vision

MDPI

Review

Meaning and Attentional Guidance in Scenes: A Review of the Meaning Map Approach

John M. Henderson [1,2,*], Taylor R. Hayes [1], Candace E. Peacock [1,2] and Gwendolyn Rehrig [2]

[1] Center for Mind and Brain, 267 Cousteau Place, University of California, Davis, CA 95618, USA; trhayes@ucdavis.edu (T.R.H.); cepeacock@ucdavis.edu (C.E.P.)
[2] Department of Psychology, University of California, Davis, CA 95618, USA; glrehrig@ucdavis.edu
* Correspondence: johnhenderson@ucdavis.edu; Tel.: +1-530-754-455

Received: 27 February 2019; Accepted: 7 May 2019; Published: 10 May 2019

Abstract: Perception of a complex visual scene requires that important regions be prioritized and attentionally selected for processing. What is the basis for this selection? Although much research has focused on image salience as an important factor guiding attention, relatively little work has focused on semantic salience. To address this imbalance, we have recently developed a new method for measuring, representing, and evaluating the role of meaning in scenes. In this method, the spatial distribution of semantic features in a scene is represented as a meaning map. Meaning maps are generated from crowd-sourced responses given by naïve subjects who rate the meaningfulness of a large number of scene patches drawn from each scene. Meaning maps are coded in the same format as traditional image saliency maps, and therefore both types of maps can be directly evaluated against each other and against maps of the spatial distribution of attention derived from viewers' eye fixations. In this review we describe our work focusing on comparing the influences of meaning and image salience on attentional guidance in real-world scenes across a variety of viewing tasks that we have investigated, including memorization, aesthetic judgment, scene description, and saliency search and judgment. Overall, we have found that both meaning and salience predict the spatial distribution of attention in a scene, but that when the correlation between meaning and salience is statistically controlled, only meaning uniquely accounts for variance in attention.

Keywords: attention; scene perception; eye movements

1. Introduction

The world contains an enormous amount of visual information, but human vision and visual cognition are severely limited in their processing capacity: Only a small fraction of the latent information can be analyzed at any given moment. Efficient visual cognition therefore requires selecting the information that is most relevant at the present moment for understanding and acting on the world. The primary way in which this selection takes place in natural vision is with overt attention via eye movements [1–9], as shown in Figure 1. Close or direct fixation of the scene region containing relevant information is typically necessary to perceive its visual details, to unambiguously determine its identity and meaning, and to encode it into short- and long-term memory. That is, what we see and understand about the world is determined by where we look [10].

Given the importance of eye movements for perception and cognition, a critical issue concerns understanding the representations and processes that guide the eyes through a visual scene in real time to support perception, cognition, and behavior [11]. Models based on image salience have provided an influential approach to eye movement guidance in scene perception. For static images, these models propose that attention is controlled by contrasts in primitive image features such as luminance, color, and edge orientation [12–14]. A key concept is the saliency map, which is generated by salience over

the primitive features. Attention is then assumed to be captured or "pulled" to the most visually salient scene regions represented by the saliency map [15–21]. Because the primitive features are semantically uninterpreted, scene regions are prioritized for attentional selection based on image properties alone. The appeal of this type of image salience model is that visual salience is both neurobiologically plausible in the sense that the visual system is known to compute the assumed primitive features, and computationally tractable in the sense that working models have been implemented that generate image salience from these features [4].

