Psychonomic Bulletin and Review (2019) 26:1683–1689
https://doi.org/10.3758/s13423-019-01642-5

BRIEF REPORT

Scene semantics involuntarily guide attention during visual search

Taylor R. Hayes[1] · **John M. Henderson**[1,2]

Published online: 24 July 2019

Abstract

During scene viewing, is attention primarily guided by low-level image salience or by high-level semantics? Recent evidence suggests that overt attention in scenes is primarily guided by semantic features. Here we examined whether the attentional priority given to meaningful scene regions is involuntary. Participants completed a scene-independent visual search task in which they searched for superimposed letter targets whose locations were orthogonal to both the underlying scene semantics and image salience. Critically, the analyzed scenes contained no targets, and participants were unaware of this manipulation. We then directly compared how well the distribution of semantic features and image salience accounted for the overall distribution of overt attention. The results showed that even when the task was completely independent from the scene semantics and image salience, semantics explained significantly more variance in attention than image salience and more than expected by chance. This suggests that salient image features were effectively suppressed in favor of task goals, but semantic features were not suppressed. The semantic bias was present from the very first fixation and increased non-monotonically over the course of viewing. These findings suggest that overt attention in scenes is involuntarily guided by scene semantics.

Keywords Scene perception · Attention · Semantics · Salience · Visual search

Eye movements are the primary way we select and extract information from the world around us (Findlay & Gilchrist, 2003). How do we determine where to look in complex, real-world scenes? This foundational question has generated two distinct theoretical frameworks: image guidance theory and cognitive guidance theory (Itti & Koch, 2001; Henderson, 2007). Under image guidance theory, our attention is primarily guided by spatial discontinuities in low-level, semantically uninterpreted image features such as color, orientation, and/or luminance (Itti & Koch, 2001; Harel, Koch, & Perona, 2006). In comparison, cognitive guidance theory posits that our attention is primarily guided by the distribution of semantic information in a scene, informed by stored scene knowledge that guides our attention to where semantic content is likely to occur (Henderson, 2003, 2017; Hayhoe & Ballard, 2005). These two theories have had a broad impact across psychology (Henderson, 2017; Wolfe & Horowitz, 2017; Itti & Borji, 2014).

Image guidance theory has recently been the dominant theoretical approach because it is easy to generate a saliency map from image features (Henderson, 2017; Itti & Borji, 2014; Itti, Koch, & Niebur, 1998). This has led to a number of different bottom-up image salience models and a vast amount of research on the role of image salience during scene viewing (see Itti & Borji, 2014, for review). Unfortunately, generating a map of the distribution of semantic features in a scene is not as straightforward. However, without a semantic analog of a saliency map, it is difficult to quantify the relative merits of image guidance and cognitive guidance theories during scene viewing (Henderson & Hayes, 2017).

To evaluate these competing theories, we recently introduced a new method for estimating the distribution of semantic features within a scene (meaning maps, Henderson & Hayes, 2017). Meaning maps draw inspiration from two classic scene-viewing studies (Antes, 1974; Mackworth &

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13423-019-01642-5) contains supplementary material, which is available to authorized users.

✉ Taylor R. Hayes
taylor.r.hayes@gmail.com

[1] Center for Mind and Brain, University of California, Davis, Davis, CA, USA

[2] Department of Psychology, University of California, Davis, Davis, CA, USA

Morandi, 1967). In these studies, images were divided into several regions and subjects were asked to rate each region based on how easy it would be to recognize (Antes, 1974) or how informative it was (Mackworth & Morandi, 1967). Critically, when a separate group of subjects freely viewed the same images, they mostly looked at the regions that were rated as highly recognizable or informative. Meaning maps scale up this general rating procedure using crowd-sourced ratings of thousands of isolated scene patches densely sampled at multiple spatial scales (Henderson & Hayes, 2017). The resulting scene meaning maps capture the spatial distribution of semantic features, just as saliency maps capture the spatial distribution of image features. While we are still a long way from having a computational model of scene semantics, meaning maps provide a foothold for studying the role of semantic features within complex, real-world scenes.

