetry.

It is important to note that we were not predicting *where* people would fixate within the image, as other studies have done (Henderson & Hayes, 2017; Itti & Koch, 2000; Torralba et al., 2006). Instead, we were predicting a *particular aspect of scene content* (in this case, to which scene category the image belonged) from fixations over time. Thus, our prediction accuracy was constrained by two factors: (i) how consistent gaze behavior was across individuals, and (ii) how similar the exemplars of the same scene category were to each other and how different they were from exemplars of a distinct category. For example, if all instances of beach images looked exactly the same, yet completely distinct from all other categories, and all people visually explored in the exact same way, then prediction accuracy would be 100%. However, not all beaches do look the same, some of them may resemble highways or cities, and not all observers explore the same images in the same way. Our prediction accuracy score on FDMs reflected these intrinsic sources of noise. We used this method as a research tool to specifically isolate the feature contributions to overall category-specific gaze guidance. This is distinct from the information we can glean from the shape of the curve, which instead reflects the relative contributions of a particular feature to bottom-up versus top-down forms of attentional guidance. With our method, we first interpret the prediction accuracy from FDMs as a ceiling, and all other prediction accuracies become a measure of how much a feature contributes to overall category-specific gaze guidance. We found that the mid-level features guided category-specific gaze more than did contrast and orientation, but less than having some higher-level category knowledge, such as is the case in the predictions using density maps obtained from the human data.

The features used by Deep Gaze matched the human gazes especially well and explained much of the bottom-up influence. Similar to human performance, the bottom-up contribution of the Deep Gaze II model is smaller for line drawings than for photographs. This is to be expected, because of the decrease of low-level information actually available in the image, and because Deep Gaze was trained on color photographs. Deep Gaze's high similarity to human gaze performance is likely due to its use of the pretrained convolutional neural network (CNN), which is capable of object categorization, followed by the retraining of its final four layers using

scene images. Deep Gaze samples from various levels of the CNN to make its fixation predictions; thus, it does not use purely low-level features to inform its algorithm. This is useful for obtaining high prediction accuracy scores similar to human performance, but unfortunately not as useful for determining which specific features contribute more to bottom-up than to top-down attentional guidance. For that reason, we explicitly computed the low-level and mid-level features directly from the images. Thus, we were able to determine the distinct bottom-up and top-down contributions of a known set of features.

Low-level features are known to act as cues for the most salient information within a particular scene (Itti et al., 1998), which captures overt attention rapidly in a bottom-up fashion (Parkhurst et al., 2002). Higher-level influences, such as the semantically informative regions of a scene, are also known to guide attention early on (Henderson & Hayes, 2017). Top-down guidance, on the other hand, makes use of higher-level categorical knowledge to inform its decisions in a more planned and methodical manner. Categorical knowledge reflects learned regularities of scene structure, which differ across categories but stay relatively stable within a given category, such as the distributions of contour junctions (Walther & Shen, 2014) or local symmetry (Wilder et al., 2018, Local symmetry facilitates scene processing, under review), and even whole objects (Greene, 2013; Oliva & Torralba, 2007). Our findings are consistent with this framework, in which the mid-level features that are useful for perceptual organization contribute to top-down category-specific gaze guidance, although symmetry also contributes to bottom-up guidance. This pattern is observed with both gray-scale photographs and line drawings.

As Wilder et al., 2018, Local symmetry facilitates scene processing (under review) suggested, symmetry serves to signal the visual system to begin grouping visual information into meaningful units. This leads to the rapid emergence of an object advantage (i.e., groups of contours being more easily perceived as a coherent object rather than separate elements) for symmetrical and parallel contours than for asymmetrical ones (Feldman, 2007). Similarly, whole objects are perceived more quickly when the objects' parts contain easily identifiable symmetry cues. Object perception is slower when the object parts must be detected by identifying the location and shape of the curvature and junction points (Panis & Wagemans, 2009). Presumably, this is because junctions cue occlusion boundaries and depth, and thus object separation and scene layout, which take a slightly longer time for the brain to process than lower-level features (Cichy, Khosla, Pantazis, & Oliva, 2017). Ultimately, this contribution of mid-level features to both bottom-up and top-down gaze guidance confirms the idea that human gaze is guided by *interactions* of features within a scene rather than by local features alone (Peters et al., 2005), allowing for contour integration and higher-level scene understanding.