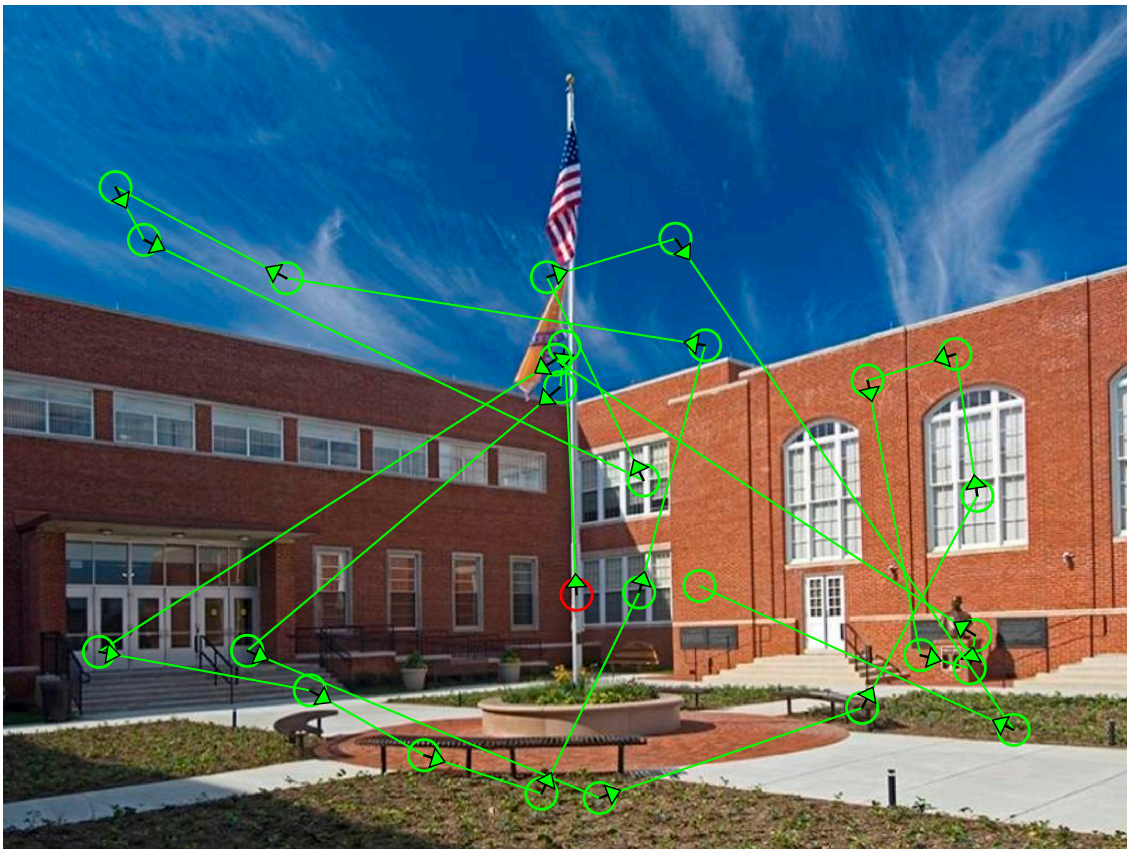

Figure 1. Scan pattern of a single viewer freely viewing a real-world scene.

We note that "salience" has different interpretations in different literatures, and so we want to be clear about which interpretation we are using. In vision science and attention research, the term is typically reserved for models in the Koch and Ullman tradition based on the idea that the human visual system computes difference maps from basic image features that are then combined and used to guide attention [19]. On the other hand, in computer vision and image processing, models that predict attention regardless of the underlying processes are also sometimes referred to as saliency models. This difference in usage can lead to confusion. For our purposes here, we specifically focus on image salience in the Koch and Ullman tradition, and to eliminate ambiguity, we use the terms "image salience" and "salience" to refer to that concept.

In contrast to models based on image salience, cognitive guidance models emphasize the important role of scene semantics in directing attention in scenes. In this view, attention is "pushed" by the cognitive system to scene regions that are semantically informative and cognitively relevant [10]. Cognitive guidance is consistent with evidence suggesting that viewers attend to semantically informative regions of a scene [5,6,22–25], as well as scene regions that are task-relevant [6,26–34]. For example, according to the cognitive relevance model [35,36], the attentional priority assigned to a scene region is based on its inherent meaning (e.g., "cup") as well as its meaning with respect to the scene (e.g., "a cup in an office") and the goals of the viewer (e.g., "I am looking for something

I can put water into"). In the cognitive relevance model, the scene image is of course important: It is the input that enables these higher-level semantic processes and interpretations. Additionally, the scene image is the input for forming a map of potential attention targets. However, for this model, the critical hypothesis is that the image-based parse generates a "flat" (that is, unranked) landscape of potential targets rather than a landscape ranked by image salience. Attentional priority is then based on the detected and predicted informativeness (i.e., meaning) of the scene regions and objects in that landscape [3,4].