Using meaning and saliency maps, we have previously compared how well scene semantics and image salience predicted overt attention during scene memorization, aesthetic judgment, scene description, brightness search, and brightness rating tasks (Henderson & Hayes, 2017, 2018; Henderson, Hayes, Rehrig, & Ferreira, 2018; Peacock, Hayes, & Henderson, 2019). In every task, meaning explained overt attention better than image salience. Importantly, the attentional advantage for meaning was present from the very first fixation, suggesting that scene gist (see Oliva & Torralba, 2006, for review) rapidly biases participants toward more semantically rich scene regions (Henderson & Hayes, 2017, 2018). This raises the question whether the attentional bias toward scene semantics is involuntary. By involuntary we simply mean that overt attention is allocated toward scene semantics even when performing a task that is independent of semantics. We are not making any claims about the role of attention in activating scene semantics.

A few previous studies have indicated that semantic features may bias attention in object arrays (Malcolm, Rattinger, & Shomstein, 2016) and real-world scenes (Cornelissen & Võ, 2017; Peacock et al., 2019) even when the semantics are not task relevant. Malcolm et al. (2016) found that in object triplet arrays, task-irrelevant semantic relationships between two objects biased attention as measured by faster reaction times in an independent target detection task. Cornelissen and Võ (2017) found increased dwell time on a semantically incongruent object in scenes while performing an unrelated letter search task. Finally, Peacock et al. (2019) found that when participants performed a task that evaluated a scene-dependent image feature (i.e., overall brightness of the scene or counting the bright regions in the scene), semantics still accounted for more variance in fixation density than image salience.

In the present study, we examined whether the bias toward meaningful scene regions is involuntary by examining how well image salience and meaning can each be suppressed in favor of a scene-independent task goal. Participants were asked to search for hard to find superimposed letter targets that were randomly placed in each scene. Importantly, in the 40 critical scenes used for analysis, there were no targets. The lack of targets kept participants searching these scenes throughout the trial while avoiding potential contaminants associated with target fixations.

Our study addressed a variety of gaps left by previous work. First, our study directly compared the influences of image salience and semantics in real-world scenes rather than object triplet arrays (Malcolm et al., 2016). This is an important difference because full scenes allow us to directly evaluate the role of scene gist in early semantic guidance. Second, our study evaluated how image salience and meaning are related to overt attention across the entire scene, rather than on a single object in each scene (Cornelissen & Võ, 2017). This allows for a continuous measure of semantics and more statistical power for detecting any meaning effects. Finally, it could be that any scene task that requires evaluating a scene-dependent feature (i.e., a property of the scene itself), even an image property like brightness such as Peacock et al. (2019), also activates scene semantics. Therefore, our study used a task that is scene-independent (i.e., the visual search task can be performed without the scene). Another key difference between the current study and Peacock et al. (2019) is in the mechanism being targeted by the experimental task. The current study is a pure test of how well image salience and semantics can each be *suppressed* to pursue a scene-independent task goal, rather than if image salience can be *enhanced* by making a salient scene-dependent feature (i.e., brightness) task relevant (Peacock et al., 2019).

To summarize, we used a scene-independent visual search paradigm in which scene semantics and image salience are both unrelated to the search to investigate whether scene semantics involuntarily guide attention.

Method

Participants

Forty University of California, Davis undergraduate students with normal or corrected-to-normal vision participated in the visual search study and 165 Amazon Mechanical Turk workers participated in the meaning map study. Each study was approved by the institutional review board at the University of California, Davis. All participants were naïve concerning the purposes of each experiment and provided verbal or written consent.

Stimuli

The visual search stimuli contained 80 digitized photographs (1024 × 768) of indoor and outdoor scenes (Fig. 1a). Forty of these scenes were the meaning mapped scenes from Henderson and Hayes (2017) and the other 40 scenes were new. The 40 new scenes contained either 1 or 2 superimposed letter L targets. The letter L targets were small (9 × 11 pixels) and were matched to the global scene luminance value to make them difficult to find and avoid 'popout' effects. The letter targets were randomly placed in each scene excluding the area within 2° of the pre-trial fixation cross. The 40 meaning mapped scenes were all assigned to the target absent condition to avoid any contamination due to target fixations during these critical trials.

Apparatus and procedure

Eye movements were recorded with an EyeLink 1000+ tower-mount eye tracker (spatial resolution 0.01°) sampling at 1000 Hz (SR Research, 2010). Participants sat 85 cm away from a 21″ monitor, so that scenes subtended 27° × 20.4° of visual angle at a resolution of 1024 × 768 pixels. Head movements were minimized using a chin and forehead rest.

