Quantifying center bias of observers in free viewing of dynamic natural scenes

Po-He Tseng Department of Computer Science, University of Southern California, Los Angeles, CA, USA

Ran Carmi Neuroscience Program, University of Southern California, Los Angeles, CA, USA

Ian G. M. Cameron Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada

Douglas P. Munoz Centre for Neuroscience Studies, Departments of Physiology, Psychology, and Medicine, Queen's University, Kingston, Ontario, Canada

Laurent Itti Neuroscience Program, Department of Computer Science, University of Southern California, Los Angeles, CA, USA

Human eye-tracking studies have shown that gaze fixations are biased toward the center of natural scene stimuli ("center bias"). This bias contaminates the evaluation of computational models of attention and oculomotor behavior. Here we recorded eye movements from 17 participants watching 40 MTV-style video clips (with abrupt scene changes every 2–4 s), to quantify the relative contributions of five causes of center bias: photographer bias, motor bias, viewing strategy, orbital reserve, and screen center. Photographer bias was evaluated by five naive human raters and correlated with eye movements. The frequently changing scenes in MTV-style videos allowed us to assess how motor bias and viewing strategy affected center bias across time. In an additional experiment with 5 participants, videos were displayed at different locations within a large screen to investigate the influences of orbital reserve and screen center. Our results demonstrate quantitatively for the first time that center bias is correlated strongly with photographer bias and is influenced by viewing strategy at scene onset, while orbital reserve, screen center, and motor bias contribute minimally. We discuss methods to account for these influences to better assess computational models of visual attention and gaze using natural scene stimuli.

Keywords: eye movement, saccade, orbital reserve, eye–head coordination, saccadic eye movement, saccade selection, fixation, eye position, ocular, visuo-motor optimizing strategy, salience, saliency maps, photographer bias, motor bias, screen center

Citation: Tseng, P., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision, 9*(7):4, 1–16, http://journalofvision.org/9/7/4/, doi:10.1167/9.7.4.

Introduction

Several studies of attentional selection in natural scenes have observed that participants' visual attention, measured by saccade direction and fixation locations, is biased toward the center of static images (Busswell, 1935; Foulsham & Underwood, 2008; Mannan, Ruddock, & Wooding, 1995, 1996, 1997; Parkhurst, Law, & Niebur, 2002; Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Tatler, 2007; Tatler, Baddeley, & Gilchrist, 2005), as well as to the center of videos (Itti, 2004). This observation has been of importance because it is unclear whether the bias is driven by the content of the images/videos or by other factors (Foulsham & Underwood, 2008; Parkhurst et al., 2002; Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Tatler et al., 2005). One possible cause of the bias is intrinsic bottom-up visual salience, as computed by saliency map models (Itti & Koch, 2000; Koch & Ullman, 1985), which is a significant predictor of where observers look in arbitrary natural scenes (Foulsham & Underwood, 2008; Itti, 2004; Parkhurst et al., 2002; Renninger, Coughlan, Verghese, & Malik, 2005; Tatler et al., 2005; Tatler, Baddeley, & Vincent, 2006; Underwood & Foulsham, 2006). However, because in natural images/videos, the distribution of subjects of interest and salience is usually biased toward the center (Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Tatler, 2007; Tatler et al., 2005), it is unclear how much the salience actually contributes in guiding attention. It is possible that people look at the center for reasons other than salience, but their gaze happens to fall on salient locations. Therefore, this

doi: 10.1167/9.7.4 Received September 6, 2008; published July 9, 2009 ISSN 1534-7362

center bias may result in overestimating the influence of salience computed by the model and contaminate the evaluation of how visual salience may guide orienting behavior. Hence, the goal of this study is to quantify the relative contributions of several suspected causes of center bias in dynamic natural scenes. In addition, we propose a method to adjust for the bias to evaluate the correlation between salience and gaze.

One of the most interesting causes of center bias is known as photographer bias (Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Schumann et al., 2008; Tatler, 2007; Tatler et al., 2005). Photographer bias is a natural tendency of photographers to place objects or actors of interest (top down) near the center of their composition and to enhance their focus and size relative to the background. In fact, what the photographer considers interesting may also be highly salient (bottom up). Schumann et al. (2008) utilized a wearable device to simultaneously record gaze-centered and head-centered videos while people were freely exploring natural environments. They demonstrated that local feature distribution of these head-centered videos contained a spatial bias, and this spatial bias was centered in the gaze-centered videos. This study suggests that photographer bias exists in guiding behavior in natural environments. Therefore, photographer bias may explain why viewers are then attracted to the center of these particular displays.

Photographer bias can lead to another bias—known as viewing strategy (Parkhurst et al., 2002)—whereby viewers may reorient at a greater frequency to the center of a scene relative to other locations, if they expect highly salient or interesting objects to be placed there. This may result in an initial bias to the center upon a novel scene but will then change as the content of the scene becomes familiar. The strategy is not tied to a particular stimulus or photographer. Instead, it is a developed strategy when people are exposed to photographer-biased stimuli repeatedly. Moreover, other causes can also contribute to this strategy. For example, upon encountering a new scene, looking at the center can maximally utilize attentional resources if they are hemifield independent (Alvarez & Cavanagh, 2005). Hence, compared to other locations, looking at the center allows viewers to acquire more information about the scene compared to other locations. We specifically investigated the influence of photographer bias with video stimuli whose saliency maps (color, intensity, orientation, flicker, and motion) and subjects of interest (rated by humans) were center-biased to different degrees. We also examined the influence of viewing strategy by assessing the gaze distribution of participants as each new scene progressed in time.

Aside from photographer bias and viewing strategy, three other potential causes of center bias have been identified: orbital reserve (Carmi & Itti, 2006; Fuller, 1996; Parkhurst & Niebur, 2003), motor bias (Foulsham & Underwood, 2008; Tatler et al., 2005), and center of screen bias (Vitu, Kapoula, Lancelin, & Lavigne, 2004). Regarding orbital reserve, previous studies have demonstrated that initial orbital positions influence saccadic latency and saccade amplitude (Van Opstal, Hepp, Suzuki, & Henn, 1995). These studies implied that a bias exists for preferentially initiating eye movements toward the central orbital position (looking straight ahead) as opposed to away from it (Fuller, 1996; Paré & Munoz, 2001). This recentering bias is initiated immediately when the eyeballs leave central orbital position, and it prevents reaching the limit of the ocular muscles and therefore facilitates the flexibility for selecting the next target (Tweed, 1997). Motor bias is the tendency for participants to make short saccades rather than long saccades (Tatler, 2007). Given that most natural scene free viewing experiments asked participants to start viewing from a central marker, gaze distribution would be clustered at the starting marker. However, Vitu et al. (2004) showed that the center of the screen (based on the boundaries of the display), rather than straight-ahead position of the eyes (orbital reserve), biased the saccade landing position. To investigate these three contributions to center bias, we first compared the gaze paths to those produced by simulation to assess motor bias (Tatler, 2007). Then we examined the influence of orbital reserve and center of screen by displaying videos at different locations on a large screen, while keeping the observers' heads fixed straightforward to the center of the large screen.

Recent efforts have been made in order to understand the roles of these factors in causing center bias in natural images. Tatler (2007) concluded that neither motor bias nor low-level features of saliency (brightness, chromaticity, contrast, and edge content) could fully account for the center-biased fixation distribution while participants freely viewed natural images. However, once participants were given a task to search for targets defined by basic features (e.g., luminance properties) the fixation distributions correlated with the feature distribution. The correlation was more obvious during early viewing, but after an initial centering response. This study showed that center bias has greater influence on free viewing than searching of images.

Here we extend the investigation of center bias to videos and explore its causes under the context of dynamic salience. Our findings demonstrate that center bias is mainly driven by photographer bias and viewing strategy. Moreover, the top-down component of photographer bias has greater influence than the bottom-up component. Orbital reserve and center of screen contribute to center bias one order of magnitude less than photographer bias, and no effect was found on motor bias, consistent with Tatler (2007). We propose that within the limit of comfortable visual exploration (shown by the small influence of orbital reserve or screen center), people are biased to central locations based on their expectations and possibly by the actual placement of objects of interest or of high salience by the photographer.

Experiment 1

Experimental procedures were approved by the Human Research and Ethics Board at Queen's University and adhere to the guidelines of the Declaration of Helsinki.

Stimuli

Forty 30-s continuous videos were filmed (Sony HandyCam DCR-HC211 NTSC) at the USC campus, a beach, Hollywood, a shopping district, a ski resort, a desert, and recorded from television and video games. Cognitively interesting objects were deliberately placed in different locations of the videos while filming. For example, for some videos, the camera was set immobile on a tripod, or panned at a constant speed (6°/s ranging 120° back and forth horizontally), to film people walking, vehicles moving, etc., so that possibly interesting events would occur at many locations in the videos; for other videos, the camera was also set to follow particular people and to put them at the center of the videos so that interesting events would happen mostly near the center of the videos.