It is useful to interpret our results within the context of the reverse hierarchy theory (Hochstein & Ahissar, 2002). In their theory, Hochstein and Ahissar described two forms of vision, “vision at a glance” and “vision with scrutiny.” “Vision at a glance” happens in an implicit bottom-up manner. The brain combines information, starting from low-level image features to more complex combinations of features, to create a detailed high-level percept, such as an object or scene. The low- and mid-level features are thus integrated within the full percept of the scene, though not as readily available to conscious perception. In our paradigm, this results in the mid-level features achieving a lower overall prediction accuracy than when observers have a more complete knowledge of the category (e.g., FDMs). Note also that the influence of bottom-up contributions decreases in line drawings as compared to photographs: In line drawings, there are fewer features for the brain to combine. Therefore, the final percept will be relatively impoverished, leading to a decreased amount of “vision at a glance” (i.e., bottom-up influence). The second form of vision, “vision with scrutiny,” functions in a top-down (explicit) manner, allowing for a more detailed analysis of the scene. This form of attention is deliberate, acts on the information obtained through “vision at a glance,” and serves to gather additional information to maximize behavioral relevance.

In our study, we were able to determine to what extent mid-level features contributed to the implicit bottom-up versus explicit top-down guidance of attention. The mid-level features contribute especially to top-down attentional guidance, because they correspond to object parts and scene layout, which need to be scrutinized in order to obtain more detailed knowledge of the scene.

Altogether, these findings allow us to further converge on a timeline of scene perception and exploration. Humans are able to perceive the gist of a scene, such as its category or level of navigability, extremely quickly (Greene & Oliva, 2009). Within the constraints of the rapid category knowledge we acquire from the environment through “vision at a glance” (Hochstein & Ahissar, 2002), we are cued to the most relevant objects. Of the relevant objects, the most salient objects, based on their low-level features and local symmetry, are likely explored first. After the most salient objects have been explored, the rest of the scene can then be explored in a logical fashion to attain one’s behavioral goal. In terms of the present study, the behavioral goal was to remember the scene for a later memory test; thus, a logical pattern of fixations after picking out the most salient content would be to follow the layout of the scene to try to extract any other content that might be diagnostic of that particular image. Another task setting with high ecological relevance would be navigation in the real world. To this end, the cues to object parts and structural scene layout provided by mid-level features are imperative (Bonner & Epstein, 2018). Our research has not only been able to

quantify these mid-level features, but we have also shown that these features do indeed guide human gaze in a content-specific manner.

## Limitations

We acknowledge certain limitations that could have an impact on the interpretation of the results. First, it is clear that an extremely salient feature (i.e., color) is missing from our stimulus set. The lack of color potentially affected participants’ fixation behavior on photographs. However, the overall accuracy and shape of the category prediction curve were similar to those found when observers had viewed the same images in color (O’Connell & Walther, 2015), suggesting that the lack of color in our image set was not profoundly impacting how participants viewed those images. Additionally, the exclusion of color had no effect on interpretation of the results from the line drawing data.

However, the lack of color in both photographs and line drawings might have a larger impact on Deep Gaze II. Since Deep Gaze’s network is trained on color photographs, it is unclear how its relevant feature space would decipher impoverished stimuli such as line drawings. Deep Gaze’s network certainly includes low-level and mid-level features that are available in line drawings, but it is difficult to examine the various layers of the network to determine how those features are being interpreted and propagated to later layers. To account for this uncertainty, we also computed simple low-level features using the Saliency Toolbox and mid-level features directly from the stimuli. Though not as accurate as Deep Gaze in matching human behavior, these controls allowed us to have full knowledge of the low- and mid-level features we were using and provided clearly interpretable results for both photographs and line drawings.