Subjects were instructed to search each scene for small embedded letter L targets. Each trial began with a pretrial fixation for 300 ms on a central fixation cross. Then each scene was presented for 12 s while subjects searched for the targets. At the end of each trial, subjects indicated how many targets they located via button press.

Fixation density maps

A fixation density map (Fig. 1c) based on the x and y coordinates of all fixations (Fig. 1b) was generated across participants for each scene. Following our previous work (Henderson & Hayes, 2017, 2018), a Gaussian low-pass filter with a circular boundary and a cutoff frequency of $-6dB$ was applied to account for foveal acuity and eye-tracker error.

Saliency maps

A saliency map was generated for each target absent scene (Fig. 1f) using the Graph-based Visual Saliency (GBVS) toolbox with default settings and no center bias (Harel et al., 2006). We chose to compare meaning to the GBVS model because it is based on known low-level mechanisms of the human visual system (Harel et al., 2006; Itti et al., 1998; Itti & Koch, 2001) and is one of the best performers (Walther & Koch, 2006). In comparison, it is less clear whether state-of-the-art deep neural network models (e.g., Deep Gaze II; Kümmerer, Wallis, Gatys, & Bethge, 2017) that learn where

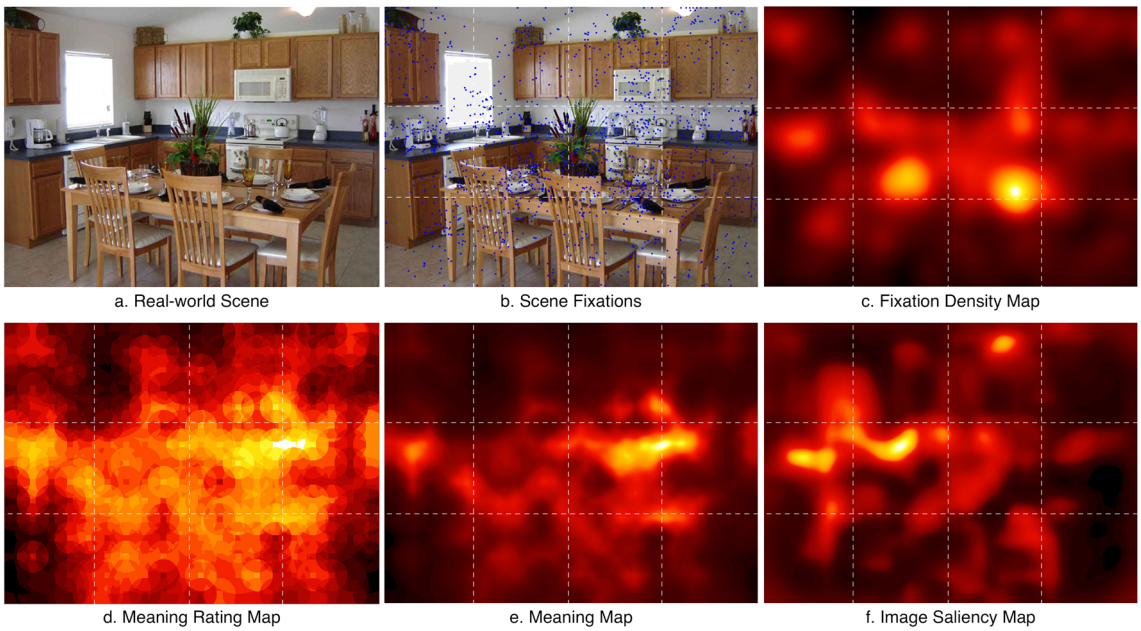

Fig. 1 *Scene and corresponding fixation density, meaning, and saliency maps.* The *top row* shows a typical scene (**a**), the individual fixations produced by all participants during the visual search task (**b**), and the resulting fixation density map (**c**) for the scene. The fixation density map was compared to the meaning map (**e**) and the saliency map (**f**). The meaning rating map (**d**) shows the raw rating data across both spatial scales. The meaning maps (**e**) and saliency maps (**f**) were each normalized using image histogram matching using the fixation density map (**c**) as the reference image. Note that grid lines are included here for easy of comparison and were not included in the experiment

people attend in scenes from training on sets of fixations over object features share this same biological plausibility. More importantly, deep learning models are a poor fit for the current work because we want to cleanly dissociate low-level image features associated with image guidance theory from high-level semantic features associated with cognitive guidance theory.

