Objects are Prioritized for Attention Based Upon Meaning During Active Scene Viewing

Elizabeth H. Hall[1,2], Candace E. Peacock[1,2], & John M. Henderson[1,2]

[1]Center for Mind and Brain

[2]Department of Psychology

University of California, Davis

Correspondence:
Candace Peacock
Center for Mind and Brain
267 Cousteau Place
University of California
Davis, CA 95618
cepeacock@ucdavis.edu

Word Count: 2052

Abstract

Although the physical salience of objects has previously been demonstrated to guide attention in real-world scene perception, it is unknown whether objects are also prioritized based on their meaning. To answer this question, we computed the average meaning and the average physical salience of objects in scenes. Using eye movement data from aesthetic judgment and memorization tasks, we then tested whether fixations are more likely to land on high-meaning objects than low-meaning objects while controlling for object salience. The results demonstrated that fixations are more likely to be directed to high meaning objects than low meaning objects regardless of object salience. Furthermore, the influence of object salience was progressively reduced as object meaning increased and was eliminated at the highest levels of meaning. Overall, these findings provide the first evidence that objects are prioritized by meaning for attentional selection during active scene viewing.

Keywords: scene perception, eye movements, meaning, physical salience, objects, attention

As we sample visual information from the world around us, the visual system must prioritize and select the most relevant information for analysis at any one moment. However, the process by which this occurs remains unclear. There are many factors that influence where we will look in real-world scenes, such as task (Castelhano et al., 2009; Cronin et al., 2020), scene context (Henderson et al., 1999, 2009; Loftus & Mackworth, 1978; Malcolm & Henderson, 2010), objects (Castelhano et al., 2009; Chen & Zelinsky, 2019; Cronin et al., 2020; Einhäuser et al., 2008; Nuthmann et al., 2020; Nuthmann & Henderson, 2010; Pajak & Nuthmann, 2013; Stoll et al., 2015; 't Hart et al., 2013), and meaning (Hayes & Henderson, 2019; Henderson et al., 2018, 2019; Henderson & Hayes, 2017, 2018a; Peacock et al., n.d., 2019b, 2019a, 2020; Rehrig, Peacock, et al., 2020; Rehrig, Hayes, et al., 2020). The present paper focuses on two of these sources, objects and meaning, which have been studied independently in prior work. It is unknown whether objects are prioritized by their meaning for attentional selection.

In basic attention studies, objects are defined as perceptual groupings of image features (Duncan, 1984; Treisman & Gelade, 1980; Wolfe et al., 1989). There are attentional costs when people shift their attention between objects relative to shifting attention within an object, suggesting that objects are important to attentional selection (Duncan, 1984; Egly et al., 1994). This basic finding has been replicated in real-world scenes (Malcolm & Shomstein, 2015). Despite the fact that objects are more difficult to define in scenes than in object arrays and therefore have been operationalized in various ways, they have consistently been shown to be important for attentional selection in scenes (Castelhano et al., 2009; Chen & Zelinsky, 2019; Cronin et al., 2020; Einhäuser et al., 2008; Nuthmann et al., 2020; Nuthmann & Henderson, 2010; Pajak & Nuthmann, 2013; Stoll et al., 2015; 't Hart et al., 2013).

Objects have been studied in conjunction with other sources of attentional guidance, such as task (Castelhano et al., 2009; Cronin et al., 2020), scene context (Henderson et al., 1999, 2009; Loftus & Mackworth, 1978; Malcolm & Henderson, 2010), and image salience (Chen & Zelinsky, 2019; Einhäuser et al., 2008; Nuthmann et al., 2020; Nuthmann & Henderson, 2010; Stoll et al., 2015; ’t Hart et al., 2013). It has been shown that fixation placement and fixation duration interact with objects as a function of task (Castelhano et al., 2009; Cronin et al., 2020), and that objects are processed differently when they are semantically inconsistent with their environments (Brockmole & Henderson, 2008; Cornelissen & Võ, 2017; De Graef et al., 1990; Henderson et al., 1999, 2009; Loftus & Mackworth, 1978; Malcolm & Henderson, 2010; Underwood & Foulsham, 2006; Võ & Henderson, 2009). Turning to image salience, it has been found that when object representations are contrasted with image salience, objects predict fixated regions above and beyond image salience (Einhäuser et al., 2008; Nuthmann & Henderson, 2010; Stoll et al., 2015; Underwood & Foulsham, 2006). Other studies have combined object and image salience representations, finding that combined object-salience models predict fixation placement better than image salience alone (Chen & Zelinsky, 2019; Stoll et al., 2015).

Recently, Nuthmann et al. (2020) combined objects and image salience and found that objects with greater salience were more likely to be fixated than objects with lower salience, suggesting that objects are prioritized for attention by their image salience. However, because image salience is correlated with scene meaning (Elazary & Itti, 2008; Henderson, 2003; Henderson et al., 2007; Henderson & Hayes, 2017, 2018b; Rehrig, Peacock, et al., 2020; Tatler et al., 2011), and given the finding that objects better predict attention than image salience (Einhäuser et al., 2008; Nuthmann & Henderson, 2010; Stoll et al., 2015; Underwood &

Foulsham, 2006), it could be that the apparent influence of image salience on object selection is actually due to meaning rather than to image salience.

