Influence of local scene color on fixation position in visual search

Kinjiro Amano* and David H. Foster

School of Electrical and Electronic Engineering, University of Manchester, Sackville Street, Manchester M13 9PL, UK
*Corresponding author: k.amano@manchester.ac.uk

Where observers concentrate their gaze during visual search depends on several factors. The aim here was to determine how much of the variance in observers' fixations in natural scenes can be explained by local scene color and how that variance is related to viewing bias. Fixation data were taken from an experiment in which observers searched images of 20 natural rural and urban scenes for a small target. The proportion R^2 of the variance explained in a regression on local color properties (lightness and the red–green and yellow–blue chromatic components) ranged from 1% to 85%, depending mainly on how well those properties were consistent with observers' viewing bias. When viewing bias was included in the regression, values of R^2 increased, ranging from 62% to 96%. By comparison, local lightness and local lightness contrast, edge density, and entropy each explained less variance than local color properties. Local scene color may have a much stronger influence on gaze position than is generally recognized, capturing significant aspects of scene structure on target search behavior.   © 2014 Optical Society of America

OCIS codes: (330.1720) Color vision; (330.2210) Vision - eye movements; (330.1880) Detection.
http://dx.doi.org/10.1364/JOSAA.31.00A254

1. INTRODUCTION

What decides where we look when searching a scene for an object or target? Depending on the situation, there are several factors known to influence where gaze is concentrated: some are external to the observer, such as the structure of the scene and its surface reflecting properties, and others are internal, such as the nature of the task and the observer's search strategy [1–8].

The relationship between these "bottom-up" and "top-down" factors is complicated [4,7–11]. Most studies of gaze behavior in natural scenes have used gray-scale images and free viewing, that is, without a target. In those conditions, scene features used to explain the positions of observers' fixations have been based on, for instance, edge density [12], intensity contrast [13,14], and intensity bispectra [9]. Models of visual attention have also related gaze behavior to combinations of low-level scene features, including luminous intensity, contrast, color, and orientation [7,8,15,16], although see [17]. Fixation duration, as opposed to fixation position, has also been examined in a relation to mean image luminance [18]. At best, in a nonparametric framework, local features explain approximately 60% of the variance in point of gaze in free viewing of gray-scale images [10]. The remaining variance is usually attributed to the effects of search strategy and other cognitive factors [3,4,19], in either a deterministic or random way [20,21].

Among natural scene properties explaining gaze behavior, color has received relatively little attention, for instance in [8–10,13,14,19,22,23], although see [16,24–26]. In free viewing, patterns of fixations have been reported as being different between colored and gray-scale images of natural scenes [25,27], and in some specific discrimination tasks, the role of color can be decisive, such as in discriminating fruit and fresh foliage from more mature foliage [28,29]. But the general importance of local scene color on gaze behavior in visual search is unclear.

A previous report on target-detection performance in natural scenes estimated that 36–40% of the variance in observers' fixations could be explained by local scene color [30]. That estimate, however, did not include the possible effect on gaze behavior of observers' viewing strategy. In particular, no account was taken of the viewing bias that occurs with circumscribed images of scenes, where, rather than being distributed uniformly, fixations tend to be directed toward the center of the image or display [7,12,31], possibly reinforced by the photographer's bias in scene composition [10]. The effect has also been found in search with abstract displays of geometric elements [32–34]. This central viewing bias could facilitate or inhibit the influence of local scene color on gaze behavior and has been argued to have a larger influence than scene structure itself [12].

The aim of this study was to determine how much of the variance in fixations in natural scenes can be explained by local scene color and how that variance is related to viewing bias. Data were taken from an experiment on visual search [30] in which images of natural scenes were presented on a color monitor. Within each scene, the target to be searched for was a small, shaded, gray sphere matched in mean luminance to its local surround. The observer's gaze position was simultaneously monitored with an infrared video eye-tracker. Fixations were classified from individual observers' gaze data by a method that required no parametric assumptions or expert judgment. Viewing bias was estimated by pooling fixations over both observers and scenes [12].

The analysis was based on multiple linear regressions, which provide a natural framework for quantifying the

variance contributions of different explanatory factors. First, the spatial distribution of fixations for each scene was regressed on the spatial distributions of the local color properties of the scene, namely lightness and the red–green and yellow–blue chromatic components, defined in a particular color appearance space [35]. As a control, observers' viewing bias was regressed on local color properties and then the fixation distribution regressed on both local color properties and viewing bias. Second, to provide an independent measure of scene influence, the distributions of the first, second, and subsequent fixations were compared with each other within and between scenes. Finally, to provide a comparison with other explanatory properties, the distribution of fixations for each scene was regressed separately on the distributions of local lightness, local lightness contrast, edge density, and entropy.

Local color properties were found generally to yield a good explanation of fixation position, better than that by other local achromatic properties. Further, when combined with viewing bias, local color properties accounted for 62%–96% of the variance in observers' systematic fixation behavior.

2. METHODS

The methods of data acquisition have been reported elsewhere [30] and are described here only in abbreviated form. The methods of gaze analysis and regression analysis, however, differ from those in [30] and are therefore described in full.

A. Apparatus

Images were presented on a 20-in. CRT color display (GDM-F520, Sony Corp., Japan) controlled by a graphics workstation (Fuel, Silicon Graphics Inc., California, USA) with spatial resolution 1600×1200 pixels, refresh rate approximately 60 Hz,

and intensity resolution 10 bits on each RGB gun. Observers' point of gaze was recorded with an infrared monocular video eye-tracker (High-Speed Video Eye-tracker Toolbox mk2, Cambridge Research Systems Ltd., Kent, UK), with sampling frequency 250 Hz, generating a sequence of horizontal and vertical coordinates at 4-ms intervals.

B. Visual Stimuli

Twenty natural scenes were rendered from hyperspectral images [36] under daylights with correlated color temperature 6500 K, corresponding to typical daylight. The test target, a gray sphere (Munsell N7), was superimposed digitally. The images on the screen subtended 17×13 deg visual angle and the target approximately 0.25 deg at a viewing distance of 1 m. Images of four example scenes are shown in Fig. 1.