Until recently, it has been difficult to compare directly the influence of image salience and meaning on attentional guidance in scenes, because to do so requires representing both of them in a format that allows for comparable quantitative predictions of attentional distribution over the scene. Saliency map models naturally provide this type of prediction [17,18,20,21,30,37]. Unfortunately, it is far more difficult to create a computational model of meaning than it is to create a computational model of image salience, a likely reason that saliency models have been so popular [4,10]. Given this difficulty, studies of meaning-based models of attention in scenes have typically focused on manipulations of one or at most a small number of specific scene regions or objects [22,38–41]. However, these types of manipulations do not allow a direct comparison of image salience and semantic informativeness across the entire scene.

The issue we have recently pursued, then, is this: How can we generate and represent the spatial distribution of semantic informativeness over a scene in a format that supports direct comparison with a saliency map? To address this issue, we introduced a new approach based on meaning maps [42]. Meaning maps were inspired by two classic scene viewing studies [23,24]. The central idea of a meaning map is that it represents the spatial distribution of semantic informativeness over a scene in the same format as a saliency map represents the spatial distribution of image salience. To create meaning maps, we use crowd-sourced responses given by large numbers of naïve subjects who rate the meaningfulness of scene patches. Specifically, photographs of real environments are divided into dense arrays of objectively defined circular overlapping patches at two spatial scales (Figure 2). The two scales and numbers of patches are chosen based on simulations showing that we can recover ground truth visual properties of scenes from them [42]. Large numbers of workers on Mechanical Turk each rate a randomly selected subset of individually presented patches taken from the set of scenes to be rated. We then construct meaning maps for each scene by averaging these ratings by pixel over patches and raters and smoothing the results (Figure 3). Like image salience, meaning is spatially distributed non-uniformly across scenes, with some scene regions relatively rich in semantic informativeness and other regions relatively sparse. Meaning maps represent this spatial distribution pixel by pixel, and so offer a foundation for directly comparing the relative roles of meaning and image salience on attentional guidance.

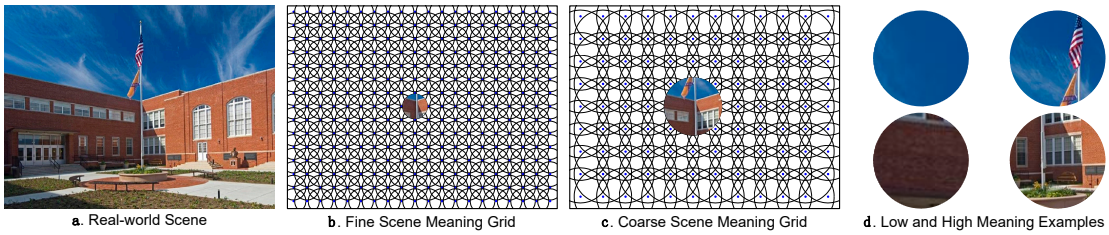

Figure 2. (**a**) A real-world scene; (**b**) fine scale, and (**c**) coarse scale patches from the patch grids; (**d**) examples of patches rated low (left column) and high (right column) in meaning.