Meaning maps represent the continuous spatial distribution of local semantic densities in scenes (Henderson & Hayes, 2017), allowing direct study of how semantics influence attention during scene viewing. Meaning maps are made by asking raters to rate the meaning of image patches, which decouples meaning from objects. The patch ratings are then combined to create a map of the spatial distribution of semantic content in the same format used for image saliency maps. Meaning maps, then, offer a useful tool in combination with object representations to test whether objects are prioritized by their meaning rather than their image salience.

Meaning has been demonstrated to predict eye movements significantly better than image salience across a variety of viewing tasks (Hayes & Henderson, 2019; Henderson et al., 2018, 2019; Henderson & Hayes, 2017, 2018a; Peacock et al., n.d., 2019b, 2019a, 2020; Rehrig, Hayes, et al., 2020; Rehrig, Peacock, et al., 2020). Therefore, it could be the case that objects are selected by their meaning for attention rather than by their physical salience. This result would be consistent with cognitive relevance theory which proposes that attention is "pushed" to scene regions or objects that are meaningful to the cognitive system rather than passively "pulled" by uninterpreted image features (Buswell, 1935; Hayhoe & Ballard, 2005; Henderson, 2003, 2017; Henderson et al., 1999, 2009; Tatler et al., 2011; Yarbus, 1967).

Current Work

The goal of the present work was to test the hypothesis that the visual system selects meaningful objects over objects that are not meaningful while controlling for the role of object salience. To test this hypothesis, we computed the average meaning of objects and the average physical salience of objects. We then tested whether people were more likely to fixate high-meaning objects than those that were low-meaning irrespective of object salience and whether these effects were influenced by the size and eccentricity of those objects.

Methods

Eye-tracking

Participants. One hundred fourteen experimentally naive University of California, Davis undergraduates with normal or corrected-to-normal vision were recruited from the UC Davis undergraduate subject pool. They received course credit in exchange for their participation. Fourteen participants' data were replaced due to poor eye tracking (25% or greater signal loss over all trials; Henderson and Hayes, 2017), leaving 100 participants' data available for analysis.

We have previously used this eye movement dataset to study general eye movement characteristics in scenes (Cronin et al., 2020) and to study how a vector-based model of semantics influences fixation probability (Hayes & Henderson, 2021). This is the first time that this dataset has been used to study object meaning.

Apparatus and Stimuli. Eye movements were monitored with a tower-mounted EyeLink 1000 eye tracker (spatial resolution 0.01° rms) sampling the right eye at 1000Hz (SR Research, 2010). Participants were seated 85 cm from a 21" CRT monitor. Participants' head movements were limited by a chin and forehead rest. One hundred luminance-matched images of real-world scenes were presented at their full resolution (1024 x 768 px), which filled the entire viewable area of the monitor (26.5 x 20 degrees of visual angle). The 100 scenes were chosen to represent 100 unique scene categories (e.g., kitchen, park), where half of the images were indoor scenes and half were outdoor scenes. The experimental stimuli were presented using the SR Research Experiment Builder software (SR Research, 2010).

Procedure. Participants viewed each of the 100 scenes for 12 seconds under one of two sets of task instructions. For 50 scenes they were told to memorize the images for a later memory test. For the other 50 scenes, they were asked to assess the aesthetic qualities of the image and, after the 12 second viewing period, responded whether they liked, felt neutral about, or disliked the image. This response was recorded by a RESPONSEPixx Handheld button box (VPixx Technologies). Task instruction order was counterbalanced across subjects and scenes such that all subjects viewed all 100 images and each of the 100 images appeared equally under the two viewing task conditions across all subjects.

Data preparation. Eye movement data were imported into MATLAB using the EDFConverter tool. Participants' eye tracking data was first assessed for missing data: any participant with track-loss of greater than 25% was removed from further analysis and replaced with a new subject. Participants who met this criterion were then assessed at the trial level: any trial in which participants' eyes were tracked for less than 75% of the duration of the trial were also excluded from further analyses. This resulted in a loss of 1.2% of experimental trials. To be consistent with previous work, we trimmed fixation durations that lasted less than 50ms or more than 1500ms from our analyses. This trim resulted in a loss of 2.3% of fixations.

Object Meaning

Meaning maps. We used the meaning map technique developed by Henderson and Hayes (2017) (see https://osf.io/654uh/ for code and instructions) to compute object meaning. To create meaning maps, scene-patch ratings were performed by 413 participants on Amazon Mechanical Turk. Participants were recruited from the United States, had a hit approval rate of 99% and 500 hits approved, and were allowed to participate in the study only once. Participants were paid $0.50 per assignment, and all participants provided informed consent. Rating stimuli were the same 100 digitized (1,024 × 768 pixels) photographs of real-world scenes used for the aesthetic judgment and memorization tasks. Each scene was decomposed into a series of partially overlapping (tiled) circular patches at two spatial scales. The full patch stimulus set consisted of 30,000 unique fine patches (87-pixel diameter) and 10,800 unique coarse patches (205-pixel diameter), for a total of 40,800 scene patches. The optimal meaning-map grid density for each patch size was previously determined by simulating the recovery of known image properties as reported in Henderson & Hayes (2018).