The mean luminance of the images on the screen was 3.6 cd m^{-2} (range 0–61.4 cd m^{-2}). The luminance of the target was matched in mean luminance to its local surround (<1.0 deg extent) to avoid producing accidental chromatic or luminance contrast cues. The small angular subtense of the target was chosen to encourage observers to inspect the entire image. The illumination and shading on the sphere were consistent with that on the scene since the sphere was embedded in the scene at the time of hyperspectral imaging. In Fig. 1, second image, the target is shown arrowed, and in the center of the close-up on the bottom right.

C. Procedure

In each trial, observers were presented with an image of a scene for 1 s, followed by a dark field. Their task was to indicate whether or not they saw the target by pressing a computer mouse button after the image disappeared. They were allowed to move their eyes freely during the trial and had
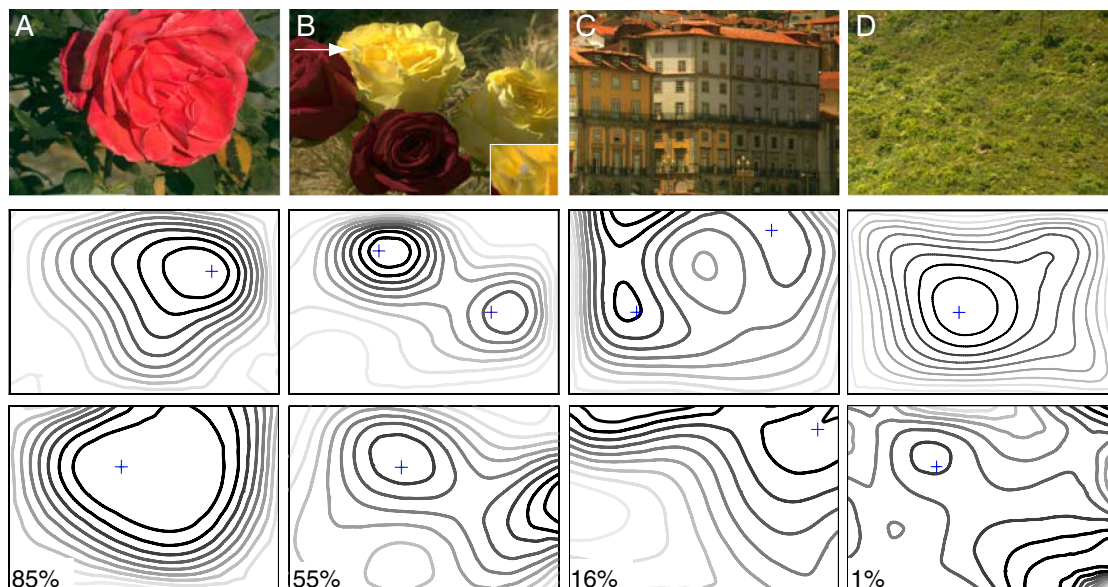

Fig. 1. Examples of natural scene images and observed and fitted spatial distributions of observers' fixations. The first row shows four of the scenes from the 20 used in the experiment, including those producing the best fit (leftmost image) and worst fit (rightmost image). In the second image, the target is shown arrowed, and in the center of the close-up on the bottom right. The middle row shows the smoothed observed fixation distributions. The bottom row shows the estimated fixation distribution obtained by fitting the distributions of local color properties in each scene [see Eq. (1) in text] and the resulting percentage R^2 of variance explained at the bottom left. Values of R^2 were adjusted for the degrees of freedom in each fit. The smoothing was based on a locally weighted quadratic regression (loess) with a Gaussian kernel of 2.5-deg standard deviation. Higher frequencies of fixations are indicated by darker contours. Positions of maxima are indicated by crosses.

unlimited time to respond. No fixation point was displayed either before or during the trial to guide gaze position, which was recorded continuously. Head movement was minimized with a forehead rest and chinrest.

Images derived from the same scenes were presented repeatedly in order to identify the systematic effects of scene structure. Repetition is thought not to affect the influence of basic features of natural images on gaze position [37–40]. The image duration of 1 s in each trial was chosen to limit the observer's fixations to about four, during which scene content has the most influence [7,16,26].

Each scene was tested in 260 trials, constituting one experimental block. Half of the trials, chosen at random, contained the target, and the other half did not. Experimental blocks were divided into four subblocks of 65 trials. For each observer, the eye-tracker was calibrated at the start, in the middle, and at the end of each subblock, and observers were allowed to take a short break between subblocks. In total each observer performed 5200 trials (20 scenes × 260 trials), with additional trials repeated if there was a failure with the eye-tracker.

In each calibration of the eye-tracker, the root-mean-square error between the 20 calibration targets and the observer's corresponding gaze positions was taken as the calibration error. Over observers and scenes, based on 1120 measurements, the mean calibration error was approximately 0.26 deg, with standard deviation 0.06 deg [30,41].

D. Observers
Data were recorded from seven observers (4 female, 3 male, aged 21–31 years). All had normal binocular visual acuity and normal color vision, verified with a series of color vision tests (Farnsworth-Munsell 100-Hue test, Ishihara pseudoisochromatic plates, Rayleigh and Moreland anomaloscopy with luminance test). All of the observers except one were unaware of the purpose of the experiment.

The procedure was approved by the University of Manchester Committee on the Ethics of Research on Human Beings, which operated in accord with the principles of the Declaration of Helsinki.

E. Gaze Analysis
Observers' fixation positions were extracted from the point-of-gaze sequences by a nonparametric classification method [41]. The disadvantage with parametric methods of classification is that they require choices of threshold values for such parameters as speed, acceleration, duration, and stability of point of gaze [31,42–44], all of which may vary with the observer and viewing conditions. The nonparametric method was based on some general distributional properties of eye movements and required neither assumptions about threshold nor expert judgment. As explained in [41], the method was primarily speed-based, but by contrast with existing methods, the optimum speed threshold for classifying saccades, and therefore fixations, was derived automatically from the data for each observer. Because speed-based methods can generate seemingly unphysiologically short fixations [45], an optimum duration threshold, also derived automatically from individual data, was used to eliminate the effects of instrumental noise. Results were verified against those from a parametric method due to Vig *et al.* [46] with hand tuning. The first,

second, third, and fourth fixations after the first saccade were included in the analysis of each trial.

Fixation data from each scene were pooled over observers to reflect the systematic effects of scene structure and viewing bias [47] rather than random variations between observers [21].

To anticipate the results, the number of fixations declined steadily from first to last, with totals over scenes and observers of 34,800, 27,500, 8900, and 1000, respectively. In all, the number of fixations was approximately 72,000.