These videos (640 × 480 pixels, MPEG-1) were cut randomly into clip snippets ("clippets") in different ways. The length of the segments for 30 out of 40 videos was uniformly distributed from 2 to 4 s, yielding a total of 291 clippets. These clippets were scrambled and reassembled randomly into 30 "MTV-style clips" (common clip set) under the constraints that no more than one clippet from the same original video could exist in one MTV-style clip, and the length of each MTV-style clip should be approximately 30 s (Carmi & Itti, 2006). The 30-s clips were thus made up of 9 to 11 clippets that run continuously without any temporal gaps in between. Using different clippet length avoided the expectation in timing of the onset of the next clippet (example). The last 10 of our 40 original videos were segmented and reassembled into three "specific clip sets" differing in clippet length. For the first specific clip set, the 10 videos were segmented into clippets whose length was uniformly distributed from 2 to 4 s (93 clippets), like the first 30 videos, and reassembled to 10 MTV-style clips; the second and the third specific clip sets were processed in the same way with a different clippet length, 1–3 s for the second set (139 clippets), and 0.5–2.5 s for the third set (200 clippets).

There were three advantages of using MTV-style stimuli to investigate the causes of center bias. First, participants might fixate anywhere at the beginning of a new scene, rather than being engaged in a typical central cross-fixation task until the clip started. Second, although the clippets were short, they were long enough for participants to explore and understand the scene; hence, using short clippets allowed experimenters to collect more data for learning the evolution of center bias over time within the same total video length. Third, shorter clippets were easier to classify in terms of the degree of photographer bias compared to longer clips where that degree may change over time, thereby providing better control of the stimuli.

Participants

Seventeen young adults (8 males and 9 females, range 20–29 years, mean 23.2 years) with normal or corrected-to-normal vision were recruited for the experiment. They were compensated and were naive to the purpose of the experiment.

Data acquisition

Stimuli were displayed on an 18-inch color monitor (363 × 271 mm), 61 cm in front of the participant (corresponding to a 35.14° × 25.88° field of view). Participants' heads were stabilized with a chin rest, and participants were instructed to "watch and enjoy the clips".

Clips from the common set and from one of the specific sets were played in random order. Participants were able to rest every ten clips (approximately after 5 min, which was the length of one session). A nine-point calibration was performed at the beginning of every session. At the beginning of every clip, participants were required to fixate a cross at the center of the screen; however, participants could look anywhere on the screen at the beginning of individual clippets.

Instantaneous eye position was tracked by a head-mounted EyeLink II camera (SR Research) in Pupil-CR mode (250 Hz, noise < 0.022°, gaze position accuracy < 0.5° average, and gaze tracking ranges of ±20° horizontal and ±18° vertical) from participants' right eye. Gaze position was shown on the experimenter's screen to monitor the participant's status and the quality of the data. Six hundred and eighty data sets (17 participants × 40 clips) of eye-movement traces were obtained. Data were removed from further analysis if they contained excessive (more than 10%) loss of tracking (39 eye-movement traces). Eye traces belonging to *specific clip sets* (170 eye-movement traces, 17 participants × 10 clips) were not analyzed because they had different clippet length and these clippets were used primarily to explore other aspects of eye movements unrelated to this study. Likewise, pan-style clips were not included in this study because optokinetic nystagmus (eye movements tended to follow global motion of the scene) induced by pan-style clips could contaminate the measurement of influence of photographer bias. When observers looked at locations of panning direction, it was unclear whether it was caused by

optokinetic nystagmus or photographer bias. The remaining eye-movement traces were further analyzed and classified as fixation, saccade, blink/artifact, saccade during blink, smooth pursuit, and eye-tracker drift/misclassification. Blinks were identified as whenever pupil diameter (recorded by the eye tracker) was zero. Gaze position outside of 1° inside of the border of the screen was labeled as an artifact. Eye movements whose minimum velocity was 30°/s and minimum amplitude was 2° were labeled as saccades. The first saccade after clip onset was removed because it started from the central fixation point. In the end, 17,136 saccades, obtained from 9,302.56 s of eye-movement recording (195 clippets and 3,022 participant–clippet pairs), were used for detailed analysis.

Data analysis and results

Photographer bias

We quantified photographer bias by a top-down center bias score (TD score; how centered were subjectively "interesting" elements of the scenes, in which an element may be of interest to one person but not to another) and a bottom-up center bias score (BU score; how centered were visually salient elements of the scene), which will be defined in the following sections. The degree of center bias in saccade endpoint distributions was characterized by a human saccade center bias score (HS score). The null hypothesis was that if photographer bias (stimuli) has nothing to do with saccade-endpoint center bias (behavior), then photographer bias (TD and BU scores) should be nonsignificant predictors to the saccade endpoint center bias (HS score); otherwise, they should be significant predictors. The scores are defined as follows.

BU score represents how center-biased the "saliency maps" of the clippets were. Saliency maps of each clippet were computed by the saliency model (Itti & Koch, 2000). Five channels, color, intensity contrast, orientation, flicker, and motion, were used to compute saliency maps of each clippet. All the saliency maps of each clippet's video frames were summed up to give the overall saliency map of the clippet, and the map was normalized by dividing its saliency values by the sum of all saliency values in the map so as to convert saliency value to probability. The BU score was then calculated as the sum of saliency values weighted by the Euclidean distance to the center. After BU scores of all the clippets were computed, the range of the scores was normalized from zero (least center-biased clippet in terms of saliency map) to one (most center-biased clippet in terms of saliency map) for further analysis.

TD score reflects how center-biased the cognitively interesting things (events, objects, etc.) were in each 2- to 4-s-long clippets. To assess this, five naive participants (excluded from the following eye-tracking experiments) were recruited to provide subjective scores from 1 to 5 in terms of how interesting things were biased toward center for every clippets. They were instructed to "Please give a score from 1, 2, 3, 4 and 5. If all the interesting things attractive to you are happening at the center of the screen, please press 1; if they are half in the center and half are around the border, please press 3; if everything interesting is happening around borders, please press 5". The score was given at the end of each short clippet, whose duration would likely not overwhelm working memory resources, and the scene did not change abruptly within each clippet. Hence, this offline estimation should not be too far from the online process. Nevertheless, some contamination might exist because raters did not always agree on their rating. We accounted for this by repeating analysis for clippets with high or low rater disagreement. The score was later normalized from zero (all interesting things happening around the borders) to one (all interesting things happening at the center) for further analysis.

TD* score is similar to TD score but independent of BU score. When raters gave TD score to each clippet, it is likely that they were also attracted by salient locations, which may have biased their scoring. Therefore, the TD score was likely contaminated by the BU score. Hierarchical regression allowed us to remove the dependency. When TD score was regressed on BU score, the residual was the part of TD score that cannot be explained by BU score. Hence, the residual was denoted as TD* score.

Examples of stimuli are shown in Figure 1. Figure 1A is highly photographer-biased because cognitively interesting things happened at the center (high TD score) and the center was also salient (high BU score); for Figure 1B, the cognitively interesting things are happening at the center (high TD score): woman carrying a baby walking along the path, but it is not salient at the center (low BU score); for Figure 1C, there was nothing interesting at the center (low TD score), but the center was salient (high BU score); for Figure 1D, there was neither cognitively interesting things at the center (low TD score) nor was the center a salient location(low BU score).

HS score reflects how center-biased the saccade endpoint distribution was. It was the average distance of all the saccade endpoints to the center of a display, and then normalized from zero (average distance to center from a uniform saccade endpoint distribution) to 100 (all saccade endpoints fall exactly at the center, whose distance to center is zero). Standard error was estimated by bootstrapping 1000 runs. HS score was widely used in this study in two ways: either computed from a single clippet or all clippets depending on the analysis.

An illustration of the influence of photographer bias is shown in Figure 2. Saccade endpoint distributions are plotted for clippets with four extreme photographer bias conditions, strong (TD*, BU), strong TD*, strong BU, and weak (TD*, BU). Clippets that rank in the top 25% of each condition are selected. For example, the strong (TD, BU) data plotted in Figure 2 are the saccade endpoint distributions while participants watched clippets whose TD* score and BU score both rank in the top 25% among all of the clippets. For the strong TD* condition in Figure 2,