Finally, since our paradigm relies on the use of distinct categories in order to obtain successful category predictions, it is difficult to extend our findings on the role of mid-level features in attentional guidance to other categories. Whether mid-level features contribute mostly to top-down gaze guidance for *any and all categories* cannot be concluded by our method, because it requires a small set of distinct categories in order to achieve accurate category predictions. The accurate predictions are imperative, because we derive the relative contributions of mid-level features to bottom-up and top-down guidance of gaze from the time course of the category predictions. To derive this information for a wider and less distinct range of stimuli, the analyses would have to be adapted.

## Conclusion

In summary, we have shown that mid-level features (junctions and symmetry) contribute to the category-specific guidance of gaze—more so than purely low-

level features, but less so than having full knowledge of the scene. There was a shift over time, from the initial use of low-level and simple mid-level (i.e., local symmetry) information to inform bottom-up attentional guidance, to the more prominent use of both simple (symmetry) and more complex mid-level information (i.e., contour junctions) to inform top-down guidance, allowing for the scrutiny of scene structure and layout. Mid-level vision is an important transition point in visual perception that is not yet well-understood. Our findings that mid-level features contribute to the guidance of eye movements underscores the need for a better mechanistic understanding of the Gestalt principles of perceptual organization and their relations and interactions with other functions of the visual system.

## References

- Anderson, B. A. (2013). A value-driven mechanism of attentional selection. *Journal of Vision*, 13(3), 7. <https://doi.org/10.1167/13.3.7>
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16, 437–443. <https://doi.org/10.1016/j.tics.2012.06.010>
- Berman, D., Golomb, J. D., & Walther, D. B. (2017). Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. *PLoS ONE*, 12, e0189828. <https://doi.org/10.1371/journal.pone.0189828>
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115.
- Bonner, M. F., & Epstein, R. A. (2018). Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. *PLoS Computational Biology*, 14, e1006111. <https://doi.org/10.1371/journal.pcbi.1006111>
- Borji, A., & Itti, L. (2013). State-of-the-art in modeling visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 185–207.
- Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3), 5. <https://doi.org/10.1167/9.3.5>
- Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3), 6. <https://doi.org/10.1167/9.3.6>
- Choo, H., & Walther, D. B. (2016). Contour junctions underlie neural representations of scene categories in high-level human visual cortex. *NeuroImage*, 135, 32–44. <https://doi.org/10.1016/j.neuroimage.2016.04.021>
- Cichy, R. M., Khosla, A., Pantazis, D., & Oliva, A. (2017). Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *NeuroImage*, 153, 346–358. <https://doi.org/10.1016/j.neuroimage.2016.03.063>
- Feldman, J. (2007). Formation of visual “objects” in the early computation of spatial relations. *Perception & Psychophysics*, 69, 816–827. <https://doi.org/10.3758/BF03193781>
- Greene, M. R. (2013). Statistics of high-level scene context. *Frontiers in Psychology*, 4, 777. <https://doi.org/10.3389/fpsyg.2013.00777>
- Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers’ task from eye movement patterns. *Vision Research*, 62, 1–8. <https://doi.org/10.1016/j.visres.2012.03.019>
- Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, 20, 464–472. <https://doi.org/10.1111/j.1467-9280.2009.02316.x>
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7, 498–504. <https://doi.org/10.1016/j.tics.2003.09.006>
- Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behavior*, 1, 743. <https://doi.org/10.1038/s41562-017-0208-0>
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791–804. [https://doi.org/10.1016/S0896-6273\(02\)01091-7](https://doi.org/10.1016/S0896-6273(02)01091-7)
- Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, 7, 213–235. <https://doi.org/10.1080/135062800394775>
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506. <https://doi.org/10.1016/S0042-698900163-7>
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1254–1259. <https://doi.org/10.1109/34.730558>
- Koch, C., & Ullman, S. (1985). Shifts in selective visual-attention—Towards the underlying neural circuitry. *Human Neurobiology*, 4, 219–227.
- Koffka, K. (1935). Principles of Gestalt psychology (International Library of Psychology, Philosophy and Scientific Method). New York: Harcourt, Brace & World.
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, 35, 1897–1916. [https://doi.org/10.1016/0042-6989\(94\)00279-U](https://doi.org/10.1016/0042-6989(94)00279-U)
- Kümmerer, M., Wallis, T. S., & Bethge, M. (2016). *DeepGaze II: Reading fixations from deep features trained on object recognition*. arXiv preprint. arXiv:1610.01563
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4, 6–14. <https://doi.org/10.1016/S1364-661301418-7>
- Marchner, J. R., & Preuschhof, C. (2018). Reward history but not search history explains value-driven attentional capture. *Attention, Perception, & Psychophysics*, 80, 1436–1448. <https://doi.org/10.3758/s13414-018-1513-z>
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45, 205–231. <https://doi.org/10.1016/j.visres.2004.07.042>
- O’Connell, T. P., & Walther, D. B. (2015). Dissociation of salience-driven and content-driven spatial attention to scene category with predictive decoding of gaze patterns. *Journal of Vision*, 15(5), 20. <https://doi.org/10.1167/15.5.20>
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, 155, 23–36. [https://doi.org/10.1016/S0079-6123\(06\)55002-2](https://doi.org/10.1016/S0079-6123(06)55002-2)
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11, 520–527. <https://doi.org/10.1016/j.tics.2007.09.009>
- Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization and identification of fragmented object outlines. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 661. <https://doi.org/10.1037/a0013547>
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 107–123. [https://doi.org/10.1016/S0042-6989\(01\)00250-4](https://doi.org/10.1016/S0042-6989(01)00250-4)
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, 45, 2397–2416.
- Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search.