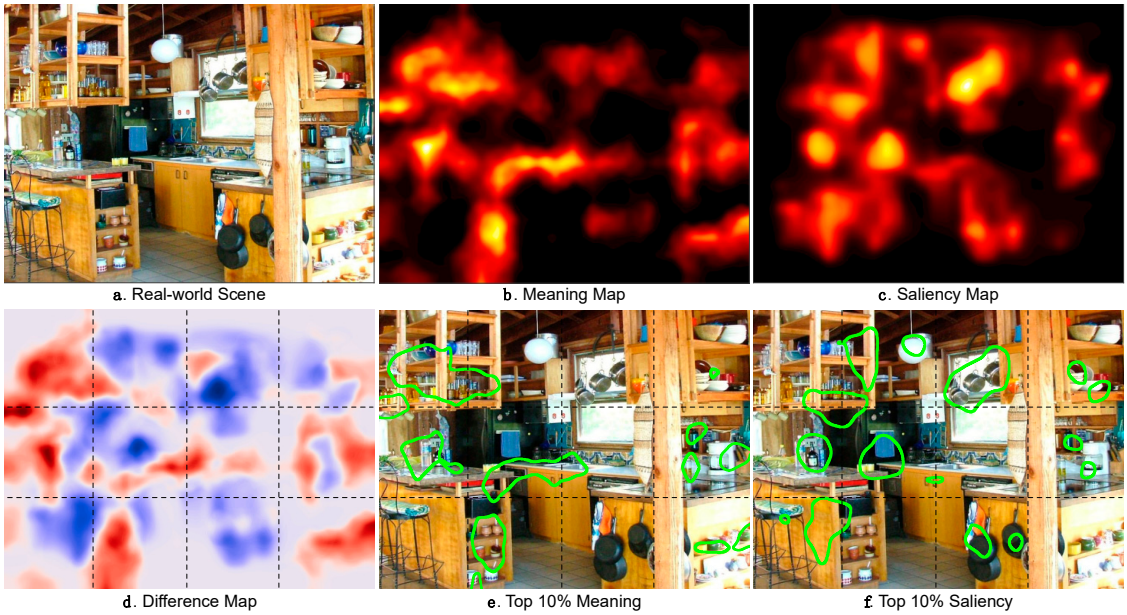

Figure 3. (**a**) Real-world scene; (**b**) scene's meaning map; (**c**) saliency map; (**d**) difference map showing regions that are more meaningful (red) and salient (blue); (**e**) regions in the top 10% of meaning values; (**f**) regions in the top 10% of saliency values.

In sum, meaning maps provide a conceptual analog of saliency maps by representing the spatial distribution of semantic features associated with informativeness across a scene. Meaning maps generate predictions concerning attentional guidance that can be tested using the same methods that have been used to test predictions from saliency maps. This allows contrasting predictions from semantic informativeness and image salience to be directly compared [42].

2. Review of Results

Meaning maps and saliency maps serve as predictions of how attention will be guided through scenes. The key empirical question is how well these predictions accord with observed distributions of attention produced by people viewing scenes.

In a first study directed toward this question, we asked subjects to view a set of scenes for 12 s each while their eye movements were recorded [42]. Subjects viewed the scenes to prepare for memory and aesthetic judgment questions that were presented after viewing. We operationalized attention maps as fixation density maps that reflected the spatial distribution of eye fixations across the scene. Importantly, the attention maps represented attention in the same format and on the same scale as the meaning and saliency maps. We then contrasted the degree to which the spatial distribution of meaning and salience predicted viewers' attention over the scenes. The results showed that both meaning and salience predicted attention, but that when the association between meaning and salience was statistically controlled with semi-partial correlations, only meaning uniquely accounted for attentional variance. Furthermore, the influence of meaning was observed both at the very beginning of scene viewing and throughout the trial. Figure 4 summarizes the data from this study. This result was observed in both scene memorization and aesthetic judgment viewing tasks. Given the strong observed correlation between meaning and salience, and the finding that only meaning accounted for variance in attention when that correlation was controlled, we concluded that the existing data are consistent with a theory in which meaning is the main factor guiding attention through scenes.