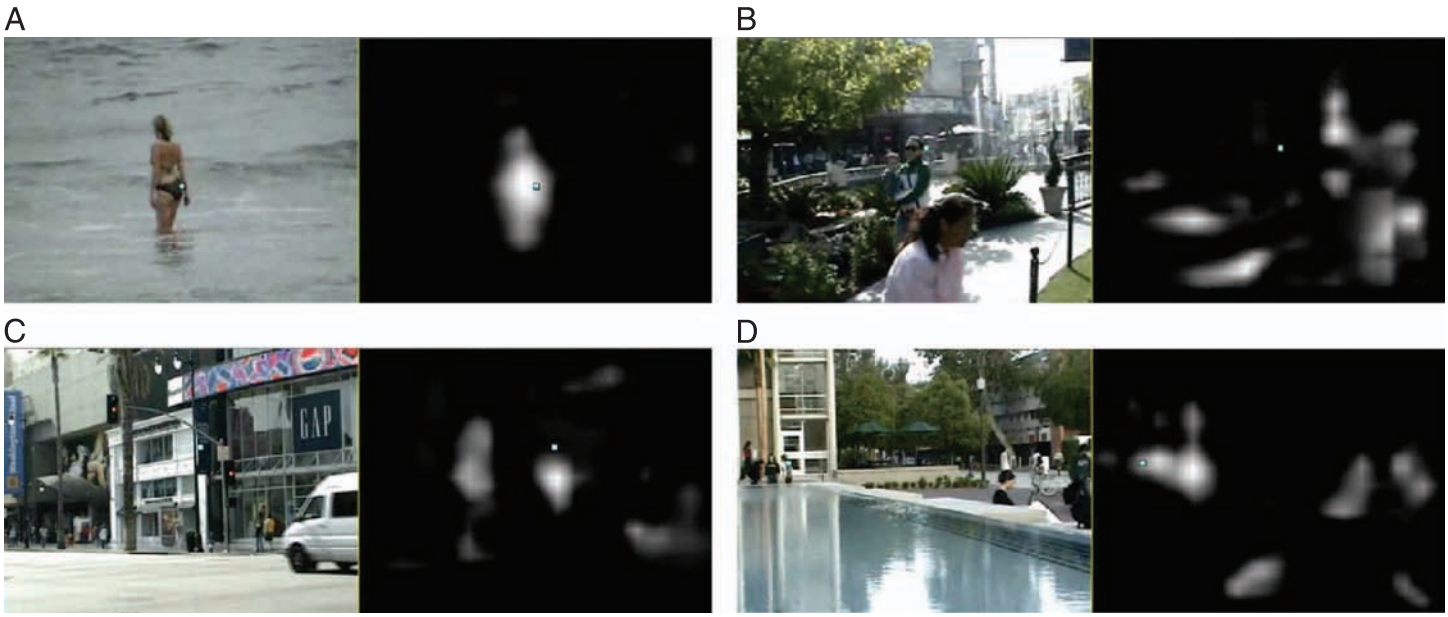

Figure 1. Examples of video stimuli of four conditions. In each panel, the left figure is a frame of videos presented to participants, and the right figure is the saliency map of that frame. The small cyan square represents the current eye position of one observer. (A) High TD and BU scores (strong (TD, BU)). (B) High TD score but low BU score (strong TD). (C) Low TD score but high BU score (strong BU). (D) Low TD and BU scores (weak (TD, BU)).

the selected clippets rank in the top 25% in TD score but lower 25% in BU score. There are 12 clippets with 989 saccades for strong (TD*, BU) condition, 11 clippets with 990 saccades for strong TD* condition, 8 clippets with 625 saccades for strong BU condition, and 14 clippets with 1430 saccades for weak (TD*, BU) condition. The impact of photographer bias on center bias is revealed by HS score. The HS score of the 4 photographer bias conditions was not the same (ANOVA, $F(3, 4028) = 156.62$, $p < 0.05$). The scores increase from weak (TD*, BU), strong BU, strong TD*, to strong (TD*, BU), but the difference between strong TD* and strong BU was not significant (post hoc ANOVA, $p = 0.20$).

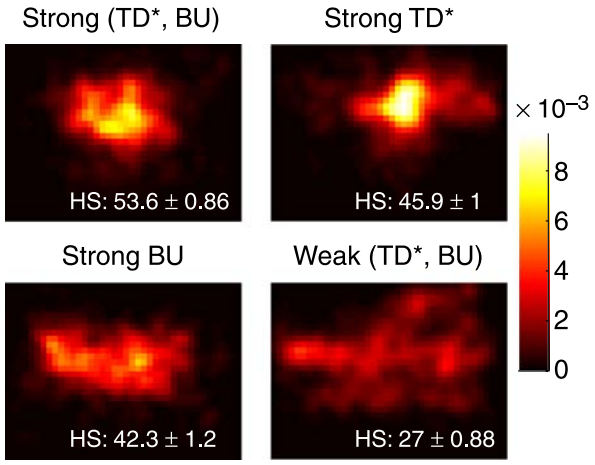

Figure 2. Saccade endpoint distributions and the HS score (degree of center bias given above each panel) of four different photographer bias conditions. TD*, unlike TD score given by raters, is independent from BU by decorrelating with BU score.

To formally quantify the correlation between photographer bias and center bias, we performed linear regressions involving TD, BU, and HS scores in the following steps. First, we computed the coefficient of determination (R^2) between BU scores and HS scores to determine the effect of salient locations on gaze behavior. Second, we regressed TD score on BU score to obtain the residual (TD* score), the fraction of the TD score that cannot be explained by the BU score. Therefore, the BU score and TD* score are independent. Third, we computed the R^2 value between the TD* score and the HS score to learn the effect of the spatial distribution of subjects of interest on gaze behavior. Lastly, to calculate the effect of a combined TD and BU on center bias, we computed the R^2 value of the combination of BU and TD* scores on HS score.

Raters giving TD score did not always find the same subjects/objects cognitively interesting. To evaluate their disagreement and to accommodate the small number of raters, we examined how different the TD scores were given by raters (Table 1). Rater disagreement of each clippet was calculated as the absolute difference of the TD score between every pair of raters. Inter-rater reliability was evaluated by computing the average Pearson's correlation between the TD score given by every pair of raters. When raters agreed with each other (lower half), the inter-rater reliability was high ($r = 0.71$). On the contrary, when they disagreed with each other, the inter-rater reliability was low ($r = 0.03$).

Average rater disagreement	Number of clippets	Inter-rater reliability (correlation)
Low (0–1.2)	104	0.71
High (1.4–2.2)	91	0.03
All (0–2.2)	195	0.39

Table 1. Inter-rater reliability as a function of average rater disagreement given by 5 raters. Low and High represent the lower and higher half of the clippets sorted by the average rater disagreement. Numbers in brackets are the range of the average rater disagreement.

Figure 3 shows the amount of center bias explainable by photographer bias. The average HS score of each clippet was regressed by TD*, BU, or both, and the corresponding R^2 was reported. Here we found that TD* and BU were both significant predictors of the degree of center bias regardless of rater disagreements (Figure 3B; F-test, all $p < 0.05$). To further investigate the influence of rater disagreement on the TD* regressor, the two regression models of low and high rater disagreement were compared and their difference was marginally significant ($F(2, 191) = 2.72$, $p = 0.068$). Nevertheless, about 31% of the variance ($0.29 < R^2 < 0.33$) in center bias could be attributed to TD*. For regression models with only the BU regressor, average rater disagreement had no influence on them ($F(2, 191) = 1.40$, $p = 0.25$). To evaluate the overall impact of photographer bias, TD* and BU were both included as regressors, and about 47% of the variance in center bias could be attributed to photographer bias ($0.40 < R^2 < 0.52$).

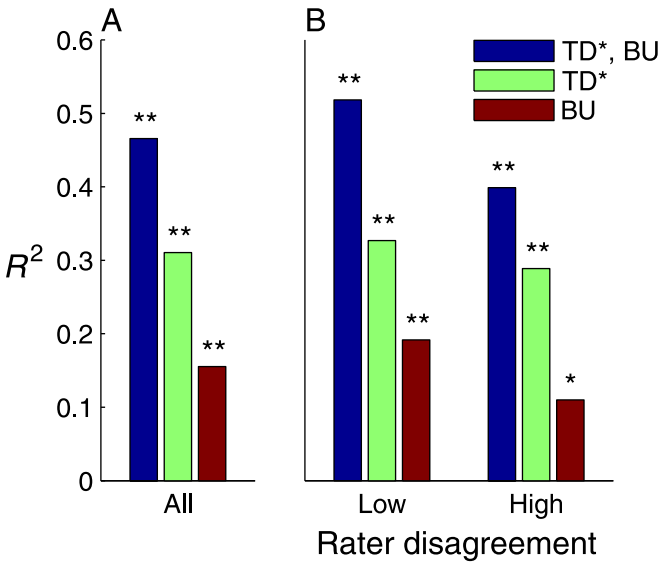

Figure 3. The proportion of variability (R^2) in HS score that could be attributed to (TD*, BU), TD*, or BU scores for (A) all clippets and (B) clippets with different rater disagreements. (*$p < 0.05$, **$p < 0.001$).

Motor bias

Another presumed cause of center bias is motor bias due to the fact that humans prefer to make shorter saccades and horizontal saccades (Tatler, 2007). Therefore, the saccade sequence could be clustered around the starting point. Given that most natural scene viewing experiments gave observers a starting marker at the center of the screen, center bias could result from the motor bias. One way to test the hypothesis is by simulating saccade sequence with a random walk model, given that each step exhibits the same motor bias as humans (Tatler, 2007).

In our simulation, the random walk model's starting positions of each clippet were given the same starting position recorded while observers initiated their first saccade in each clippet. The number of saccades in each clippet was also matched to that of human observers. The amplitude and direction of each simulated saccade were randomly sampled from those of human observer to mimic motor bias. The overall saccade endpoint distribution of human saccades and the distribution simulated by random walk model are shown in Figure 4A. The simulated distribution was more uniform over space than human saccade endpoints (two-sample t-test, $p < 0.05$), which suggested that motor bias is not likely a crucial factor.