Each participant rated 300 random patches extracted from 100 scenes. Participants were instructed to assess the meaningfulness of each patch based on how informative or recognizable it was. They were first given examples of two low-meaning and two high-meaning scene patches, to make sure they understood the rating task, and then they rated the meaningfulness of scene patches on a 6-point Likert scale (very low, low, somewhat low, somewhat high, high, very high). Patches were presented in random order and without scene context, so ratings were based on context-free judgments. Each unique patch was rated three times by independent raters for a total of 128,520 ratings. However, due to the large degree of overlap across patches, each patch contained rating information from 27 independent raters for each fine patch and 63 independent raters for each coarse patch. The ratings for each pixel at each scale in each scene were averaged, producing an average fine and coarse rating map for each scene. The average fine and course rating maps were then combined into a single map using the simple average and a light Gaussian filter was applied using the MATLAB function ‘imgaussfilt.m’ set at 10.

Object Meaning. Scenes were segmented and non-occluding objects were identified by an undergraduate research assistant in each scene (Castelhano et al., 2009; Cronin et al., 2020; Nuthmann et al., 2017; Nuthmann & Henderson, 2010). From there, we computed the size of each segmentation and removed segmentations that were greater than the third quartile of object sizes (third quartile = 13,369 pixels) in order to isolate objects that were of small to medium size, consistent with Nuthmann et al. (2020). Each object was given a weight of the average meaning value of the segmentation and added to a 1024 x 768 array of zeros to create object meaning maps (Figure 1).

Object Salience

Saliency maps. Following Nuthmann et al. (2020), saliency maps were computed using the adaptive white saliency (AWS) model (Garcia-Diaz, Fdez-Vidal, et al., 2012; Garcia-Diaz, Leborán, et al., 2012). The saliency maps from the AWS model were generated using the MATLAB code provided by the authors (http://persoal.citius.usc.es/xose.vidal/research/aws/AWSmodel.html). Parameters were kept at the authors' default values.

Object Salience. Each object was given a weight of its average salience value and added to a 1024 x 768 array of zeros to create object salience maps (Figure 1).

Object Eccentricity

Object eccentricity served as a global representation of how close each object in the scene was from scene center. To compute object eccentricity, we first generated a center proximity map which represented the inverted Euclidean distance from the center pixel of the scene to all other pixels in the scene image. Each object was given a weight of its average center proximity and added to a 1024 x 768 array of zeros to create object eccentricity maps (Figure 1). The object eccentricity measure was used in the mixed-effects models described below to account for and control the role of center bias, the tendency to fixate centrally (Bindemann, 2010; Hayes & Henderson, 2021; Tatler, 2007; Tseng et al., 2009).

Object Size

To measure object size, we computed the area of each object. Each object was then given a weight of its area in pixels and added to a 1024 x 768 array of zeros (Figure 1). The object size measure was included as a predictor in the mixed-effects models described below to account for and control the role of object size, which might influence whether objects are fixated or not, beyond the role of object meaning (Nuthmann et al., 2020).