Meaning maps

Meaning maps were generated as an estimate of the spatial distribution of semantic information in each scene (Henderson & Hayes, 2017). Meaning maps were created for each target-absent scene by decomposing the scene into a dense array of overlapping circular patches at a fine spatial scale (300 patches with a diameter of 87 pixels) and coarse spatial scale (108 patches with a diameter of 207 pixels). Participants ($N = 165$) on Amazon Mechanical Turk then provided ratings of thousands (16320) of scene patches based on how informative or recognizable they thought they were on a six-point Likert scale. Patches were presented in random order and without scene context, so ratings were based on context-independent judgments. Each unique patch was rated by three unique raters.

A meaning map (Fig. 1e) was generated for each target absent scene by averaging the rating data at each spatial scale separately, then averaging the spatial scale maps together, and finally smoothing the average rating map (Fig. 1d) with a Gaussian filter (i.e., Matlab 'imgaussfilt' with sigma = 10).

Map-level correlation metric

We used map-level correlation to quantify the strength of the relationship between the saliency maps, meaning maps, and fixation density maps. Linear correlation is one of the most widely used and best overall metrics due to its minimal assumptions and sensitivity (Bylinskii, Judd, Oliva, Torralba, & Durand, 2016). We present the correlation data as squared correlation because R^2 is broadly understood in psychology as the proportion of variance explained, which is helpful when comparing the shared, overall, and unique variance explained. Finally, using R^2 allows for a direct comparison between the current scene-independent task and all our previous work using scene-dependent viewing tasks.

Results

We first quantified the relationship between the saliency and meaning maps themselves. The left plot in Fig. 2a shows the squared correlation (R^2) between the saliency and meaning maps for all 40 scenes. The squared correlation between the saliency and meaning maps was 0.15 ($SD = 0.12$). A one-sample t test confirmed that the squared correlation was significantly greater than zero, $t(39) = 7.75$, $p < 0.001$, 95% CI [0.11, 0.19]. These results indicate that meaning and saliency maps share a significant amount of overlap, but are more different than they are similar in the absence of the shared GBVS center bias (Henderson & Hayes, 2017).

Next, we tested how well the saliency and meaning maps accounted for fixation density during our scene-independent visual search task. Figure 2a shows the overall and unique variance explained by the meaning and saliency maps in the fixation density maps. Each data point shows the R^2 value for the observed fixation density maps for image salience (blue) and meaning (red). On average across the 40 scenes, meaning accounted for 30% of the variance ($SD = 0.15$) and image salience accounted for 8% of the variance in the fixation density maps ($SD = 0.08$). A two-tailed t test revealed that this difference was statistically significant, $t(78) = 8.42$, $p < 0.001$, 95% CI [0.17, 0.28]. In addition, meaning maps captured twelve times as much unique variance ($M = 0.24$, $SD = 0.13$) as image salience ($M = 0.02$, $SD = 0.02$). A two-tailed t test confirmed that this difference was statistically significant, $t(78) = 11.13$, $p < 0.001$, 95% CI [0.19, 0.27]. These results suggest that while performing a scene-independent

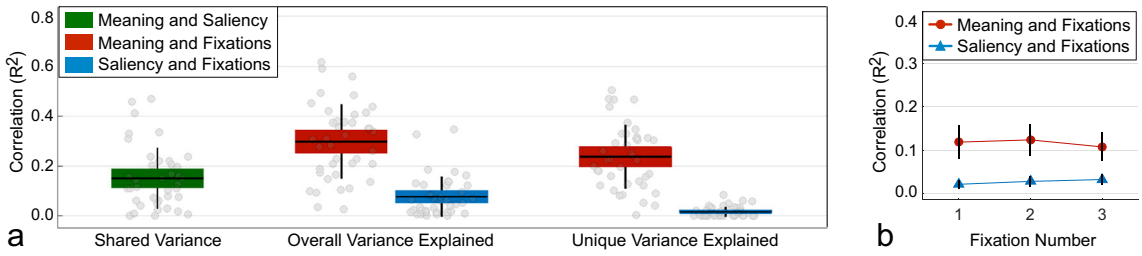

Fig. 2 Squared correlation between fixation density maps and meaning and saliency maps. The scatter box plots (**a**) indicate the shared variance between the meaning and saliency maps, and the overall and unique variance in fixation density they each explained. The scatter box plots show the corresponding grand correlation mean (*horizontal line*) across all 40 scenes (*circles*), 95% confidence intervals (*box*) and 1 standard deviation (*vertical line*). The line plot (**b**) shows the squared correlation between the meaning and saliency maps and the fixation density maps for the first three scene fixations. *Error bars* indicate 95% confidence intervals

orthogonal task, image salience is effectively suppressed, but meaning is not.