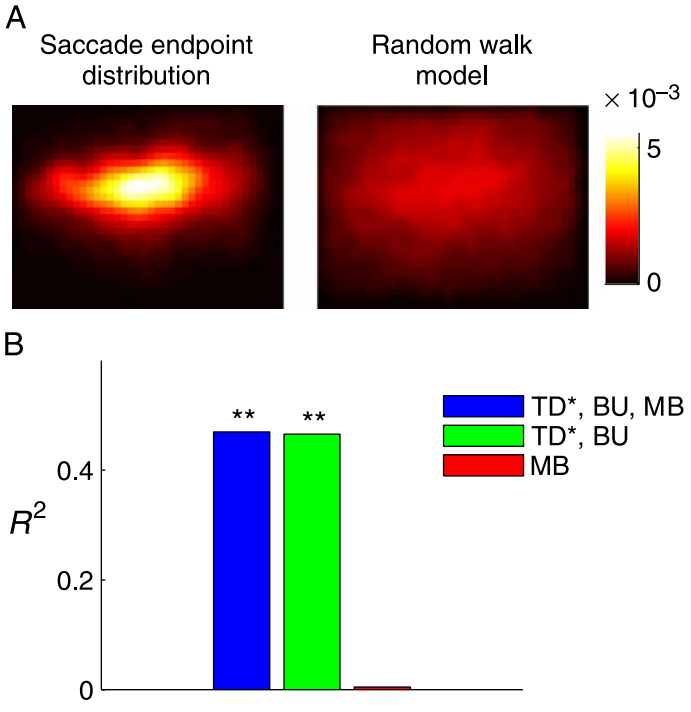

Figure 4. (A) Saccade endpoint distribution and the simulated distribution generated by random walk model provided the same starting positions and number of saccades of each clippets. Saccade amplitude and direction of the simulated saccade sequence are randomly sampled from that of human observers. (B) Contribution of photographer bias (TD*, BU) and motor bias (MB) to center bias.

To further quantify the influence of motor bias on center bias, a motor-bias score (MB score) was calculated from the simulated saccade endpoint distribution of each clippet. MB score was calculated exactly the same way as HS score. However, they were different in that HS scores were derived from the observed saccade endpoint distribution while MB scores were derived from the simulated saccade endpoint distribution. Next, the contribution (R^2) of motor bias on center bias was obtained by regressing HS score on MB score (Figure 4B). The contribution of motor bias was confirmed to be very small ($R^2 = 0.0046$) and nonsignificant ($F(1,193) = 0.9$, $p = 0.34$). Therefore, motor bias is not a contributor to center bias and agrees with previous study (Tatler, 2007).

Viewing strategy

An initial centering response upon a new scene was observed in every participant and revealed in Figure 5. The figure showed HS score across participants as a function of saccade number from the start of the clippets. If observers explored a scene equally in space as the clippet progressed, then HS score should have remained the same throughout saccade sequence. However, a one-way ANOVA showed an effect of saccade sequence on center bias ($F(4, 12899) = 162.85$, $p < 0.05$). A post hoc paired *t*-test with Bonferroni correction showed that the HS scores of the first and second saccades were higher than that of the following saccades ($p < 0.05$). However, the HS scores from the third to fifth saccades were not significantly different ($p > 0.11$). These results suggested that people tended to look closest to the center at the beginning of clippets (higher HS score) and then explored the scene more uniformly in subsequent saccades (lower HS score), consistent with previous findings (Tatler, 2007).

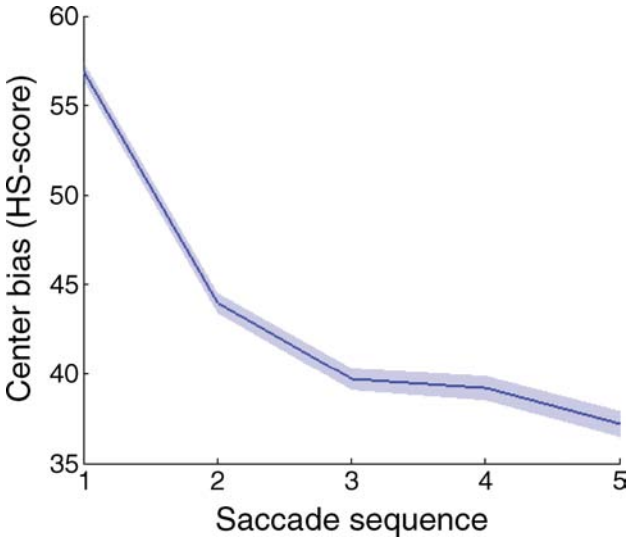

Figure 5. An initial centering response upon a new scene, indicated by the HS score across saccade sequence. The light region is standard error.

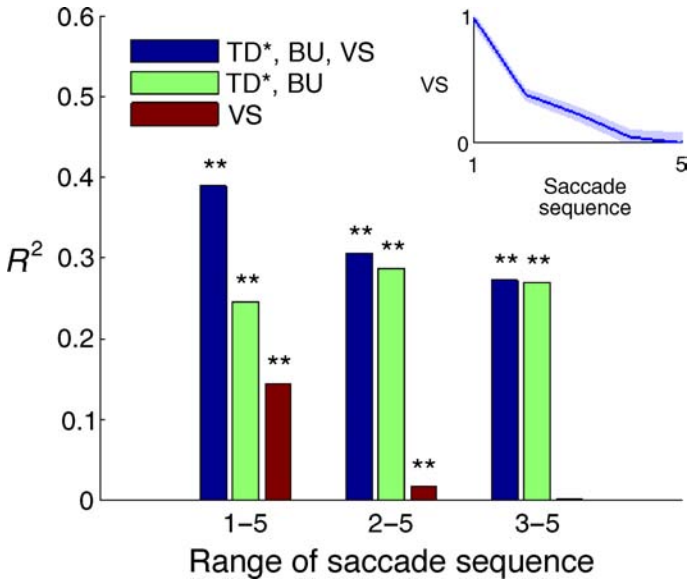

Figure 6. The influence of viewing strategy (VS) compared to photographer bias (TD*, BU). Upper right subplot is the VS estimated from the other clippet set (specific set).

This initial centering response and the decreasing trend revealed a viewing strategy in exploring a scene. Here we attempted to roughly estimate the influence of the strategy. The viewing strategy was estimated from another set of eye-movement data to show its generality. The data were obtained while participants were watching videos belonging to the specific sets that were separated from the main analysis (7741 saccades from 5,101.54 s of eye-movement recording). The trend of viewing strategy (VS) was first computed as HS scores of the first 5 saccades across participants and clippets, then was normalized from 0 to 1 by their minimum and maximum HS scores (upper right corner of Figure 6). By pairing the VS score of the first 5 saccades to each clippet from the common set, VS score was not correlated with TD* and BU scores ($F(2, 952) = 0.12$, $p = 0.88$, $R^2 = 0.0003$). However, VS was found to be a significant predictor to HS score (center bias; $F(1, 953) = 162.32$, $p < 0.05$, $R^2 = 0.15$). To compare the influence of VS to photographer bias under the same condition, the HS scores of the first 5 saccades were regressed on the TD* and BU scores of each clippet. Photographer bias was still a significant predictor ($F(2, 952) = 156.00$, $p < 0.05$, $R^2 = 0.25$). However, the influence of viewing strategy diminished quickly; if the first saccade of viewing strategy was dropped, viewing strategy became a much weaker but was still a significant predictor ($F(1, 778) = 14.42$, $p < 0.05$, $R^2 = 0.02$). Moreover, if the first two saccades were dropped, viewing strategy was no longer a significant predictor ($F(1, 583) = 1.39$, $p = 0.24$). These results imply that (1) the influence of viewing strategy was comparable to that of photographer bias initially, however (2) viewing strategy was no longer influential after the third saccade whereas photographer bias continued to be so. These results are summarized in Figures 6 and 7.

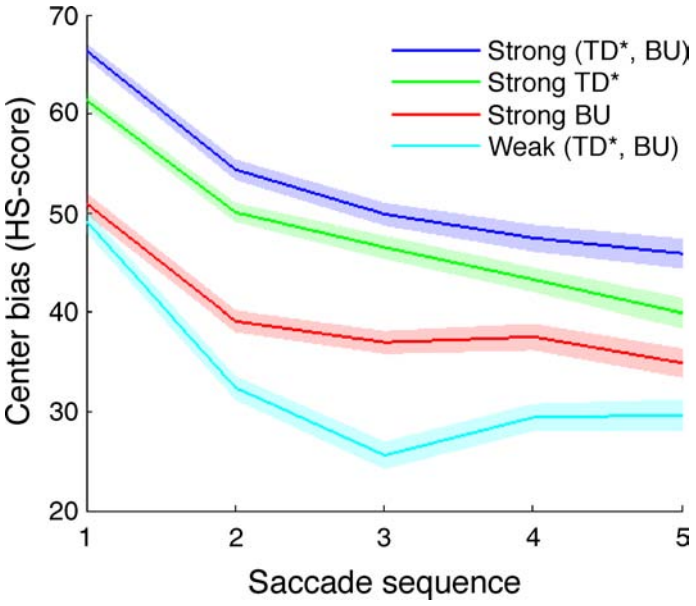

Figure 7. HS score across saccade sequence in 4 different photographer bias conditions.

So far it is shown that photographer bias and viewing strategy were the two major contributors to center bias. Their combined contributions are plotted in Figure 7. Clippets were sorted to one of the four photographer bias conditions based on their TD* and BU scores compared to median. The decreasing trends in 4 conditions were clear and well separated. A two-way ANOVA showed main effects in both saccade sequence ($F(4, 12896) = 165.78$, $p < 0.05$) and photographer bias conditions ($F(3, 12896) = 251.94$, $p < 0.05$), which clearly suggested that viewing strategy contributes to center bias on top of photographer bias. Moreover, in each photographer bias condition, the HS score of the first saccade was higher than that of the following saccades (t-test, Bonferroni-corrected $p < 0.05$). Hence, the initial centering response existed regardless of photographer bias conditions.