For computational purposes, the observed fixation positions were assigned to 130 square bins of side 1.5 deg forming an imaginary 13×10 grid in each scene. Pooled over observers, the number of fixations in each bin varied from 0 to 278, with less than 5% of bins empty. These naïve bin estimates of fixation frequency were improved by smoothing [48]. A locally weighted quadratic regression, loess [49], was used as the smoother, but with a relatively large Gaussian smoothing kernel of 2.5-deg standard deviation. This value was chosen under the assumption that the kernel defined the local region within which surface color attributes were processed in parallel [50]. This local regression smoothing should not be confused with the regression modeling on local color properties described in detail in Section 2.H. A Gaussian kernel was used in preference to a tricube kernel [30] but produced similar results.

As a control, the analysis was repeated with larger and smaller values of the standard deviation of the Gaussian kernel. An independent test of oversmoothing was provided by data obtained from successive fixations. A logarithmic transformation of the data produced little change in the fits.

Although the fixation data were acquired during visual search [30], no use was made in this analysis of the recorded detection performance (cf. [51]). Performance was well above chance with discrimination index d' from signal detection theory averaged over observers ranging from 0.2 to 1.9, depending on the scene. The mean d' was 1.2, with a standard deviation of 0.3. The mean hit rate was 0.17, and the mean false-alarm rate was 0.02. For other details see [30], where the relationship between detection performance and fixation is also briefly discussed.

F. Local Color Properties
The local color properties of each scene were calculated according to the CIE color appearance space CIECAM02 [35]. This space was chosen for its appearance attributes and its approximate perceptual uniformity. Among other attributes, CIECAM02 provides lightness J and the red–green and yellow–blue chromatic (chroma) components, a_C and b_C, in a Cartesian coordinate system (cf. [15]). The parameters of the CIECAM02 space were set so that the white point was D65, the luminance of the background was 20% of the white level, and the surround level was "average" [35].

In each scene, values of the three quantities J, a_C, and b_C at each pixel were averaged within each of the 13×10 square bins and then smoothed by the same locally weighted quadratic regression as for fixation position (Section 2.E).

G. Local Contrast, Edge Density, and Entropy
Local achromatic properties of each scene were calculated in the following way. All were derived from the lightness J. Local contrast was taken as the standard deviation of the values of J at each pixel within each of the 13×10 square bins [11,13,22].

Scaling by the mean value of J over the whole scene was omitted since it could not be extended to the chromatic components a_C and b_C, which can have nonpositive values.

Local edge density was taken as the proportion of pixels in each bin that were edge pixels, as determined by a Canny edge detector with parameters as in [52] applied to lightness J. Local entropy was taken as the entropy of the values of J at the pixels within each bin [53,54]. The value of the entropy was estimated by an asymptotically bias-free, k-nearest-neighbor estimator due to Kozachenko and Leonenko [55,56]. This measure was used instead of subband entropy [52], which had been found earlier to account for less variance with the scene images used here [57].

Values were smoothed by the same locally weighted quadratic regression as for fixation position (Section 2.E).

As a control, local color contrast, edge density, and entropy were similarly derived from the three local color properties J, a_C, and b_C at each pixel.

H. Regression Analysis
The regression of fixation position on local color properties was performed per scene in the following way. For each scene i, suppose that $f_i(x, y)$ is the smoothed value of the observed fixation distribution at position (x, y) and that $J_i(x, y)$, $a_{Ci}(x, y)$, and $b_{Ci}(x, y)$ are the corresponding smoothed values of the distributions of lightness and the red–green and yellow–blue chromatic components, respectively (the notation differs somewhat from that in [30]). Then the estimated value $\hat{f}_i(x, y)$ of $f_i(x, y)$ in scene i obtained by fitting $J_i(x, y)$, $a_{Ci}(x, y)$, and $b_{Ci}(x, y)$ to $f_i(x, y)$ is given by

$$\hat{f}_i(x, y) = \alpha_{1,i} J_i(x, y) + \alpha_{2,i} a_{Ci}(x, y) + \alpha_{3,i} b_{Ci}(x, y) + \beta_i, \quad (1)$$

where the scalar parameters $\alpha_{1,i}$, $\alpha_{2,i}$, $\alpha_{3,i}$, and β_i are estimated by minimizing the sum of squared residuals $f_i(x, y) - \hat{f}_i(x, y)$ over (x, y).

The control regression of viewing bias on local color properties was performed analogously. Suppose that $o(x, y)$ is the smoothed value of the viewing bias at location (x, y), assumed to be the same for all scenes and observers and estimated as the average of $\hat{f}_i(x, y)$ over all scenes i [12]. Then the estimated value $\hat{o}_i(x, y)$ of $o(x, y)$ for scene i obtained by fitting $J_i(x, y)$, $a_{Ci}(x, y)$, and $b_{Ci}(x, y)$ to $o(x, y)$ is given by

$$\hat{o}_i(x, y) = \alpha_{1,i} J_i(x, y) + \alpha_{2,i} a_{Ci}(x, y) + \alpha_{3,i} b_{Ci}(x, y) + \beta_i, \quad (2)$$

where the scalar parameters $\alpha_{1,i}$, $\alpha_{2,i}$, $\alpha_{3,i}$, and β_i are generally different from those in Eq. (1).

The regression of fixation position on both local color properties and viewing bias was based on a straightforward extension of Eq. (1). That is, the estimated value $\hat{\hat{f}}_i(x, y)$ of $f_i(x, y)$ in scene i obtained by fitting $J_i(x, y)$, $a_{Ci}(x, y)$, $b_{Ci}(x, y)$, and $o(x, y)$ to $f_i(x, y)$ is given by

$$\hat{f}_i(x, y) = \alpha_{1,i} J_i(x, y) + \alpha_{2,i} a_{Ci}(x, y) + \alpha_{3,i} b_{Ci}(x, y)$$
$$+ \alpha_{4,i} o(x, y) + \beta_i, \quad (3)$$

where the scalar parameters $\alpha_{1,i}$, $\alpha_{2,i}$, $\alpha_{3,i}$, and β_i are generally different from those in Eq. (1) and depend on the scalar parameter $\alpha_{4,i}$.

The regressions of fixation position on other explanatory variables, including local lightness J, local lightness and color contrast, edge density, and entropy, were based on obvious modifications of Eq. (1).