- Quarterly Journal of Experimental Psychology*, 62, 1457–1506. <https://doi.org/10.1080/17470210902816461>
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, 10, 341–350. [https://doi.org/10.1088/0954-898X\\_10\\_4\\_304](https://doi.org/10.1088/0954-898X_10_4_304)
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45, 643–659. <https://doi.org/10.1016/j.visres.2004.09.017>
- Torrallba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113, 766–786. <https://doi.org/10.1037/0033-295X.113.4.766>
- Torrallba, A., Walther, D. B., Chai, B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2013). Good exemplars of natural scene categories elicit clearer patterns than bad exemplars but not greater BOLD activity. *PLoS ONE*, 8, e58594. <https://doi.org/10.1371/journal.pone.0058594>
- Wagemans, J. (1993). Skewed symmetry: A nonaccidental property used to perceive visual forms. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 364–380. <https://doi.org/10.1037/0096-1523.19.2.364>
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, 138, 1172. <https://doi.org/10.1037/a0029333>
- Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19, 1395–1407. <https://doi.org/10.1016/j.neunet.2006.10.001>
- Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, 108, 9661–9666. <https://doi.org/10.1073/pnas.1015666108>
- Walther, D. B., & Shen, D. (2014). Nonaccidental properties underlie human categorization of complex natural scenes. *Psychological Science*, 25, 851–860. <https://doi.org/10.1177/0956797613512662>
- Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). New York: Humanities Press. (Original work published 1923)
- Wilder, J. D., Rezanejad, M., Dickinson, S., Jepson, A., Siddiqi, K., & Walther, D. B. (2017a). The role of symmetry in scene categorization by human observers, Paper presented at the Computational and Mathematical Models in Vision Conference, St. Pete Beach.
- Wilder, J., Rezanejad, M., Dickinson, S., Siddiqi, K., Jepson, A., & Walther, D. (2017b). The perceptual advantage of symmetry for scene perception. *Journal of Vision*, 17(10), 1091. <https://doi.org/10.1167/17.10.1091>
- Wu, C. C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, 5, 54. <https://doi.org/10.3389/fpsyg.2014.00054>
- Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and Vision* (pp. 171–211). New York: Springer. [https://doi.org/10.1007/978-1-4899-5379-7\\_8](https://doi.org/10.1007/978-1-4899-5379-7_8)