Discussion

There are three major observations from Experiment 1. First, center bias (HS score) and photographer bias (TD* and BU scores) were significantly correlated ($R^2 = 0.47$). Second, motor bias does not contribute to center bias. Third, a viewing strategy was observed such that people tended to look at locations closer to the center (Figure 5) immediately after the beginning of new scenes, and the influence had the same order of magnitude as photographer bias.

Contradicting our result, Tatler (2007) found that fixation distributions were not different between images whose features were centrally versus peripherally biased while participants freely viewed them. However, the conclusion was obtained by categorizing images into two conditions post hoc (central/peripheral). Our study, on the other hand, used linear regression that took into account the continuous changes in the degree of center bias. Our method is more sensitive in detecting differences, which might lead us to a different conclusion than Tatler's (2007). Tatler et al. (2005; Tatler, 2007) and Parkhurst et al. (2002) also observed an early central fixation bias in viewing natural images. We show that this early central bias (viewing strategy) also exists in viewing dynamic natural scenes but diminishes quickly (Figure 5). This early central bias is observed in all participants, but the speed of diminishment may vary among participants. Nevertheless, viewing strategy and photographer bias together (Figure 7) suggest that when a new scene is presented, people tend to begin exploring the scene by looking at interesting (top-down) and salient (bottom-up) objects closest to the center. This strategy probably developed by the fact that visual media in our daily life usually present subjects of interest near center, and we are implicitly trained over time to look near center and expect to gain most information when we have no clues about the scene.

Interestingly, Figure 7 showed that TD* and HS scores were correlated from the very first saccade. This result was consistent with previous findings that the "gist" (semantic category of a scene) can be quickly extracted and guides eye movements (Biederman, Mezzanotte, & Rabinowitz, 1982; De Graef, Christiaens, & d'Ydewalle, 1990; Henderson & Hollingworth, 1999; Palmer, 1975; Torralba, 2003). Moreover, the results support the notion that interesting objects can be identified in early stages of scene processing. Loftus and Mackworth (1978) reported a tendency for participants to fixate earlier and longer on objects inconsistent with a line drawing natural scene, which implied semantic process in the early phase of scene perception. However, using the same experimental paradigm but different task instructions and more complex scene, several studies (De Graef et al., 1990; Henderson & Hollingworth, 1999) showed that eye movements are relatively independent of peripheral object identification process in the early phase of scene exploration. Task instructions and complexity of a scene seem to influence the conclusion. In this study, the complexity of a scene was not controlled, but in many of the scenes major actors and events could be easily identified. Moreover, there was no given task ("watch and enjoy the clips") so that participants did not rush to make saccades (saccade latency after scene onset was 355.7 ± 6 ms). Furthermore, TD score was an overall score of the distribution of interesting objects, rather than an individual object. Hence, it is possible that the first saccade was guided by semantic processing (e.g., gist, early phase of object identification). However, other explanations are also plausible.

Carmi and Itti (2006) and Parkhurst and Niebur (2003) suggested another potential cause of center bias that needs to be considered: orbital reserve. Behavioral (Fuller, 1996) and physiological (Paré & Munoz, 2001) studies demonstrated this physiological recentering mechanism.

However, Vitu et al. (2004) also demonstrated that saccade landing position was biased by the center of the screen, rather than orbital reserve. Therefore, we designed the second experiment to measure the difference in terms of saccade endpoint distributions and fixation distributions when the stimuli were no longer displayed right in front of the participants and at screen center. The hypothesis is that if orbital reserve and center of the screen play strong roles in center bias, then the distributions should be biased toward the gaze location of central orbital position and the center of the screen in our setting.

Experiment 2

Experimental procedures were approved by USC Institutional Review Board (IRB).

Stimuli

The same set of MTV-style video stimuli of Experiment 1 were used in Experiment 2. However, only the common clip set and the first specific clip set, whose clippet lengths were all uniformly distributed from 2 to 4 s, were presented to participants.

Participants

Five young adults (4 males and 1 female, range 21–32; mean 25.4 years) with normal or corrected-to-normal vision were recruited for the experiment. They were compensated and were naive to the purpose of the experiment.

Data acquisition

Experiment 2 was conducted in a different laboratory (USC) from Experiment 1 (Queen's University) because it employed a larger screen, which was necessary for the investigation of orbital reserve. Stimuli were displayed on a 46-inch LCD monitor (Sony Bravia XBR-III, 1016 × 571.5 mm), 97.8 cm in front of the participants (corresponding field of view is 54.7° × 32.65°). Participants adjusted the height of the seat to comfortably rest their chin on a chin rest and positioned their eyes in front of the center of the screen so that straight-ahead position and center of the screen were aligned. Participants were instructed to "watch and enjoy the clips". In addition, participants were asked to keep their head fixed so that the eye tracker would not lose track of their eyes. If participants rotated their heads, our table-mounted eye tracker would lose track of their gaze, and the corresponding eye trace was removed from further analysis. Nevertheless, small head movements might occur.

The stimuli were displayed as the same field of view (35.14° × 25.88°) as Experiment 1 rather than being displayed in full screen. They were displayed at 5 different locations, which are the 4 corners (corner displays) and the center (center display) of the screen (Figure 8). Each participant watched the same 40 MTV-style clips, but each clip was randomly placed in one of the five display locations. Because there were 5 locations, only 20% of the total clips were watched in each location for each participant, and each clip was only showed at one particular location for a given participant (thus no clips were displayed at the same location across participants). In short, a 5 (subset of clips) by 5 (display location) Latin-Square design was used.

The forty MTV-style clips were played in random order, calibration was done for every five clips (about 2.5 min), and participants were allowed to leave the chin rest and take a break for every ten clips (about 5 min). At the beginning of each clip, a red cross was displayed at the center of the screen to remind participants to keep their head fixed. When they were ready for the next clip, participants pressed the space bar and the red cross at the center disappeared, and a blinking red cross indicated where the next clip would appear at the center of the next display location. Participants were asked to follow the blinking red cross and then "watch and enjoy the clips". In this way, participants started viewing clips at the center of the stimuli as in Experiment 1; however, in Experiment 2, the stimuli could appear at one of the five display locations.

Eye position was tracked by an ISCAN RK-464 (ISCAN) in pupil-CR mode (240 Hz, gaze position accuracy < 1°) to left eye. Nine-point display calibration was used to compute the affine transform from the eye-tracker coordinates to the stimulus coordinates in the least-square sense. Small nonlinear residual errors in the transformation were corrected by a thin-plate-spline warping algorithm (Bookstein, 1989). Outlier calibration points were eliminated before computing the transformation. If a calibration session had less than 6 valid calibration points, the corresponding eye position data were discarded. Eye-movement traces from two clips out of 200 (5 participants × 40 clips) were discarded due to loss of tracking, or their corresponding clips belonged to specific clip sets (50 eye-movement traces, 5 participants × 10 clips) for the same reason as Experiment 1. Next, the calibrated eye position was further labeled as fixation, saccade, blink/artifact, saccade during blink, smooth pursuit, and drift/misclassification as the same criteria as described in Experiment 1 (8,067 saccades). The first saccade after clip onset, and eye-movement traces from pan-style clips were discarded for the same reason as Experiment 1. In the end, 5,061 saccades, obtained from 2,941.75 s of eye-movement recording (955 participant–clippet pairs), were used for detailed analysis.

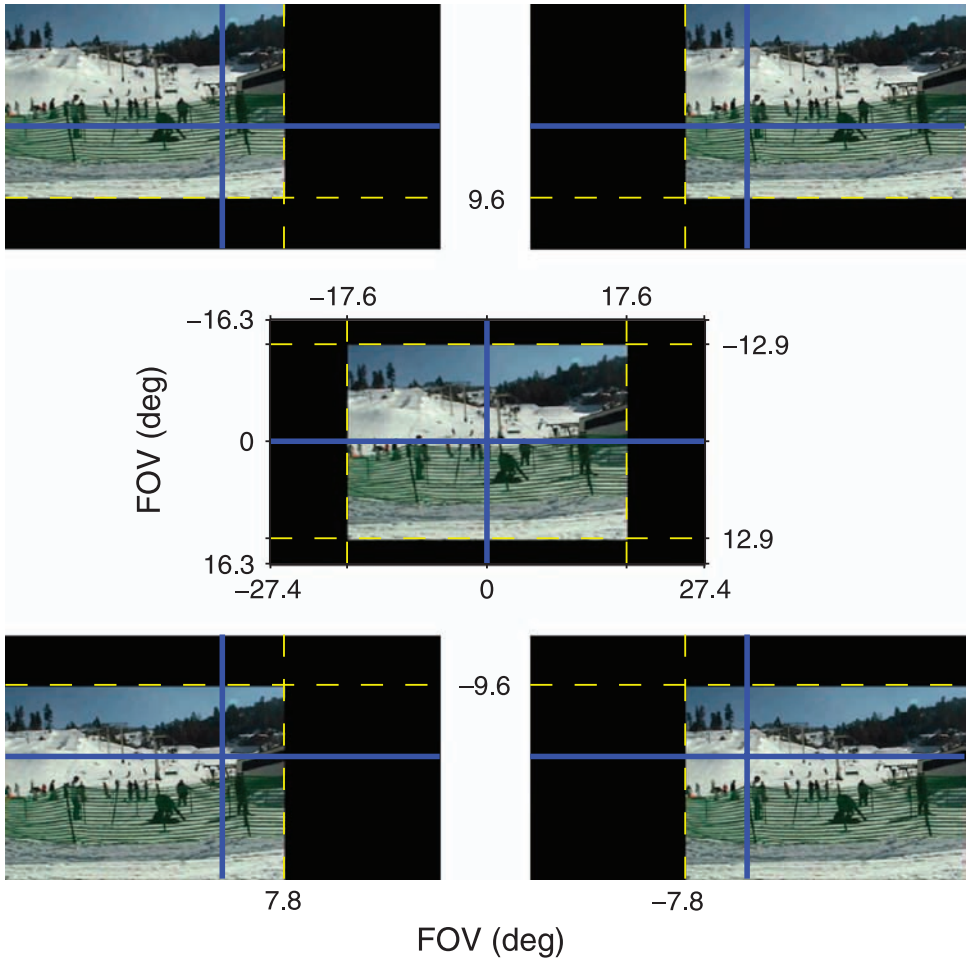

Figure 8. Five display locations. Video was only displayed on the part of the screen whose dimension corresponds to the size of each figure; the region that was not presenting stimuli was left black. Blue crosses show the center of the screen, and the yellow dashed lines are the boundary of the stimuli on the screen. The blue crosses and the yellow dashed lines were not visible to participants.