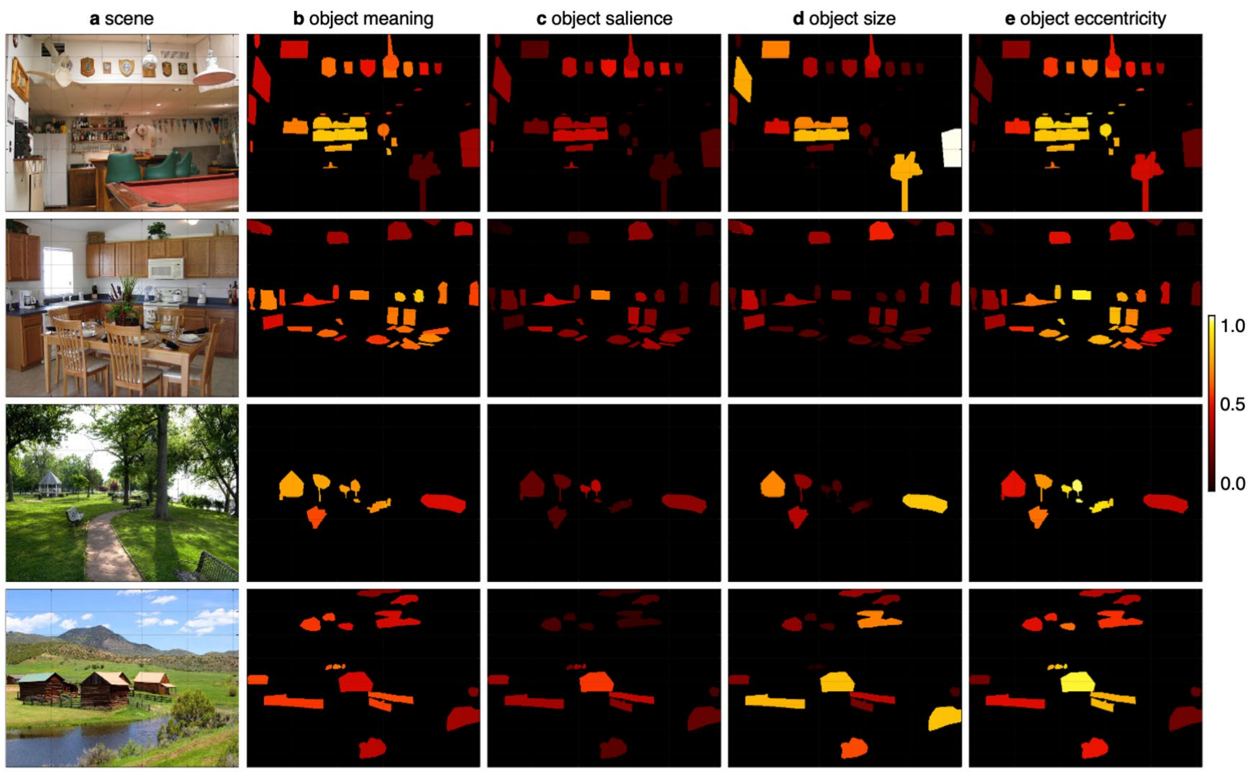

Figure 1. *Object representations.* The figure visualizes an example of each object representation for a few example indoor and outdoor scenes. Each column represents an example scene (a) with each respective object meaning (b) object salience (c), object size (d), and object eccentricity representation (e).

Eyetracking Analysis

To understand whether people are more likely to fixate objects with greater over those with less meaning irrespective of image salience, while also taking object eccentricity, object size, scene-by-scene variation, and participant-by-participant variation into account, we used a general linear mixed effects (GLME) model with the link logit ('binomial') distribution (Hayes & Henderson, 2021; Nuthmann et al., 2017, 2020). Before submitting the data to the GLME, we z-normalized object meaning, object salience, object eccentricity, and object size to a common scale.

For each participant and scene, we computed the average meaning, salience, eccentricity, and size value of each fixated object. Any fixations that did not land on the objects selected were excluded from analysis. To represent scene features that were not associated with overt attention, we randomly sampled an equal number of objects that were not fixated for that participant in that scene. In total, there was an average of 36.78 fixated and non-fixated objects per scene for each participant (*SD* = 14.51). There were 355,862 total observations (i.e., fixated and non-fixated objects) across the entire dataset.

The dependent variable was whether an object was fixated or not. The fixed effects were the object meaning values, the object salience values, the object size values, and the object eccentricity values. The primary effect of interest was the interaction between object meaning and object salience to test whether objects are selected by their meaning irrespective of object salience. However, we also modeled the four-way interaction between object meaning x object salience x object eccentricity x object size in order to account for the roles of object size and eccentricity. Additionally, we included a random intercept of scene and participant as random effects.

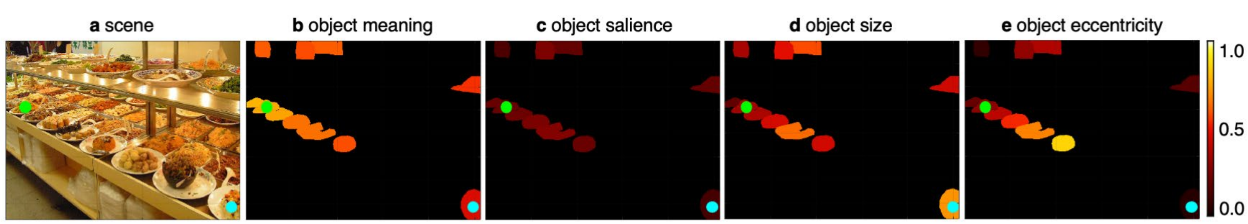

Figure 2. *Analysis and predictions.* The figure shows an example scene (a), object meaning (b), object salience (c), object size (d), and object eccentricity (e) with hypothetical fixated (green) versus non-fixated (cyan) objects.

Results

The goal of the present study was to test whether objects are selected for fixation based upon meaning irrespective of image salience while taking the roles of object size and object eccentricity into account. Using a GLME, we modeled the four-way interaction between object meaning x object salience x object size x object eccentricity. We hypothesized that objects are selected on the basis of meaning beyond the contributions of object salience and that this effect would not be due to object eccentricity or size. The GLME model results are summarized in Figure 3 and Table 1. The four-way interaction was significant ($\beta = 0.02$, CI = [0.01, 0.04], $Z = 3.84$, $p < 0.001$). The two-way interaction of interest between meaning x image salience was also significant ($\beta = -0.10$, CI = [-0.11, -0.09], $Z = -16.50$, $p < 0.001$).