I. Accounting for Variance
There exist many ways of evaluating how well the variance in a set of data may be explained by a regression equation [58], but the proportion R^2 of variance explained is used almost universally ([58], Section 11.2). It involves no assumptions about the distribution of the error and allows the effectiveness of different explanatory variables to be readily assessed, even if they yield an incomplete account of the variance.

If goodness of fit does need to be tested, then under the usual normality assumptions, the residual sum of squares divided by an estimate of the pure error variance should be distributed approximately as χ^2 with an appropriate estimate of the residual degrees of freedom (d.f.) [59]. As noted in [16,60], normality assumptions may well fail, as was found here with most estimates of pure fixation error.

An alternative method of assessing goodness of fit without any distributional assumptions is to compare the value of R^2 with the results of a nonparametric bootstrap analysis [61]. In this bootstrap analysis, values of the proportion R^2 of the variance in fixations in each scene explained by fixations resampled with replacement from the same scene were estimated over 1000 bootstrap iterations. These values of R^2 coincide with the coefficients of determination (the squares of the product moment correlation coefficients) and provide a reference for goodness of fit. Because of the strong contribution of nonscene effects to gaze position, none of the combinations of local scene properties was expected to give a complete account of the variance.

Other approaches to the analysis of the variance were considered. These included the use of the information-theoretic divergence of distributional differences, which loses the connection to a simple regression model, and the use of the area under a receiver operating characteristic (ROC) curve as the performance measure, which requires a classificatory approach [16,47]. These measures were not taken further, although values of the area under the ROC curve (AUC) were estimated for the four example scenes to allow comparison with the R^2 estimates. ROC curves were estimated as in [62] but with saliency modeled by Eq. (1), with no allowance for degrees of freedom, and with raw (i.e., unsmoothed, unbinned) fixation data. An AUC value of 100% corresponds to perfect prediction and 50% to chance level. Some limitations of the measure have been noted in [63,64].

Values of R^2 in each scene i were estimated from the following sums of squared residuals, based on the values $\hat{f}_i(x, y)$ fitted to the observed fixation distribution $f_i(x, y)$, the values $\hat{o}_i(x, y)$ fitted to the observed bias distribution $o(x, y)$ (Section 2.H), and $\bar{f}_i$, the mean of $\hat{f}_i(x, y)$ over (x, y):

$$\mathrm{RSS}_0 = \sum_{x,y} [f_i(x, y) - \bar{f}_i]^2,$$
$$\mathrm{RSS}_1 = \sum_{x,y} [f_i(x, y) - \hat{f}_i(x, y)]^2,$$
$$\mathrm{RSS}_2 = \sum_{x,y} [o(x, y) - \hat{o}_i(x, y)]^2.$$

The proportion of variance in fixations explained by local color properties is given by $R^2 = 1 - \text{RSS}_1/\text{RSS}_0$, and the proportion of variance in viewing bias explained by local color properties is given by $R^2 = 1 - \text{RSS}_2/\text{RSS}_0$. Analogous expressions were used for other combinations of explanatory variables.

The similarities between the distributions of first, second, and subsequent fixations within and between scenes were quantified by the corresponding coefficients of determination, also denoted by R^2, which, as with the bootstrap analysis, give the proportion of the variation in one distribution explained by another.

Providing they remained positive, values of R^2 were adjusted for losses in d.f. in smoothing. If the d.f. of the smoothed distribution is n, computed from the trace of the hat matrix ([59], App. B), and if the number of estimated coefficients is k, then the adjusted value is given ([58], Section 5.2) by $1 - (1 - R^2)(n - 1)/(n - k)$. For example, for Eq. (1), $n = 41.2$ and $k = 4$. Only adjusted values of R^2 are reported in the following.

3. RESULTS AND COMMENT

A. Influence of Local Color Properties and Viewing Bias
Figure 1 shows the observed fixation distributions for four example scenes and fits to those distributions by local color properties. The top row shows the rendered scene image under daylight of correlated color temperature 6500 K; the middle row the observed fixation distribution, with maxima indicated by crosses; and the bottom row the estimated fixation distribution obtained by fitting the distributions of local color properties, with R^2 the proportion of variance explained shown in the bottom left. Higher values are indicated by darker contours. These R^2 values of 85%, 55%, 16%, and 1% corresponded to AUC values of 79%, 74%, 57%, and 53%, respectively.

Figure 2 shows the viewing-bias distribution and the corresponding fits by local color properties for the same four scenes.

The scenes where local color properties gave the best and worst fits to the fixation distributions had R^2 values of 85% and 1%, respectively (the best is in the leftmost column of Fig. 1). The corresponding fits to the viewing-bias distribution had similar R^2 values of 86% and 2%, respectively (the best is in

the leftmost column of Fig. 2). Over all the scenes, the mean value of R^2 for the fixation distributions was 36% and for the bias distribution 32%.

With smaller and larger standard deviations of 2.0 and 3.0 deg for the Gaussian kernel defining the local regions (Section 2.E), the mean values of R^2 for fits of local color properties to the fixation distributions were 33% and 39%, respectively. The corresponding values for fits to the bias distribution were 29% and 35%. A very small standard deviation of 0.5 deg was also tested, but with a local linear rather than quadratic smoother, as there were too few data points in each local region. The mean values of R^2 for fits to the fixation distributions and bias distribution were 21% and 20%, respectively.

To compare these fits of local color properties to fixation distributions with the fits of local color properties to the viewing-bias distribution, the R^2 values were plotted against each other for all 20 scenes. Figure 3 shows the result for the Gaussian kernel with standard deviation of 2.5 deg. The dotted line is a linear regression. The dependence was strong, with the regression accounting for 64% of the variance (reliably greater than chance; bootstrap $p < 0.001$). With the smaller and larger standard deviations of 2.0 and 3.0 deg, the regression explained 60% and 69% of the variance, respectively.

Nevertheless, viewing bias did not dominate performance. There were marked departures from the regression, as indicated in Fig. 3 for the scenes B and C and as revealed in the distributions in Figs. 1 and 2, second and third columns, respectively. These departures might be explained in the following way. Where local scene color varies strongly, as in scene B, it can override an inconsistent viewing bias (i.e., one fitted poorly by local color properties), producing a relatively high R^2 value (but not as high as in scene A, where it is consistent with viewing bias, i.e., fitted well by local color properties). By contrast, where local scene color varies only moderately, as in scene C, it cannot override an inconsistent viewing bias and produces a relatively low R^2 value.