Data analysis and results

Orbital reserve and center of the screen

To learn how different display locations affect human gaze behavior, the same regression framework was used as that in Experiment 1. HS scores were regressed on display locations encoded as dummy variables. The five locations were encoded as [1 0 0 0] for center display, [0 1 0 0] for upper right corner display, [0 0 1 0] for upper left corner display, [0 0 0 1] for lower right corner display, and [0 0 0 0] for lower left corner display. Then, a linear regression was performed to evaluate the influence of orbital reserve and screen center on center bias.

Figure 9 shows an example of the influence of display locations on saccade endpoint distribution. Although the distributions appeared to be similar, the HS scores of the distribution differed significantly (ANOVA, $F(4, 5056) = 9.22$, $p < 0.05$). However, post hoc paired t-tests (Bonferroni corrected) showed that the HS score of the center display was not different from that of corner displays (Bonferroni-corrected $p > 0.23$) except in the upper right corner ($p = 0.03$). To quantify the influence of display location, linear regression was performed with regressors TD*, BU, and location (Loc) to predict HS score (Figure 10). Display location had a small ($R^2 = 0.014$) but significant effect ($F(4, 950) = 3.29$, $p < 0.05$) on center bias. However, compared to the influence of photographer bias ($R^2 = 0.26$, $F(2, 952) = 167.98$, $p < 0.05$), the effect was one order of magnitude smaller than that of photographer bias; hence, it does not affect attentional selection greatly under this experiment setup.

Shift of display affects gaze distribution

In addition to analyzing the influence of display locations on shifts of attention and gaze, we also analyzed the effect of display locations on the time that participants spent at different locations of the stimuli. If people were more comfortable looking straightforward due to orbital reserve, then we expected they would spend more time watching the quadrant of the scene closer to the center of the physical screen. Figure 11 shows the difference of gaze distribution between corner and center displays.

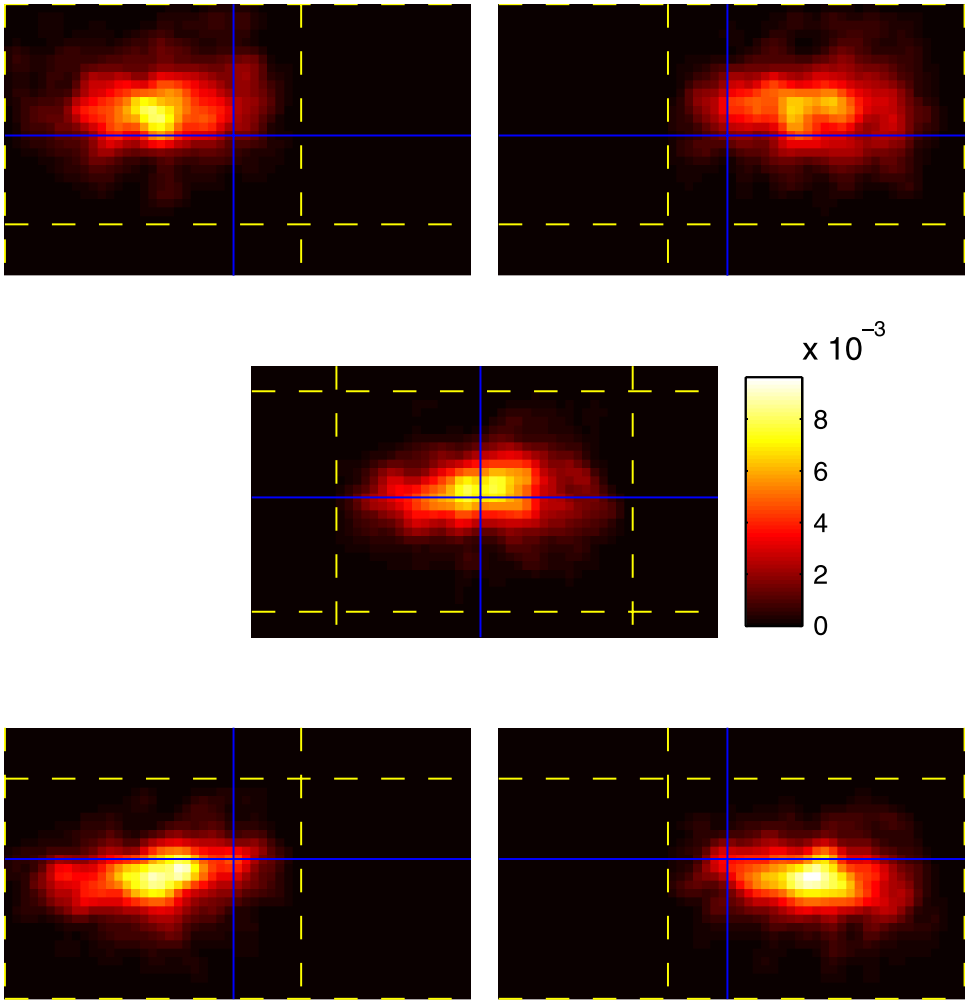

Figure 9. Saccade endpoint distributions of participants watching clippets shown at five different display locations (see Figure 8 for annotations of lines). The distributions were blurred with a Gaussian function (std. = 1°) to account for the radius of fovea. Figures at the corners represent corner displays in their corresponding corner of the screen, and the one at the center is the center display.

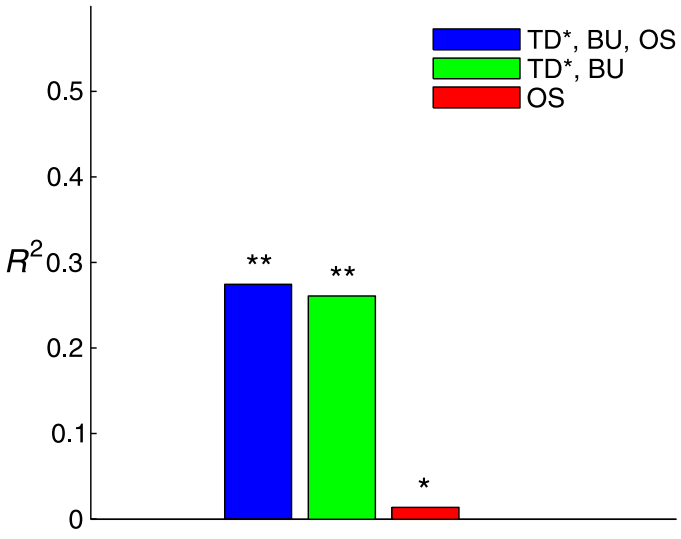

Figure 10. Influence of the combined effect of orbital reserve and screen center (OS) compared to photographer bias (TD*, BU).

Participants spent more time (red) watching at quadrants closer to the screen center and less time (blue) at the quadrants farther away from the screen center. Each quadrant for the four-corner display was labeled as the quadrant that was: (1) closest to the center of screen, (2) horizontally closer to the center of screen, (3) vertically closer to the center of screen, and (4) farthest from the center of screen. Next, the time spent in each type of quadrant was compared to the corresponding quadrant in center display (baseline). A two-tail paired *t*-test in permutation form (10,000 runs) was performed to learn the effect of display location. For quadrants at the locations closest to the center, people spent significantly ($p < 0.025$) more time looking at them than at the same stimuli shown at the center display. Moreover, for the quadrants farthest from the center of the screen, people spent significantly ($p < 0.025$) less time looking at them. However, for those neighboring quadrants in the same horizontal or vertical line to the center, the time people spent was not different from that of the center display

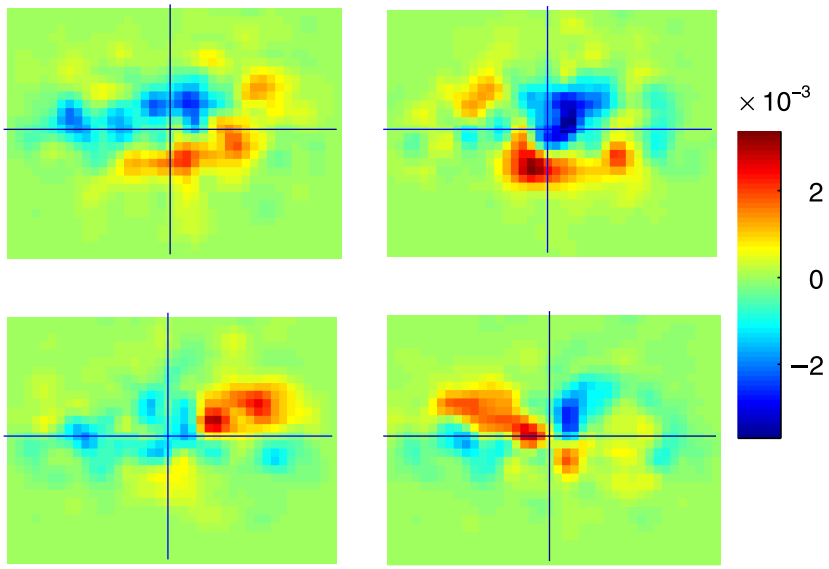

Figure 11. Difference in gaze distribution on stimuli between corner displays and center display. Location of the four histograms represents its corresponding corner on the screen. The red color means the probability that participants looked at that location in a corner display is higher than that in the center display, and the blue color means the opposite.