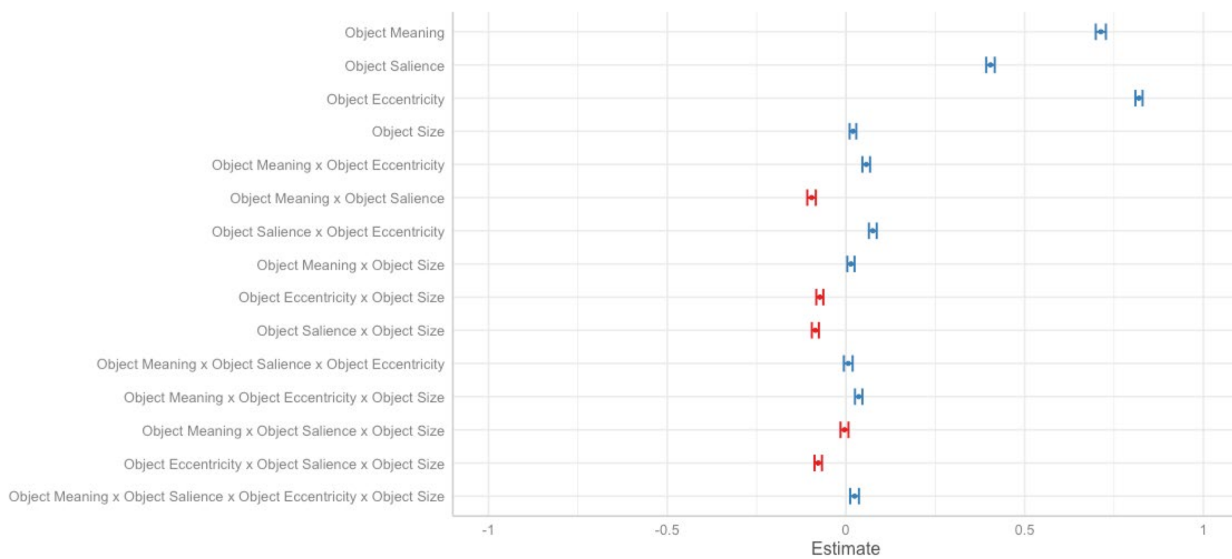

Figure 3. *Model fits.* The estimates for the model are visualized. An estimate of 0 indicates that neither positive nor negative values of a predictor are likely to occur with fixated objects. An estimate of greater than 0 (blue) indicates that positive values of a predictor are more associated with fixated objects whereas an estimate of less than 0 (red) indicates that negative values of a predictor are associated with fixated objects. Error bars reflect 95% confidence intervals.

Table 1. *Object meaning x object salience x object size x object eccentricity GLME results.* Beta estimates (β), 95% confidence intervals (CI), standard errors (SE), z-values, and p-values (p) for each fixed effect and standard deviations (SD) for the scene random effect.

	Fixed Effects					Random Effects, SD	
Predictors	Estimate	95% CI	SE	Z	p	By Scene	By Subject
Intercept	0.12	[0.002, 0.24]	0.06	1.99	0.05	0.60	0.11
Object Meaning	0.72	[0.70, 0.73]	0.007	99.57	<0.001		
Object Salience	0.41	[0.39, 0.42]	0.006	66.95	<0.001		
Object Eccentricity	0.81	[0.80, 0.82]	0.005	159.47	<0.001		
Object Size	0.02	[0.01, 0.03]	0.005	4.65	<0.001		
Object Meaning x Object Eccentricity	0.06	[0.05, 0.07]	0.006	10.26	<0.001		
Object Meaning x Object Salience	-0.10	[-0.11, -0.09]	0.006	-16.50	<0.001		
Object Salience x Object Eccentricity	0.08	[0.07, 0.09]	0.006	13.99	<0.001		
Object Meaning x Object Size	0.02	[0.01, 0.03]	0.005	2.87	0.004		
Object Eccentricity x Object Size	-0.07	[-0.08, -0.06]	0.005	-14.32	<0.001		
Object Salience x Object Size	-0.09	[-0.10, -0.08]	0.005	-17.26	<0.001		
Object Meaning x Object Salience x Object Eccentricity	0.01	[0.0002, 0.03]	0.006	1.98	0.05		
Object Meaning x Object Eccentricity x Object Size	0.03	[0.02, 0.05]	0.005	6.54	<0.001		
Object Meaning x Object Salience x Object Size	-0.01	[-0.02, 0.01]	0.006	-0.99	0.32		
Object Salience x Object Eccentricity x Object Size	-0.08	[-0.09, -0.07]	0.005	-14.88	<0.001		
Object Meaning x Object Salience x Object Eccentricity x Object Size	0.02	[0.01, 0.04]	0.006	3.84	<0.001		

Since four-way interactions are difficult to interpret, we visualized the two three-way interactions between object meaning x object salience x object size, and object meaning x object salience x object eccentricity, to investigate how the two-way interaction of interest (object meaning x object salience) was influenced by object size or object eccentricity. The object meaning x object salience x object size interaction was not significant (β = -0.01, CI = [-0.02, 0.01], Z = -0.99, p = 0.32) and the object meaning x object salience x object eccentricity interaction was marginally significant (β = 0.01, CI = [0.0002, 0.03], Z = 1.98, p = 0.05). Correspondingly, Figure 4 demonstrates that regardless of object size, fixations were more likely to be directed to high meaning objects than low meaning objects across all levels of object salience. Figure 4 also shows that the same general trend existed across all levels of object eccentricity (i.e., fixations were directed to high meaning over low meaning objects across all levels of object salience) but that the interaction between object meaning x object salience was marginally decreased when objects were in scene peripheries.