There has been some evidence suggesting that early attentional guidance may be more strongly driven by image salience (O'Connel & Walther, 2015; Anderson, Donk, & Meeter, 2016) and also disagreement on the time course of semantic feature bias (de Groot, Huettig, & Olivers, 2016; Malcolm et al., 2016). To examine early attention for image salience and meaning effects, we repeated the same analysis for just the first three fixations. The squared correlation was computed as described above, but was based on fixation density maps that aggregated across the eye movement data as a function of the fixations up to that point. That is, for each scene, we computed the fixation density map that contained only the first fixation for each subject, then the first and second fixation for each subject, and so on to form the fixation density map at each time point.

The first three fixations from the accumulating fixation analysis are shown in Fig. 2b. The squared correlation was stronger between the meaning and fixation density maps for all time steps. Meaning accounted for 11.9, 12.4, and 10.8% of the variance in the first three fixations, whereas image salience accounted for 2.1, 2.8, and 3.2%, respectively. Two-sample, two-tailed t tests were performed for each time point (fixation 1, $t(78) = 4.93$, $p < 0.001$, 95% CI [0.06, 0.14]; fixation 2, $t(78) = 4.99$, $p < 0.001$, 95% CI [0.06, 0.13]; fixation 3, $t(78) = 4.33$, $p < 0.001$, 95% CI [0.04, 0.11]) with p values corrected for multiple comparisons using the false discovery rate (FDR) correction (Benjamini & Hochberg, 1995). Additionally, we confirmed that the amount of variance in fixation density explained by meaning was greater than 0, establishing a significant semantic bias for these critical early fixations (fixation 1, $t(39) = 5.95$, $p < 0.001$, 95% CI [0.08, 0.16]; fixation 2, $t(39) = 6.42$, $p < 0.001$, 95% CI [0.09, 0.16]; fixation 3, $t(39) = 6.22$, $p < 0.001$, 95% CI [0.07, 0.14]). These findings indicate that scene semantics bias even the earliest scene fixations.

We also performed two post hoc analyses. The first analysis examined whether the increase in the squared correlation between fixation density and meaning over time (Fig. 3a, red circles) was best explained by a linear or polynomial curve. We used the Akaike information criterion (AIC) to determine the best fitting model (Akaike, 1974). The maximum AIC was achieved by a 4th-order polynomial model (AIC $= -460.36$, $R^2 = 0.999$). An F-test comparing the linear and 4th-order polynomial models was significant, $F(3, 50) = 1126.2$, $p < 0.001$. These findings suggest the strength of the relationship between scene semantics and attention changed non-monotonically over the course of scene viewing.

The second post hoc analysis quantified the spatial distribution and relative strength of the semantic bias for the first scene fixation. Each scene meaning map was normalized to have a mean of 0 and standard deviation of 1, and then each fixation location and meaning value for that location were aggregated across subjects and scenes to produce Fig. 3b. Each dot represents a first fixation where red indicates fixations on higher meaning map regions and blue indicates fixations on lower meaning map regions compared to the mean meaning value for that scene. The results confirmed there was a significant bias toward more semantically meaningful regions across all fixations, $t(1519) = 34.47$, $p < 0.001$, 95% CI [0.76, 0.85]. In addition, we measured the percentage of foveal (<= 3°, indicated by dotted black line) and extrafoveal (>3°) semantically biased fixations (meaning values > 0). We found that 38.6% (474/1229) were foveal and 61.4% were extrafoveal (755/1229). This suggests that scene gist can bias overt attention toward informative scene regions within and outside of foveal attention.