($p = 0.034$ for horizontally closer quadrant, $p = 0.071$ for vertically closer quadrant). Therefore, observers spent significantly more time watching displays closer to their initial orbital position under this experimental setup.

Discussion

In Experiment 2, we found small, but significant, support that orbital reserve and center of the screen contribute to attentional selection and a component of the center bias commonly observed in studies of visual attention. By a linear regression model, they contributed together ($R^2 < 0.1$) one order of magnitude smaller than photographer bias. However, we found that orbital reserve does increase the time participants spent on locations closer to the central orbital position (Figure 11).

It was interesting to note that when stimuli were not displayed at the center of the screen, saccade endpoint distributions were clustered closer to the center of stimuli rather than the center of the screen or straight-ahead position. This observation suggested that looking at the center of stimuli might allow participants to gain more information of the scenes than by looking at other locations. Moreover, if attentional resources are hemifield independent (Alvarez & Cavanagh, 2005), looking at the center of stimuli may utilize limited attentional capacity maximally.

Paré and Munoz (2001) showed that pre-target activation of neurons in the intermediate layer of the superior colliculus (SC) facilitated eye movements toward central orbital positions. It was not possible to measure precise reaction times to specific stimuli in the current study, as the free viewing experiment did not require the participants to respond to any pre-defined events. Hence, the experiment did not allow us to measure the onset and offset of specific visual events. Nevertheless, our analysis of the time spent in each quadrant does show an increase at the quadrant closest to the central orbital position. Future experiments would be worthwhile in measuring the reaction time with well-labeled natural scene stimuli to learn more about the influence of orbital reserve.

General results

The relative contribution (based on R^2) of possible causes to center bias is summarized in Figure 12. Photographer bias and viewing strategy were the primary causes of center bias and their influences were in the same order of magnitude. First, gaze was strongly attracted to salient (BU) and/or interesting stimuli (TD; together: photographer bias). Viewing strategy was also influential at the beginning of stimulus presentation but diminished quickly. Hence, when the scenes changed (a new clippet started), observers tended to explore the new scenes by starting with attractive locations closer to the center. Motor bias did not contribute to center bias. Orbital reserve and screen center had an effect on attentional selection, but it was found to be roughly ten times smaller than photographer bias under this experimental setup. However, orbital reserve did have a significant effect on viewing time; more time was spent on locations closer to the central orbital position than expected.

Having found that the two major influences were photographer bias and viewing strategy, we explore a

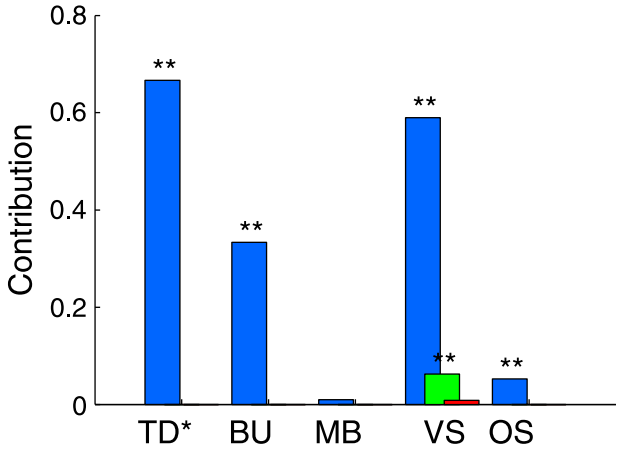

Figure 12. Summary of contribution (based on R^2) of center bias factors relative to photographer bias (TD*, BU). Factors are top-down bias (TD*), bottom-up bias (BU), motor bias (MB), viewing strategy (VS), and the combination of orbital reserve and screen center (OS). The three bars in VS from left to right are the VS estimated from saccade 1 to 5, 2 to 5, and 3 to 5, respectively.

possible method of incorporating these influences into two common methods that evaluate bottom-up influences on gaze. The first method compares the saliency values accumulated along an eye trace (scanpath) against saliency values uniformly (unbiased) sampled in the saliency maps to yield a baseline for comparison (Parkhurst et al., 2002). The advantage of this method is that predictions of the saliency model are compared to a baseline without any behavioral knowledge (about photographers or about observers) and may be considered to be the fairest test because saliency models typically do not have any built-in center bias (i.e., salient locations could be detected anywhere in the display). However, the risk of this measurement is that if there are factors other than salience that may attract observers toward the center (and if the center is also of high salience) then the measurement would overestimate the bottom-up influence on gaze allocation. Alternatively, the second method compares saliency values collected along a scanpath against saliency values sampled from the overall probability distribution of fixations that observers exhibited over the entire experiment to give a baseline (behaviorally derived; e.g., Tatler et al., 2005). The advantage of this method is that it takes into account all the causes of center bias by comparing the model to a baseline derived from the behavior. However, the risk is that if bottom-up attention is the primary cause of center bias (i.e., salient objects are often in central positions due to photographer bias), then this method would underestimate the bottom-up influence on gaze (Carmi & Itti, 2006).

Given that there was an initial center bias upon presentation of a new scene that then diminishes (e.g., viewing strategy, Figure 5), we propose that a biased baseline should be used initially, followed by an unbiased baseline as the saccade sequence progresses. As shown in Figure 12, after the third saccade, viewing strategy did not contribute to center bias, and thus an unbiased baseline could be used after this time (participants were exploring more of the scene). However, for the first 2 saccades, one would need to account for center bias largely created by viewing strategy.

Viewing strategy contributed to center bias as well as photographer bias. Moreover, viewing strategy could be treated as a constant strength toward the center as there was no correlation between viewing strategy and photographer bias in a given clippet ($F(2, 171) = 0.22$, $p = 0.80$). Therefore, the constant strength could be applied to all clippets. The constant strength was assumed to be a 2-D Gaussian function. The estimated saccade endpoint distribution of the first saccade was calculated as the distribution of the late saccades weighted by a Gaussian function centered at the center of the display. Here we used the specific clippet set to generate the 2-D Gaussian function as follows. Saccade endpoint distribution of the first saccade and late saccades were downscaled from 480×640 to 30×40. Next, a 2-D Gaussian-weighted late saccade endpoint distribution was fitted to the first saccade endpoint distribution to minimize the residual sum of squares between the two values. The fitted Gaussian had a standard deviation of 7.08 degrees (FOV) in the horizontal axis and 5.09 degrees in the vertical axis, and the estimated HS score of the first saccade was not different from the data (t-test, $p = 0.36$).

To validate the model, the Gaussian function was applied to each individual clippet in the common set to estimate the HS score of the first saccade from its late saccades (red dots in Figure 13). Blue dots are the real HS scores from 174 clippets. The figure shows that 157 out of 174 clippets had their first saccade more center-biased than their late saccades (above diagonal, green dashed line). Blue and red dashed lines are the linear regression lines of the data and the estimate, and they are different ($F(2, 344) = 4.01$, $p = 0.02$) by testing coincident regression line. Hence, the model cannot explain all the individual clippet variability in the data. There are three possible causes. First, video changes dynamically. The content of the video while making the third saccade is not the same as making the first saccade. Second, because each participant only watched each clippet once, the estimated HS score of the first saccade was predicted by a small sample of late saccades (37.08 ± 7.67 saccades). Moreover, the HS scores of the first saccade also came from small samples (15.53 ± 1.12 saccades). Third, there is more than a simple Gaussian contributing to the viewing strategy.

Nevertheless, the model did account for some of the variance, as the red line was much closer to blue line compared to the green line ($F(2, 344) = 140.75$, $p < 0.0001$), which assumed the HS score of the first saccade equaled that of the late saccades. Moreover, 84 out of 174 clippets had an estimated HS score of the first saccade within the 95% confidence interval of the data (2-tail permutation test,

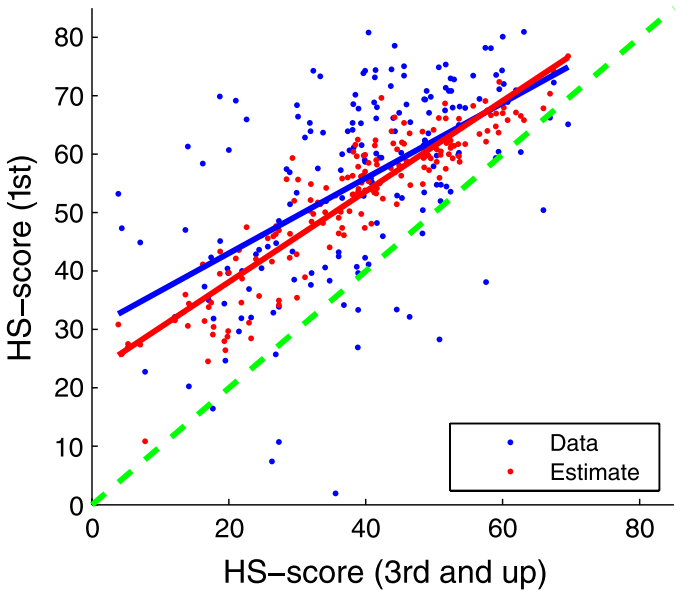

Figure 13. HS score of estimated (red dots) and true (blue dots) first saccade versus the HS score of late saccades. Green dashed line assumes that the HS score of the first saccade is equal to that of late saccades. The first saccade is more center-biased if it is above the green dashed line. Red and blue lines are the regression lines for estimated and true HS score of the first saccade.