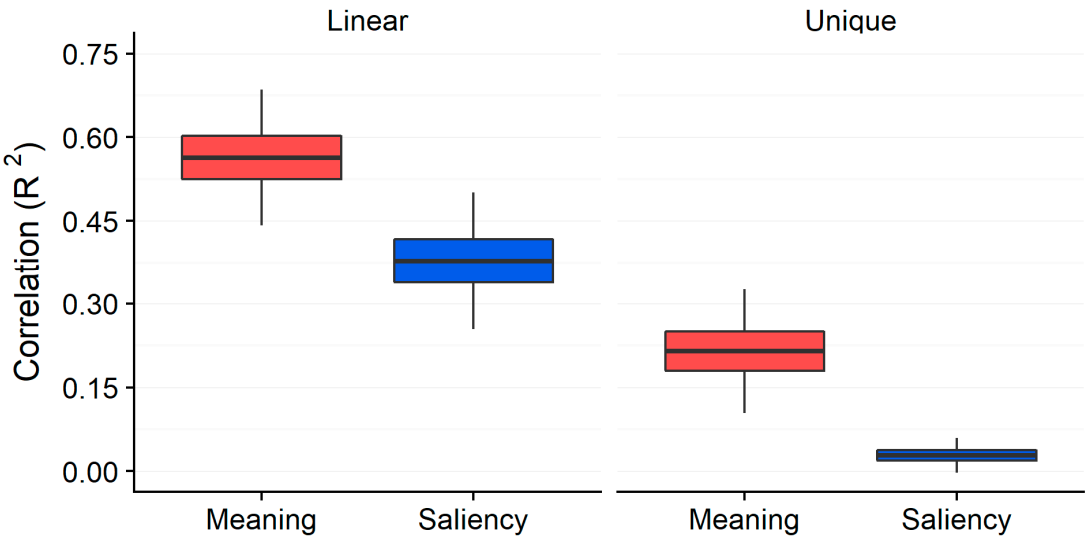

Figure 4. Squared linear correlation and semi-partial (unique) correlation for meaning and image salience across 40 scenes from a scene memorization task in Henderson and Hayes (2017) [42]. Box plots show the grand mean (black horizontal line), 95% confidence intervals (colored box), and one standard deviation (black vertical line).

In our initial study, we focused on the spatial distribution of attention by measuring the locations of fixations [42]. However, fixations also differ by duration, and fixation duration is known to reflect ongoing visual and cognitive activity [43–49]. Therefore, in a reanalysis of the data from Henderson and Hayes (2017), we examined the relative influences of semantic features and image salience on the distribution of attention taking into account attentional dwell time at each location [42,50]. This was accomplished by creating attention maps that weighted each fixation by its duration. Again, the question was whether these duration-weighted attention maps would be best accounted for by meaning maps or saliency maps. Using the same analysis methods with these duration-weighted attention maps, we replicated all of the basic data patterns that we observed in the original study. Specifically, we found that both meaning and salience were associated with attention, but that when the correlation between meaning and salience was statistically controlled, only meaning accounted for variance in attention. Once again, the influence of meaning was observed both at the beginning of scene viewing and throughout the trial. Therefore, whether the measure of attention is based on location only or includes dwell time, the answer with respect to meaning and image salience is the same [50].

In our initial experiments demonstrating the advantage of meaning maps over saliency maps, subjects viewed scenes in order to prepare for memory and aesthetic judgment questions that were presented after viewing [42,50]. In those tasks, the responses were off-line with respect to scene perception, so viewers may not have guided attention as quickly or under as much control as they might in a task requiring real-time responses during viewing. Therefore, in the next study we investigated how well meaning and image salience account for attention when the subject is actively engaged with and responding to the scene continuously in real time, and the guidance of attention is directly relevant to the real-time task. For this purpose, we used two scene description tasks [51].

We drew on evidence that language production is incremental in the sense that speakers interweave planning and speaking instead of planning the entire production and then executing the plan [52]. That is, speakers typically plan and produce small units of speech (words and phrases) that are tied to each scene region that they fixate. Scene description therefore allows us to examine the relative influences of semantic information and image salience under conditions in which on-line changes in attention to specific scene regions are functional and necessary. We used this basic paradigm in two experiments. In one experiment, subjects described what someone might do in each scene, and in a second experiment, they simply described each scene. In both experiments, subjects were

asked to begin their description when the scene appeared and to continue speaking for 30 s of scene presentation. Their eyes were tracked throughout each trial. The main result was that, once again, both meaning and salience were associated with the spatial distribution of attention, but when the correlation between meaning and salience was statistically controlled, only meaning accounted for variance in attention. This basic result was seen in both experiments. Therefore, once again we found no evidence for a unique influence of image salience on attention that could not be explained by the relationship between image salience and meaning, whereas the unique influence of meaning could not be attributed to salience.