Furthermore, irrespective of object eccentricity or size, the influence of object salience was progressively reduced as object meaning increased and was eliminated at the highest levels of meaning. In other words, when object meaning was low, object salience had its largest effect, but the overall probability of fixation was low. As meaning increased, there was a dramatic increase in the probability of fixation, and the influence of salience was reduced. These results suggest that salience influenced object prioritization only when object meaning was low, and even at low levels of meaning, the effect of salience was relatively small compared to the effect of meaning. These effects did not change as a function of object size or eccentricity.

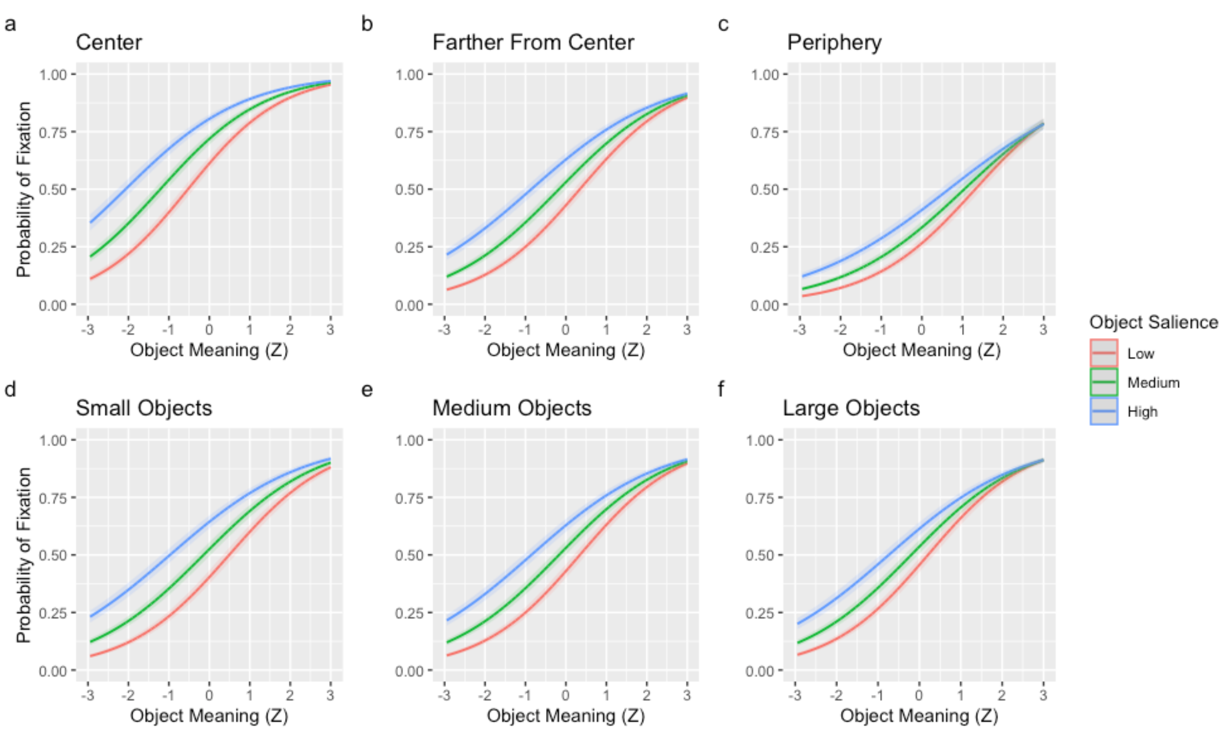

Figure 4. *Visualization of whether object salience modulates whether meaningful objects are fixated as a function of object eccentricity and size.* This figure visualizes whether object salience modulates whether meaningful objects are fixated at scene centers (a), farther from center (b), and for peripheral objects (c). This is also visualized for small objects (d), medium-sized objects (e), and large objects (f). Error bands reflect 95% confidence intervals.

Discussion

Recent work found that objects were selected for fixation based upon their physical salience (Nuthmann et al., 2020). However, given the correlation between image salience and meaning (Elazary & Itti, 2008; Henderson, 2003; Henderson et al., 2007; Henderson & Hayes, 2017, 2018b; Rehrig, Peacock, et al., 2020; Tatler et al., 2011), the apparent influence of salience on object selection could instead be due to meaning. The goal of the present work was to test whether objects are selected by meaning for fixation controlling for object salience.