Monte Carlo, N = 10,000). Thus, the model performed much better than using the saccade endpoint distribution of late saccades. This result indicates it is more appropriate to use a center-biased baseline for initial saccades.

General discussion

In this study we examined the roles of five factors (photographer bias, viewing strategy, orbital reserve, screen center, and motor bias) suspected in creating a center bias when viewing natural scenes. Importantly, we examined fixation patterns and used a model of bottom-up (BU) saliency, combined with measures of interesting locations (TD), to identify the major factors (photographer bias and viewing strategy) in center bias. We also considered the implications of these findings and proposed a pilot solution to accommodate viewing strategy into the baseline for evaluating the correlation between gaze and predictors such as salience.

The proposed solution involves using both unbiased and behaviorally derived baselines. Comparing saliency values at fixations to either baseline derivation (unbiased vs. behaviorally derived) has been shown previously to result in a significant predictive power of bottom-up saliency maps. One of the differences between the two methods lies in how the bottom-up influence evolves over time. With the unbiased baseline, Parkhurst et al. (2002) showed that bottom-up factors have the strongest impact at the beginning of stimuli presentation and then decreases over time. However, other evidence suggests that by a behaviorally derived baseline, bottom-up remains at the same level across long time periods (Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Tatler et al., 2005). The cause of the difference is center bias upon initial image presentation (Tatler, 2007). The current study extends this finding to show that the same center bias exists while people are watching dynamic visual scenes. Moreover, people immediately direct their first saccade in a new visual scene (independent of explicit reorienting cues) toward subjects of interest or salient locations closer to the center. Thus, it is necessary to incorporate a behaviorally derived baseline for initial saccades. The cutoff between initial and late saccades might not be constant and may be dependent on the particular experimental stimuli. We suggest that the cutoff between early and late saccades should be determined as when the HS score starts to stabilize. Another difference in using unbiased and behaviorally derived baselines is the estimation of overall bottom-up influence. Usually the baseline by fixation distribution is obtained from the fixation distribution while participants watch other stimuli (Mannan et al., 1996; Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Tatler et al., 2005). However, the degree of photographer bias of those stimuli is uncontrolled (Figure 7). Taking an extreme example, if all stimuli have strong photographer bias (e.g., one dot at the center with blank background), the baseline by fixation distribution will be highly clustered at the center. Therefore, the baseline is already strongly contaminated with photographer bias. Comparing salience and gaze against this contaminated baseline cannot reveal the true influence of salience. To estimate bottom-up influences properly, our study suggests that using a mixture of these two baselines is more suitable.

An improvement of the baseline would be to incorporate the top-down component of photographer bias. However, adjusting for top-down factors is difficult and would require subjective ratings of object of interest for a particular scene. Moreover, because interesting objects are usually salient (Elazary & Itti, 2008a), it is even more difficult to tell from gaze whether observers shift their attention due to top-down or bottom-up causes exclusively. However, Peters and Itti (2007) use a top-down model that learns the task based on the correlation between the "gist" of the whole scene and eye movements. The model predicted human eye movements better than Itti and Koch's (2000) saliency model by a factor of 2. This demonstrates how strong the cognitively interesting and task-relevant objects could affect eye movement patterns. Therefore, it again indicates the need for integrating task modeling (Frintrop, Backer, & Rome, 2005; Navalpakkam & Itti, 2005; Peters & Itti, 2007), object recognition (Elazary & Itti, 2008b; Lowe, 1999; Riesenhuber & Poggio, 1999) for eye movement prediction to advance.

In summary, this study attempts to distinguish quantitatively the relative contributions of five factors (photographer bias, motor bias, viewing strategy, orbital reserve, and screen center) to center bias in free viewing of natural videos. Based on the understanding of their relative contribution, a new baseline is proposed for evaluating fairly the correlation between salience and gaze. Future work could be built on this study and pose additional questions. For example, by giving participants a task, how would the relative contributions of these factors change, and what will the proper baseline be? In addition, beyond the five factors discussed in this study, one may also ask how other factors would affect attentional selection, or how to better control natural scene stimuli (e.g., objective estimation of objects/regions of interest) as strictly as is customary in rigorous psychophysical experiments with less complex stimuli. Quantitative evaluation of their contributions serves as a key to understanding the processes guiding visual attention.

Acknowledgments

This work was supported by grants from the National Science Foundation, the Human Frontier Science Program, the National Geospatial-Intelligence Agency, and the Defense Advanced Research Projects Agency. The authors affirm that the views expressed herein are solely their own and do not represent the views of the United States government or any agency thereof.

Commercial relationships: none.
Corresponding author: Po-He Tseng.
Email: ptseng@usc.edu.
Address: 3641 Watt Way, #6, Los Angeles, CA 90089, USA.

References

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science, 16,* 637–643. [PubMed]

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergrouping relational violations. *Cognitive Psychology, 14,* 143–177. [PubMed]

Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11,* 567–585. [Article]

Busswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: University of Chicago Press.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research, 46,* 4333–4345. [PubMed]

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52,* 317–329. [PubMed]

Elazary, L., & Itti, L. (2008a). A bayesian model of visual search and recognition [Abstract]. *Journal of Vision, 8*(6):841, 841a, http://journalofvision.org/8/6/841/, doi:10.1167/8.6.841.

Elazary, L., & Itti, L. (2008b). Interesting objects are visually salient. *Journal of Vision, 8*(3):3, 1–15, http://journalofvision.org/8/3/3/, doi:10.1167/8.3.3. [PubMed] [Article]

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*(2):6, 1–17, http://journalofvision.org/8/2/6/, doi:10.1167/8.2.6. [PubMed] [Article]

Frintrop, S., Backer, G., & Rome, E. (2005). Goal-directed search with a top-down modulated computational attention system. In *Pattern recognition, proceedings: Lecture notes in computer science* (vol. 3663, pp. 117–124). Springer.

Fuller, J. H. (1996). Eye position and target amplitude effects on human visual saccadic latencies. *Experimental Brain Research, 109,* 457–466. [PubMed]

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243–271. [PubMed]

Itti, L. (2004). Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing, 13,* 1304–1318. [PubMed]

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506. [PubMed]

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227. [PubMed]

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565–572. [PubMed]

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the 7th IEEE International Conference on Computer Vision* (vol. 2, pp. 1150–1157). Springer.

Mannan, S., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9,* 363–386. [PubMed]

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision, 10,* 165–188. [PubMed]

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision, 11,* 157–178. [PubMed]

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45,* 205–231. [PubMed]

Palmer, S. E. (1975). The effect of contextual scenes on the identification of objects. *Memory & Cognition, 3,* 519–526.

Paré, M., & Munoz, D. P. (2001). Expression of a re-centering bias in saccade regulation by superior colliculus neurons. *Experimental Brain Research, 137,* 354–368. [PubMed]

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 107–123. [PubMed]

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16,* 125–154. [PubMed]

Peters, R. J., & Itti, L. (2007). Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8). IEEE.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network, 10,* 341–350. [PubMed]

Renninger, L. W., Coughlan, J., Verghese, P., & Malik, J. (2005). An information maximization model of eye movements. *Advances in Neural Information Processing System, 17,* 1121–1128. [PubMed]

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025. [PubMed]

Schumann, F., Einhäuser-Treyer, W., Vockeroth, J., Bartl, K., Schneider, E., & König, P. (2008). Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. *Journal of Vision, 8*(14):12, 1–17, http://journalofvision.org/8/14/12/, doi:10.1167/8.14.12. [PubMed] [Article]

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14):4, 1–17, http://journalofvision.org/7/14/4/, doi:10.1167/7.14.4. [PubMed] [Article]

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. [PubMed]

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research, 46,* 1857–1862. [PubMed]

Torralba, A. (2003). Contextual priming for object detection. *International Journal of Computer Vision, 53,* 153–167.

Tweed, D. (1997). Visual-motor optimization in binocular control. *Vision Research, 37,* 1939–1951. [PubMed]

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology (Colchester), 59,* 1931–1949. [PubMed]

Van Opstal, A. J., Hepp, K., Suzuki, Y., & Henn, V. (1995). Influence of eye position on activity in monkey superior colliculus. *Journal of Neurophysiology, 74,* 1593–1610. [PubMed]

Vitu, F., Kapoula, Z., Lancelin, D., & Lavigne, F. (2004). Eye movements in reading isolated words: Evidence for strong biases towards the center of the screen. *Vision Research, 44,* 321–338. [PubMed]