In short, local color properties are generally able to explain fixation position providing that they are consistent with the viewing bias [12]. Unsurprisingly, therefore, when local color properties were combined with the viewing-bias distribution in the fit to fixation distributions, values of R^2 increased, ranging from 62% to 96%, with a mean of 81%. The higher R^2 values
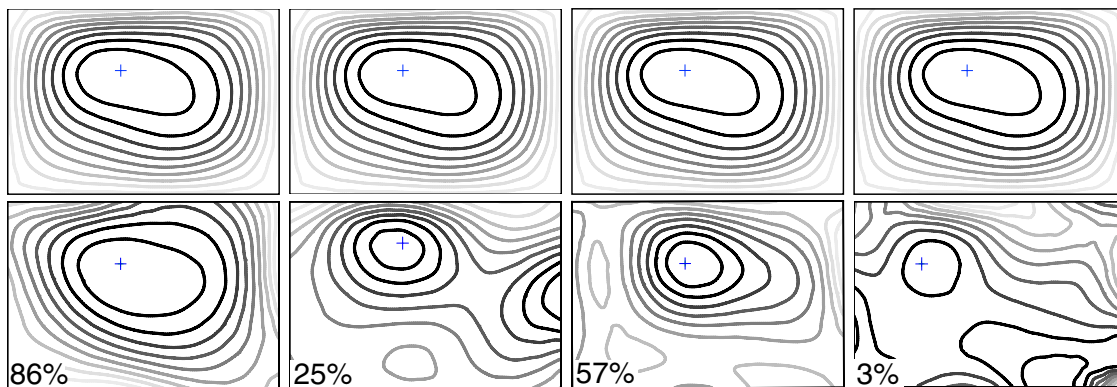

Fig. 2.   Observed and fitted viewing-bias distributions for the scenes shown in Fig. 1. The first row shows the bias distribution estimated by pooling fixations over scenes and observers (and therefore common to all scenes). The second row shows the estimated fixation distribution obtained by fitting the distributions of local color properties in each scene [see Eq. (2) in text]. Other details as for Fig. 1.
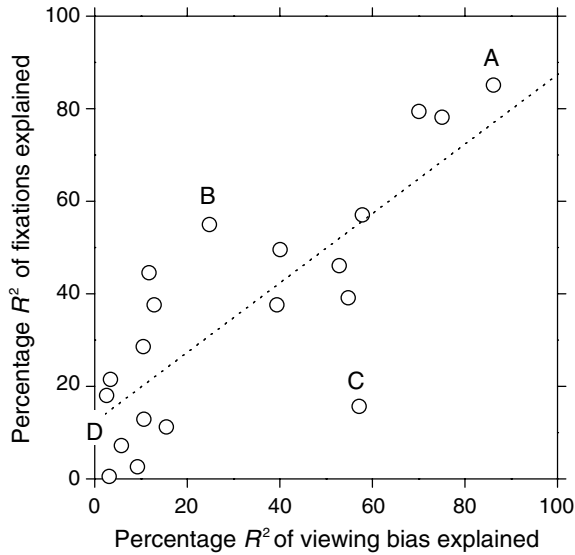
Fig. 3.   Percentage R^2 of variance in fixations explained by local color properties as a function of percentage R^2 of variance in viewing bias explained by local color properties. Each symbol represents a pair of R^2 values for a single scene out of the 20 tested. The dotted line is a linear regression. Data for the scenes in Fig. 1 are shown labeled. Values of R^2 were adjusted for the degrees of freedom in each fit.

may be compared with the bootstrap references for goodness of fit (Section 2.I), for which R^2 ranged from 88% to 97%.

As for the control measurements with the smaller and larger standard deviations of 2.0 and 3.0 deg for the Gaussian

kernel, the mean values of R^2 changed little, at 77% and 84%, respectively. Even with the standard deviation of 0.5 deg, it only fell to 56%.

B. Successive Fixations

Figure 4 shows the observed distributions of first, second, third, and fourth fixations for the same four examples of Fig. 1. Values of the coefficient of determination R^2, that is, the proportion of the variation in one distribution explained by another, averaged over all six possible pairings of successive distributions within a scene (i.e., $3 \times 2 \times 1$), are shown in the bottom left of each plot on the bottom row.

Over all 20 scenes, the mean value of R^2 for successive fixations within scenes ranged from 56% to 92%, whereas between scenes it ranged from 22% to 45%. The two ranges were manifestly disjoint.

By definition, viewing bias was assumed independent of scene. If it did dominate fixations, or, equivalently, if the Gaussian kernel defining the local regions had led to oversmoothing of the fixation distributions, then some overlap of the ranges of R^2 for the distributions within and between scenes should have been evident.