To investigate this hypothesis, we computed the average meaning of objects in scenes using the meaning mapping method from Henderson & Hayes (2017), and the average salience of objects in scenes using AWS (Garcia-Diaz, Fdez-Vidal, et al., 2012; Garcia-Diaz, Leborán, et al., 2012). The results demonstrated that regardless of object salience, people were more likely to fixate meaningful objects over those that were not meaningful. Furthermore, the role of object salience progressively declined as the meaning of objects increased, and the influence of salience was eliminated when object meaning was at its highest. Finally, even when object meaning was at its lowest, the role of object salience was relatively modest. Overall, these findings suggest that the visual system selects objects for fixation primarily based upon their meaning. This finding is consistent with cognitive relevance theory, which suggests that factors that are relevant to the cognitive system (e.g., objects and meaning), will be prioritized for analysis over those that are irrelevant (e.g., physical salience) (Buswell, 1935; Hayhoe & Ballard, 2005; Henderson, 2003, 2017; Henderson et al., 2009; Henderson & Hollingworth, 1999; Tatler et al., 2011; Yarbus, 1967).

Because meaning can be defined in different ways, there are likely other ways that meaning acts through objects to determine fixation priority. In previous work, we found that objects with higher similarity in semantic vector space to other objects were more likely to be fixated than objects with lower vector similarity (Hayes & Henderson, 2021). The present findings converge with the Hayes & Henderson (2021) result, and together the results suggest that objects are selected for fixation based upon their informativeness and recognizability and their meaning with respect to the other objects in the scene.

Nuthmann et al. (2020) did not test whether objects were selected for fixation on the basis of meaning. That study was partly focused on whether older and younger adults would select objects for fixation similarly, and they argued that they could not examine the role of meaning in this process because it was unclear whether older adults would rate meaning in the same way as younger adults. Recently, we found that older adults were less likely to fixate scene meaning than younger adults when meaning was rated by younger adults (Rehrig et al., preprint). Therefore, it could be that older adults indeed assign scene meaning differently than younger adults. In future work testing whether younger and older adults differ in how object meaning influences fixations, it may be necessary to obtain separate meaning ratings for older adults and younger adults.

We also recently defined meaning based upon the grasping affordances of entities in scene regions (Rehrig, Peacock, et al., 2020). Here, we found that graspability and meaning equally predicted attention in scenes with reachable objects. It could be the case, then, that objects are selected for attention based upon their grasping affordances and based upon how informative and recognizable they are. Understanding how object meaning interacts with other object properties (such as object graspability) during attentional selection will be important avenues for future work.

Conclusion

The present study provided the first insight into whether objects are selected based upon their meaning for fixation. We found that the visual system prioritizes meaningful objects for attention over those that are not meaningful irrespective of object salience, a finding that is consistent with cognitive relevance theory.

Funding

This research was supported by the National Eye Institute of the National Institutes of Health (grant 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. This project also received support from the UC Davis Vision Science Training Program (T32 EY015387).

Declaration of Competing Interest

The authors declare no competing financial interests. Additionally, these data have not previously been published.

Open Practices Statement

The meaning mapping code and materials are available at https://osf.io/654uh/. The eye-tracking data are not currently available as the authors are mining the dataset. The experiment was not preregistered.

Acknowledgements

We thank Michelle Galat, Erika Agbabian, Aimee Nguyen, Mehvish Javed, Nat Finnegan, and Viviana Padilla for assisting in data collection. We would also like to thank Anthony Lagunda for annotating the scenes used in the present study.

References

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, *50*, 2577–2587. https://doi.org/10.1016/j.visres.2010.08.016

Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-world scenes: Effects of object–scene consistency. *Visual Cognition*, *16*(2–3), 375–390. https://doi.org/10.1080/13506280701453623

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art* (U. of C. Press, Ed.).

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3), 6–6. https://doi.org/10.1167/9.3.6

Chen, Y., & Zelinsky, G. J. (2019). Is there a shape to the attention spotlight? Computing saliency over proto-objects predicts fixations during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(1), 139–154. https://doi.org/10.1037/xhp0000593

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. https://doi.org/10.3758/s13414-016-1203-7

Cronin, D. A., Hall, E. H., Goold, J. E., Hayes, T. R., & Henderson, J. M. (2020). Eye Movements in Real-World Scene Photographs: General Characteristics and Effects of Viewing Task. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02915

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*(4), 317–329. https://doi.org/10.1007/BF00868064

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161–177.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18–18. https://doi.org/10.1167/8.14.18

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8*(3), 3–3. https://doi.org/10.1167/8.3.3

Garcia-Diaz, A., Fdez-Vidal, X. R., Pardo, X. M., & Dosil, R. (2012). Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image and Vision Computing*, *30*(1), 51–64. https://doi.org/10.1016/j.imavis.2011.11.007

Garcia-Diaz, A., Leborán, V., Fdez-Vidal, X. R., & Pardo, X. M. (2012). On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. *Journal of Vision*, *12*(6), 17–17. https://doi.org/10.1167/12.6.17

Hayes, T. R., & Henderson, J. M. (2019). Scene semantics involuntarily guide attention during visual search. *Psychonomic Bulletin and Review*.