Discussion

We have previously shown that scene semantics are much better predictors of where people look in scenes than

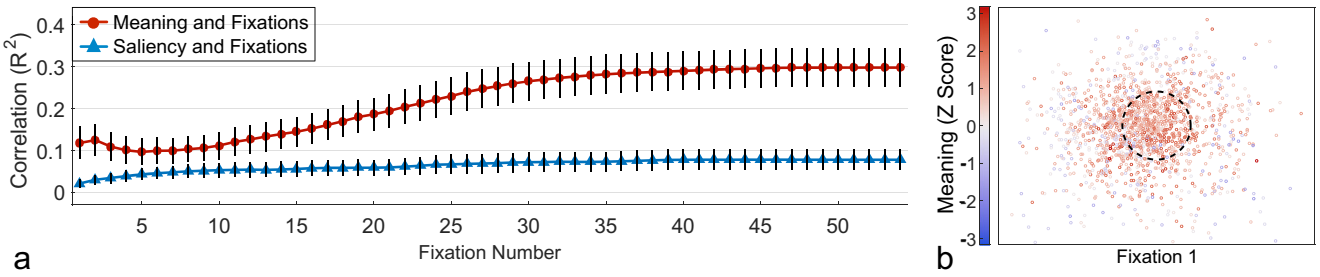

Fig. 3 Squared correlation over time and the spatial distribution and strength of semantic bias for the first fixation. The line plot (**a**) shows the squared correlation between the fixation density maps and the meaning and saliency maps accumulating across fixations within a scene. *Error bars* indicate 95% confidence intervals. The scatter plot (**b**) shows the spatial distribution of all first scene fixations as a function of their meaning rating. The *dotted black line* indicates foveal attention (3° radius)

image salience during a variety of scene-dependent viewing tasks (Henderson & Hayes, 2017, 2018; Henderson et al., 2018; Peacock et al., 2019). Here we tested whether scene semantics involuntarily guide overt attention using a scene-independent, orthogonal visual search task. We found that fixation density was still more strongly correlated with meaning maps than image saliency maps. Critically, the bias toward scene semantics was observed across all fixations, including the first fixation, indicating a faster semantic bias in scenes than in object arrays (Malcolm et al., 2016). These findings support cognitive guidance theories of attention and provide new evidence that semantic features involuntarily guide overt attention in real-world scenes.

The present results raise two interesting questions: How is overt attention biased toward scene semantics so quickly in scenes, and why would viewers be biased by scene semantics while performing an orthogonal, scene-independent visual search task? We believe the most plausible answer to both of these questions is scene gist.

Scene gist is generated rapidly (<100 ms) and provides the observer with information about the likely scene category, coarse spatial layout, broad actions in the scene, and sometimes specific objects within the scene (see Oliva & Torralba, 2006 for review). Moreover, scene gist extraction is so efficient it requires only minimal attention to be allocated to the scene (Cohen, Alvarez, & Nakayama, 2011). Therefore, scene gist is fast, requires minimal attention, and contains sufficient information to bias even the first fixation toward scene regions that are more likely to be semantically informative (e.g., the counter tops in the kitchen scene in Fig. 1a). We speculate that scene gist initially produces only a coarse representation of where meaningful information is likely to occur in the scene, and that subsequent fixations then refine this coarse semantic representation. The increase in correlation between the fixation density and meaning maps as the trial unfolds (Fig. 3a) suggests that this semantic refinement process increases the likelihood to fixate and/or refixate meaningful scene regions, and that the refined semantic map is harder to suppress.

Scene gist also offers a plausible explanation for why the semantic bias is involuntary: scene gist extraction is involuntary (Greene & Fei-Fei, 2014). Greene and Fei-Fei (2014) recently used a modified Stroop paradigm where images of objects and scenes were presented with superimposed congruent or incongruent nouns. They found slower classification for nouns when placed on an incongruent scene, and this effect was modulated by the degree of attention allocated to the scene. This suggests that as long as attention is at least partially allocated to the scene, scene categorization is involuntary. Our findings that the first fixation is biased toward semantically informative scene regions in a scene-independent task adds converging evidence from eye movement data that scene gist extraction is involuntary. Additionally, our analysis of the spatial distribution and meaning values of the first fixations suggests that the semantic bias from scene gist can quickly guide fixations to meaningful scene regions within and outside of foveal attention.