The range of R^2 values just reported for within scenes, namely 56%–92%, may be compared with the range of R^2 values for fits to the same fixation distributions of local color properties and viewing bias, namely 62%–96% (Section 3.A), which were remarkably similar. For any particular scene, however, there need be no connection between the two values of R^2.
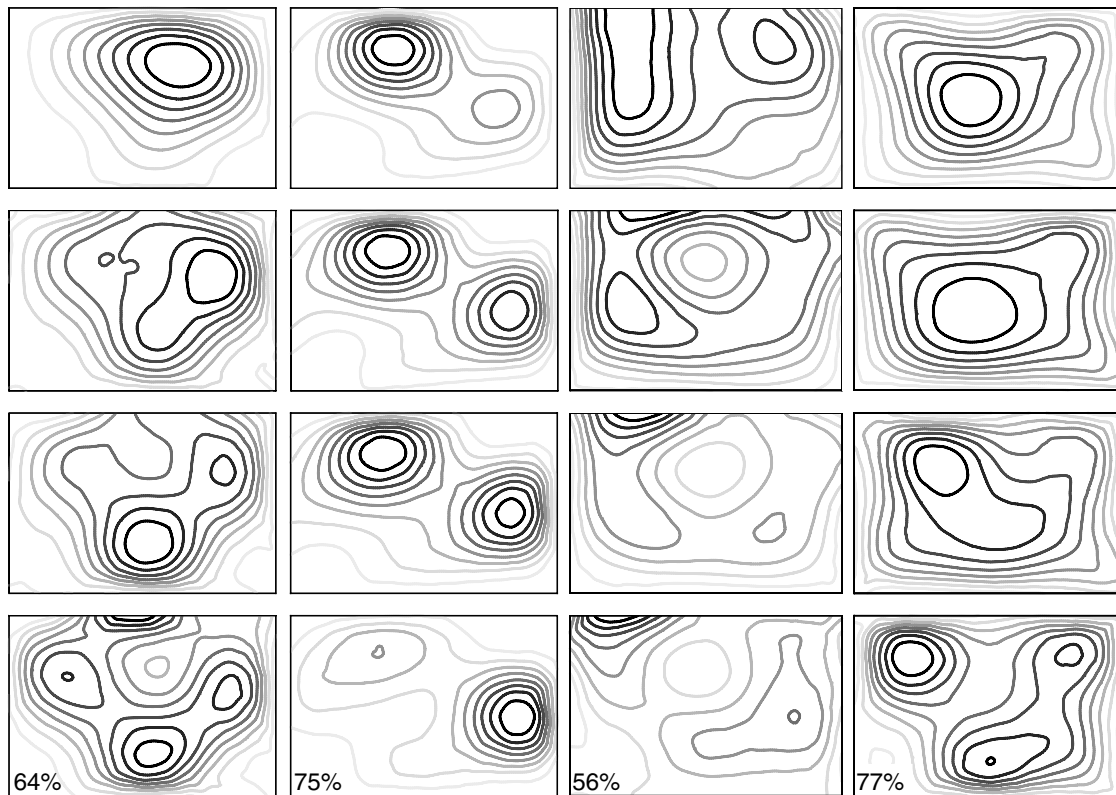

Fig. 4.   Observed spatial distributions of successive fixations for the scenes shown in Fig. 1. The first, second, third, and fourth rows show the smoothed first, second, third, and fourth fixation distributions. The values of the coefficients of determination R^2 averaged over all six possible pairings of successive distributions within a scene (i.e., $3 \times 2 \times 1$) are shown at the bottom left in the bottom row. Other details as for Fig. 1.
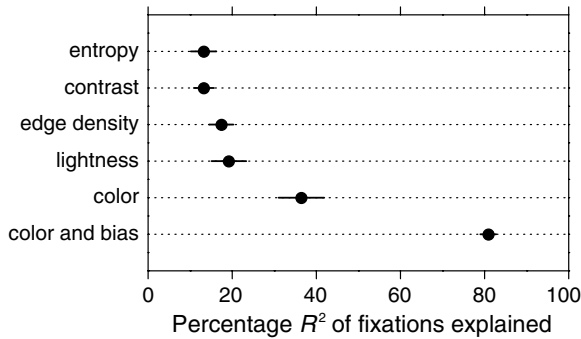
Fig. 5.   Percentage R^2 of variance in fixations explained by different scene properties. Mean values of R^2 are shown for fits to the fixation distributions in each scene of local lightness entropy, contrast, and edge density, local lightness, and color and color combined with viewing bias (for definitions, see Section 2.G). Symbols represent values averaged over 20 scenes, and horizontal bars ±1 standard error of the mean. Values of R^2 were adjusted for the different d.f. in each fit.

C. Influence of Lightness Contrast, Edge Density, and Entropy

Figure 5 shows a dot plot of the mean values of R^2 over scenes for fits to the fixation distributions in each scene of the distributions of local lightness J and local lightness contrast, edge density, and entropy (Section 2.G). Mean values of R^2 are also shown for the fits of local color properties and of local color properties combined with viewing bias (Section 3.A). Horizontal bars show ±1 standard error (SE) of the mean.

These local achromatic properties are not independent of each other, and so comparisons are confounded (the estimated SEs of the differences in R^2 values for two local properties were less than the square root of the sums of the corresponding squared SEs). Nevertheless, it is clear from Fig. 5 that the mean values of R^2 for fits of local lightness contrast, edge density, and entropy were not greater than the mean value of R^2 for the fit of local lightness. Although not shown in Fig. 5, similar results were obtained for local color contrast, edge density, and entropy derived from the three local color properties J, a_C, and b_C.

Crucially, the ordering of the effectiveness of the local scene properties in Fig. 5 remained much the same for smaller and larger Gaussian kernel standard deviations of 2.0 and 3.0 deg (Section 2.E).

4. DISCUSSION

Given the importance of color in foraging in animal species [28,29] and in related animal activity [65], it is surprising that this scene attribute has not received more attention in the analysis of gaze behavior in natural scenes. Where it has been analyzed, it has usually been reported as making a modest contribution [7,16]. Importantly, the red–green and yellow–blue chromatic components have often been combined in a single quantity, such as chroma or a more general saliency measure, which may have diminished their explanatory power. It has also not always been possible to distinguish between the effects of local scene properties from top-down behavior driven by the task and individual observers' viewing strategies [7], which may contain both systematic and random effects [19,20]. The choice of a suitable scale or smoothing kernel for defining local properties may also be relevant.

The present work attempted to deal with these issues by averaging fixation positions over trials and observers [47] and comparing the regressions of fixation position and central viewing bias [12,31,64] on local scene color defined by all three variables, lightness and the red–green and yellow–blue chromatic components, with appropriate allowance for the number of explanatory variables in the fits.

The proportion R^2 of the variance in fixations explained by local color properties ranged from 1% to 85%, depending mainly on how well those properties were consistent with observers' viewing bias. When viewing bias was included in the regression, values of R^2 increased, ranging from 62% to 96%. By contrast, local lightness and local lightness contrast, edge density, and entropy all explained variance less well than local color. Although the smoothing kernel was chosen specifically in relation to surface color processing [50], the explanatory advantage for color properties held with different sizes of the kernel. Moreover, the advantage seemed not to be an artifact of fixations being oversmoothed with this kernel: despite successive fixation distributions being closely correlated within scenes, they were not so between scenes [66,67].

Of course inferences about the role of color in natural scenes necessarily depend on the choice of scenes. The collection of 20 natural scenes used in this analysis included the main vegetated and nonvegetated land-cover classes, namely woodland, vegetation (e.g., grasses, ferns, and flowers), cultivated land, and urban residential and commercial buildings. There was evidently enough variety for a wide range of fits to fixation distributions, but the number of scenes may have been too few to test all the relevant interactions between fixation position, viewing bias, and local color properties, at least of the kind implied by the fits shown in Figs. 1 and 2.