Hayes, T. R., & Henderson, J. M. (2021). Looking for semantic similarity: What a vector space model of semantics can tell us about attention in real-world scenes. *Psychological Science*.

Hayhoe, M. M., & Ballard, D. H. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. https://doi.org/10.1016/j.tics.2003.09.006

Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, *21*(1), 15–23. https://doi.org/10.1016/j.tics.2016.11.003

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. L. (2007). Visual saliency does not account for eye movements during visual search in real world scenes. In R. P. G. V. Gompel, M. H. Fischer, S. Murray, & Wayne (Eds.), *Eye Movements: A Window on Mind and Brain* (pp. 537–562). Elsevier Ltd. https://doi.org/10.1167/9.3.6

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, *1*, 743–747. https://doi.org/10.1038/s41562-017-0208-0

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

Henderson, J. M., & Hayes, T. R. (2018b). Meaning guides attention in real-world scenes: Evidence from eye movements and meaning maps. *Journal of Vision*, *18*(6), 1–18. https://doi.org/10.1089/jmf.2012.0243

Henderson, J. M., Hayes, T. R., Peacock, C. E., & Rehrig, G. (2019). Meaning and Attentional Guidance in Scenes: A Review of the Meaning Map Approach. *Vision*, *3*(2), 19. https://doi.org/10.3390/vision3020019

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

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

Henderson, J. M., Malcolm, G. L., & Schandl, Charles. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin and Review*, *16*(5), 850–856. https://doi.org/10.3758/PBR.16.5.850

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228. https://doi.org/10.1037/0096-1523.25.1.210

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

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, *10*(2), 4–4. https://doi.org/10.1167/10.2.4

Malcolm, G. L., & Shomstein, S. (2015). Object-based attention in real-world scenes. *Journal of Experimental Psychology: General*, *144*(2), 257–257.

Nuthmann, A., Einhäuser, W., & Schütz, I. (2017). How Well Can Saliency Models Predict Fixation Selection in Scenes Beyond Central Bias? A New Approach to Model Evaluation Using Generalized Linear Mixed Models. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00491

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8), 20–20. https://doi.org/10.1167/10.8.20

Nuthmann, A., Schütz, I., & Einhäuser, W. (2020). Salience-based object prioritization during active viewing of naturalistic scenes in young and older adults. *Scientific Reports*, *10*(1), 22057. https://doi.org/10.1038/s41598-020-78203-7

Pajak, M., & Nuthmann, A. (2013). Object-based saccadic selection during scene perception: Evidence from viewing position effects. *Journal of Vision*, *13*(5), 2–2. https://doi.org/10.1167/13.5.2

Peacock, C. E., Hayes, T. R., & Henderson, J. M. (2019a). Meaning guides attention during scene viewing even when it is irrelevant. *Attention, Perception, and Psychophysics*, *81*(1), 20–34. https://doi.org/10.3758/s13414-018-1607-7

Peacock, C. E., Hayes, T. R., & Henderson, J. M. (2019b). The role of meaning in attentional guidance during free viewing of real-world scenes. *Acta Psychologica*.

Peacock, C. E., Hayes, T. R., & Henderson, J. M. (2020). Center Bias Does Not Account for the Advantage of Meaning Over Salience in Attentional Guidance During Scene Viewing. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01877

Peacock, C. E., Singh, P., Hayes, T. R., & Henderson, J. M. (n.d.). *Meaning guides attention during visual search in real-world scenes*.

Rehrig, G., Hayes, T. R., Henderson, J. M., & Ferreira, F. (2020). When scenes speak louder than words: Verbal encoding does not mediate the relationship between scene meaning and visual attention. *Memory & Cognition*. https://doi.org/10.3758/s13421-020-01050-4

Rehrig, G., Peacock, C. E., Hayes, T. R., Henderson, J. M., & Ferreira, F. (2020). Where the action could be: Speakers look at graspable objects and meaningful scene regions when describing potential actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. http://dx.doi.org/10.1037/xlm0000837

Stoll, J., Thrun, M., Nuthmann, A., & Einhäuser, W. (2015). Overt attention in natural scenes: Objects dominate features. *Vision Research*, *107*, 36–48. https://doi.org/10.1016/j.visres.2014.11.006

’t Hart, B. M., Schmidt, H. C. E. F., Roth, C., & Einhauser, W. (2013). Fixations on objects in natural scenes: Dissociating importance from salience. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00455

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), 4–4.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5–5.

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

Tseng, P.-H., 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*(4). https://doi.org/10.1167/9.7.4

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, *59*(11), 1931–1949. https://doi.org/10.1080/17470210500416342

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision2*, *9*(3), 1–15. https://doi.org/10.1167/9.3.24

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 419–433.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye Movements and Vision* (pp. 171–211). Springer. https://doi.org/10.1007/978-1-4899-5379-7_8