While our results suggest scene semantics bias attention in a scene-independent task, they also demonstrate that image salience is effectively suppressed. Weak image-based guidance has been previously observed for isolated, local scene regions (Henderson, Malcolm, & Schandl, 2009; Vincent, Baddeley, Correani, Troscianko, & Leonards, 2009), and across the entire scene in scene-dependent viewing tasks (Henderson & Hayes, 2017, 2018; Henderson et al., 2018; Peacock et al., 2019). In our scene-independent viewing task, image salience accounted for very little unique variance above and beyond meaning. This result suggests that image salience may be almost completely suppressed while performing a scene-independent task.

Eye movements are the primary way we interact with our environment. This makes understanding the factors that guide our visual attention in complex scenes a central issue in cognitive science. Here we have provided new evidence that semantic features are rapidly prioritized and may involuntarily guide attention during even a scene-independent viewing task. Moving forward it will be important to look beyond image salience and task relevance for the deeper role of high-level scene semantics in how we understand attentional guidance in real-world scenes.

Acknowledgements This research was supported by the National Eye Institute of the National Institutes of Health under award number R01EY027792. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Open Practices Statement All of the scenes, fixation density, saliency, and meaning maps are included in the supplement. The study was not preregistered.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Anderson, N. C., Donk, M., & Meeter, M. (2016). The influence of a scene preview on eye movement behavior in natural scenes. *Psychonomic Bulletin & Review*, *23*(6), 1794–1801.

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, *103*(1), 62–70.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2016). What do different evaluation metrics tell us about saliency models? arXiv preprint arXiv:1604.03605

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science*, *22*(9), 1165–1172.

Cornelissen, T. H. W., & Võ, M. L. H. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception & Psychophysics*, *79*(1), 154–168.

de Groot, F., Huettig, F., & Olivers, C. N. L. (2016). When meaning matters: The temporal dynamics of semantic influences on visual attention. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(2), 180–196.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.

Greene, M. R., & Fei-Fei, L. (2014). Visual categorization is automatic and obligatory: Evidence from Stroop-like paradigm. *Journal of Vision*, *14*(1), 1–11.

Harel, J., Koch, C., & Perona, P. (2006). Graph-based Visual Saliency. In *Neural information processing systems* (pp. 1–8).

Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*, 219–222.

Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, *21*(1), 15–23.

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes revealed by meaning maps. *Nature Human Behaviour*, *1*, 743–747.

Henderson, J. M., & Hayes, T. R. (2018). Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *Journal of Vision*, *18*(6:10), 1–18.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, *16*, 850–856.

Henderson, J. M., Hayes, T. R., Rehrig, G., & Ferreira, F. (2018). Meaning guides attention during real-world scene description. *Scientific Reports*, *8*, 1–9.

Itti, L., & Borji, A. (2014). Computational models: Bottom-up and top-down aspects. In A. C. Nobre, & S. Kastner (Eds.), *The Oxford Handbook of Attention* (pp. 1122–1158). Oxford: Oxford University Press.

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

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259.

Kümmerer, M., Wallis, T. S. A., Gatys, L. A., & Bethge, M. (2017). Understanding low- and high-level contributions to fixation prediction. In *2017 IEEE international conference on computer vision* (pp. 4799–4808).

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics*, *2*(11), 547–552.

Malcolm, G. L., Rattinger, M., & Shomstein, S. (2016). Intrusive effects of semantic information on visual selective attention. *Attention, Perception, and Psychophysics*, *78*, 2066–2078.

O'Connel, T. P., & Walther, D. B. (2015). Dissociation of salience-driven and content-driven spatial attention to scene category with predictive decoding of gaze patterns. *Journal of Vision*, *15*(5), 1–13.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155 B*, 23–36.

Peacock, C. E., Hayes, T. R., & Henderson, J. M. (2019). Meaning guides attention during scene viewing even when it is irrelevant. *Attention, Perception, and Psychophysics*, *81*, 20–34.

SR Research. (2010). *EyeLink 1000 user's manual, version 1.5.2*. Mississauga: SR Research Ltd.

Vincent, B. T., Baddeley, R., Correani, A., Troscianko, T., & Leonards, U. (2009). Do we look at lights? Using mixture modeling to distinguish between low- and high-level factors in natural image viewing. *Visual Cognition*, *17*(6–7), 856–879.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*, 1395–1407.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*, 1–8.