There is another potential issue to do with the interpretation of successive fixations. The similarity of their distributions within scenes may have been an artifact of the way performance was pooled over observers. Suppose, for example, that observers fixated positions a, b, and c in a particular scene in an order that varied randomly either from trial to trial or across observers (e.g., in one trial in the order b, a, c; in the next trial in the order c, a, b; and so on). Then the distribution of first fixations pooled over a sufficiently large number of trials or observers would be centered on the positions a, b, and c, and would coincide with the distribution of second fixations, and so on. This behavior would be consistent with a model of gaze shifts based on a random walk guided by the local properties of the scene [20,21]. Repeated iterations of the model would yield similar distributions with the same scene, whether starting from the first, second, or subsequent fixations, and would yield different distributions with different scenes, precisely as observed.

Whether fixations are taken in order or not, it seems that local scene color has a much stronger influence on gaze position than is generally recognized, capturing significant aspects of scene structure on observers' systematic target search behavior. Central viewing bias modifies that behavior, but in a predictable way. Moreover, the explanatory power of the regressions suggests that visual color representations are approximately linearly related to the attributes of a uniform color appearance space. But how local scene color information is represented visually, that is, as local surface

color descriptors or as proto-objects or something else [17,68,69], remains to be established.

ACKNOWLEDGMENTS

REFERENCES

1. J. M. Henderson, "Human gaze control during real-world scene perception," Trends Cogn. Sci. **7**, 498–504 (2003).
2. J. M. Wolfe, M. L.-H. Võ, K. K. Evans, and M. R. Greene, "Visual search in scenes involves selective and nonselective pathways," Trends Cogn. Sci. **15**, 77–84 (2011).
3. M. S. Castelhano, M. L. Mack, and J. M. Henderson, "Viewing task influences eye movement control during active scene perception," J. Vis. **9**(3):6, 1–15 (2009).
4. A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search," Psycholog. Rev. **113**, 766–786 (2006).
5. M. J. Bravo and K. Nakayama, "The role of attention in different visual-search tasks," Percept. Psychophys. **51**, 465–472 (1992).
6. J. M. Henderson, G. L. Malcolm, and C. Schandl, "Searching in the dark: cognitive relevance drives attention in real-world scenes," Psychon. B. Rev. **16**, 850–856 (2009).
7. D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," Vis. Res. **42**, 107–123 (2002).
8. W. Einhäuser, U. Rutishauser, and C. Koch, "Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli," J. Vis. **8**(2):2, 1–19 (2008).
9. G. Krieger, I. Rentschler, G. Hauske, K. Schill, and C. Zetzsche, "Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics," Spatial Vis. **13**, 201–214 (2000).
10. W. Kienzle, F. A. Wichmann, B. Schölkopf, and M. O. Franz, "A nonparametric approach to bottom-up visual saliency," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, eds. (MIT, 2007), pp. 689–696.
11. D. J. Parkhurst and E. Niebur, "Scene content selected by active vision," Spatial Vis. **16**, 125–154 (2003).
12. S. K. Mannan, K. H. Ruddock, and D. S. Wooding, "The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images," Spatial Vis. **10**, 165–188 (1996).
13. P. Reinagel and A. M. Zador, "Natural scene statistics at the centre of gaze," Netw. Comput. Neural Syst. **10**, 341–350 (1999).
14. A. Açık, S. Onat, F. Schumann, W. Einhäuser, and P. König, "Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories," Vis. Res. **49**, 1541–1553 (2009).
15. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell. **20**, 1254–1259 (1998).
16. B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist, "Visual correlates of fixation selection: effects of scale and time," Vis. Res. **45**, 643–659 (2005).
17. A. Borji, D. N. Sihite, and L. Itti, "Objects do not predict fixations better than early saliency: a re-analysis of Einhäuser et al.'s data," J. Vis. **13**(10):18, 1–4 (2013).
18. J. M. Henderson, A. Nuthmann, and S. G. Luke, "Eye movement control during scene viewing: immediate effects of scene luminance on fixation durations," J. Exp. Psychol.-Hum. Percept. Perform. **39**, 318–322 (2013).
19. M. Nyström and K. Holmqvist, "Semantic override of low-level features in image viewing—both initially and overall," J. Eye Movement Res. **2**(2):2, 1–11 (2008).
20. G. Boccignone and M. Ferraro, "Modelling gaze shift as a constrained random walk," Physica A **331**, 207–218 (2004).
21. G. Boccignone and M. Ferraro, "Ecological sampling of gaze shifts," IEEE Trans. Cybern. **44**, 266–279 (2013).
22. B. M. 't Hart, H. C. E. F. Schmidt, I. Klein-Harmeyer, and W. Einhäuser, "Attention in natural scenes: contrast affects rapid visual processing and fixations alike," Phil. Trans. R. Soc. B **368**, 20130067 (2013).
23. R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," Vis. Res. **45**, 2397–2416 (2005).
24. A. Hurlbert, P. H. Chow, and A. Owen, "Colour boosts performance in visual search for natural objects," J. Vis. **12**(9):105, 105 (2012).
25. H.-P. Frey, K. Wirz, V. Willenbockel, T. Betz, C. Schreiber, T. Troscianko, and P. König, "Beyond correlation: do color features influence attention in rainforest?" Front. Hum. Neurosci. **5**:36, 1–13 (2011).
26. T. Jost, N. Ouerhani, R. von Wartburg, R. Müri, and H. Hügli, "Assessing the contribution of color in visual attention," Comput. Vis. Image Underst. **100**, 107–123 (2005).
27. H.-P. Frey, C. Honey, and P. König, "What's color got to do with it? The influence of color on visual attention in different categories," J. Vis. **8**(14):6, 1–17 (2008).
28. A. D. Melin, D. W. Kline, C. M. Hickey, and L. M. Fedigan, "Food search through the eyes of a monkey: a functional substitution approach for assessing the ecology of primate color vision," Vis. Res. **86**, 87–96 (2013).
29. P. W. Lucas, N. J. Dominy, P. Riba-Hernandez, K. E. Stoner, N. Yamashita, E. Loría-Calderón, W. Petersen-Pereira, Y. Rojas-Durán, R. Salas-Pena, S. Solis-Madrigal, D. Osorio, and B. W. Darvell, "Evolution and function of routine trichromatic vision in primates," Evolution **57**, 2636–2643 (2003).
30. K. Amano, D. H. Foster, M. S. Mould, and J. P. Oakley, "Visual search in natural scenes explained by local color properties," J. Opt. Soc. Am. A **29**, A194–A199 (2012).
31. B. W. Tatler, "The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions," J. Vis. **7**(14):4, 1–17 (2007).
32. F. L. Engel, "Visual conspicuity, directed attention and retinal locus," Vis. Res. **11**, 563–575 (1971).
33. J. Wolfe, P. O'Neill, and S. Bennett, "Why are there eccentricity effects in visual search? Visual and attentional hypotheses," Percept. Psychophys. **60**, 140–156 (1998).
34. M. Carrasco, D. L. Evert, I. Chang, and S. M. Katz, "The eccentricity effect: target eccentricity affects performance on conjunction searches," Percept. Psychophys. **57**, 1241–1261 (1995).
35. CIE, Technical Committee 8-01, "A colour appearance model for colour management systems: CIECAM02," CIE 159:2004 (Commission Internationale de l'Eclairage, Vienna, Austria, 2004).
36. D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," J. Opt. Soc. Am. A **23**, 2359–2372 (2006).
37. J. M. Wolfe, N. Klempen, and K. Dahlen, "Postattentive vision," J. Exp. Psychol. Hum. Percept. Perform. **26**, 693–716 (2000).
38. G. Harding and M. Bloj, "Real and predicted influence of image manipulations on eye movements during scene recognition," J. Vis. **10**(2):8, 1–17 (2010).
39. K. Kaspar and P. König, "Viewing behavior and the impact of low-level image properties across repeated presentations of complex scenes," J. Vis. **11**(13):26, 1–29 (2011).
40. D. Noton and L. Stark, "Scanpaths in saccadic eye movements while viewing and recognizing patterns," Vis. Res. **11**, 929–942 (1971).
41. M. S. Mould, D. H. Foster, K. Amano, and J. P. Oakley, "A simple nonparametric method for classifying eye fixation," Vis. Res. **57**, 18–25 (2012).
42. I. van der Linde, U. Rajashekar, A. C. Bovik, and L. K. Cormack, "DOVES: a database of visual eye movements," Spatial Vis. **22**, 161–177 (2009).

43. W. Kienzle, M. O. Franz, B. Schölkopf, and F. A. Wichmann, "Center–surround patterns emerge as optimal predictors for human saccade targets," J. Vis. **9**(5):7, 1–15 (2009).

44. D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, Palm Beach Gardens, Florida (ACM, 2000), pp. 71–78.

45. M. Nyström and K. Holmqvist, "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data," Behav. Res. Methods **42**, 188–204 (2010).

46. E. Vig, M. Dorr, and E. Barth, "Efficient visual coding and the predictability of eye movements on natural movies," Spatial Vis. **22**, 397–408 (2009).

47. N. Wilming, T. Betz, T. C. Kietzmann, and P. König, "Measures and limits of models of fixation selection," PLoS ONE **6**(9), e24038 (2011).

48. B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability (Chapman & Hall/CRC Press, 1986).

49. J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, Monographs on Statistics and Applied Probability (Chapman & Hall/CRC Press, 1996).

50. D. H. Foster, S. M. C. Nascimento, K. Amano, L. Arend, K. J. Linnell, J. L. Nieves, S. Plet, and J. S. Foster, "Parallel detection of violations of color constancy," Proc. Natl. Acad. Sci. USA **98**, 8151–8156 (2001).

51. J. Najemnik and W. S. Geisler, "Simple summation rule for optimal fixation selection in visual search," Vis. Res. **49**, 1286–1294 (2009).

52. R. Rosenholtz, Y. Li, and L. Nakano, "Measuring visual clutter," J. Vis. **7**(2):17, 1–22 (2007).

53. N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: an information theoretic approach," J. Vis. **9**(3):5, 1–24 (2009).

54. C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: comparison with eye fixations," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 970–982 (2000).

55. L. F. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," Prob. Peredachi Inf. **23**(2), 9–16 (1987).

56. M. N. Goria, N. N. Leonenko, V. V. Mergel, and P. L. Novi Inverardi, "A new class of random vector entropy estimators and its applications in testing statistical hypotheses," J. Nonparametr. Stat. **17**, 277–297 (2005).

57. M. S. Mould, "Visual search in natural scenes with and without guidance of fixations," Ph.D. thesis (University of Manchester, Manchester, UK, 2011).

58. N. R. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed. (Wiley, 1998).

59. T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*, Monographs on Statistics and Applied Probability (Chapman & Hall/CRC Press, 1990).

60. R. Baddeley, "Searching for filters with 'interesting' output distributions: an uninteresting direction to explore?" Netw. Comput. Neural Syst. **7**, 409–421 (1996).

61. B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability (Chapman & Hall/CRC Press, 1993).

62. M. Cerf, E. P. Frady, and C. Koch, "Faces and text attract gaze independent of the task: experimental data and computer model," J. Vis. **9**(12):10, 1–15 (2009).

63. Q. Zhao and C. Koch, "Learning a saliency map using fixated locations in natural scenes," J. Vis. **11**(3):9, 1–15 (2011).

64. T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," in Computer Science and Artificial Intelligence Laboratory Technical Report MIT-CSAIL-TR-2012-001 (MIT, 2012).

65. T. Troscianko, C. P. Benton, P. G. Lovell, D. J. Tolhurst, and Z. Pizlo, "Camouflage and visual perception," Phil. Trans. R. Soc. B **364**, 449–461 (2009).

66. C. Kayser, K. J. Nielsen, and N. K. Logothetis, "Fixations in natural scenes: interaction of image structure and image content," Vis. Res. **46**, 2535–2545 (2006).

67. J. M. Henderson, P. A. Weeks, Jr., and A. Hollingworth, "The effects of semantic consistency on eye movements during complex scene viewing," J. Exp. Psychol. Hum. Percept. Perform. **25**, 210–228 (1999).

68. W. Einhäuser, M. Spain, and P. Perona, "Objects predict fixations better than early saliency," J. Vis. **8**(14):18, 1–26 (2008).

69. A. F. Russell, S. Mihalaş, R. von der Heydt, E. Niebur, and R. Etienne-Cummings, "A model of proto-object based saliency," Vis. Res. **94**, 1–15 (2014).