So far, across all four tasks we have described (memorization, aesthetic judgment, scene description, and action description), meaning was highly relevant. Perhaps for tasks in which image salience is necessary and meaning is irrelevant, salience would better predict attention. To test this hypothesis, in a final set of experiments we examined the role of meaning and salience in two experiments using tasks for which meaning was completely irrelevant and saliency was critical: a brightness rating task in which participants rated each scene for its overall brightness, and a brightness search task in which participants counted the number of bright patches in each scene [53]. If meaning was used to guide attention in the previous tasks because those tasks emphasized the semantic content of the scenes, then the relationship between meaning and attention should no longer hold in the tasks that focus on the image itself. On the other hand, if the use of meaning to guide attention is a fundamental property of the operation of the attention system when faced with real-world scenes, then we should continue to see a relationship between meaning and attention even if meaning is irrelevant. Using the same methods as the previous studies, the striking finding was that in both the brightness rating and brightness search tasks, the results were very similar to the prior experiments: When the correlation between meaning and salience was controlled, only meaning uniquely accounted for significant variance in attention. These results showed that the relationship between meaning and attention is not restricted to viewing tasks that require the viewer to analyze meaning. The results support theories in which scene semantics play a central role in attentional guidance in scenes.

3. Discussion

We have reviewed the meaning map approach and shown how we have used it to investigate the relative roles of semantic informativeness and image salience on attentional guidance. Our results strongly suggest a fundamental and mandatory role for meaning in attentional guidance in real-world scenes. These results are consistent with a growing understanding that both overt and covert visual attention are often under the influence of the meaning of the visual stimulus, even when that meaning seems irrelevant to the task. Examples of this type of semantic effect are influences of object meaning on eye movements in the visual world paradigm [54], and influences of semantic object relationships on covert attention [55]. Furthermore, we found that the observed relationship between meaning and attention was about as strong as it could be given the variability in attention maps across subjects [42]. In the remainder of this section we consider and discuss some additional related issues.

First, although it has sometimes been proposed that image salience and semantic content are likely to be correlated in scenes, it has been difficult to test this hypothesis directly. The meaning map approach provides such a method. Further, as we have shown across several studies, that correlation is robust. An important implication of this finding is that previous results demonstrating a relationship between saliency maps and attention cannot be taken as evidence for a functional role of salience in guiding attention. Indeed, as we have reviewed above, essentially the entire relationship between image salience and attention can be attributed to the association between image salience and semantic content.

Second, it is important to be clear that meaning maps are not a theory of scene semantics. They are simply a method for tapping into and representing human judgments concerning the relative informativeness of scene regions continuously over space. Meaning maps provide an operational definition of the spatial distribution of meaning that can be quantified, but they do not offer direct insight into the nature of scene semantics or how meaning is represented in the mind and brain.

That said, it may be that the meaning map approach can be used as a tool for beginning to get a handle on the nature of scene semantics. For example, the type of scene meaning we have investigated so far has been based on ratings of scene patches that were presented to raters independently of the scenes from which the patches were taken. These experiments therefore focus on the role of what we call scene-intrinsic context-free meaning on attention [51]. We might collect ratings using other types of questions that could provide insight into other types of semantic representations. For example, we could compare context-free to contextualized meaning in which the degree of meaning associated with a scene region is based on its relationship to the scene as a whole rather than on its own. Similarly, we might consider goal-related meaning, in which the meaning of a scene region is based on its relationship to the viewer's goals rather than intrinsic to the scene itself. Using judiciously chosen ratings, we might be able to unravel how different types of meaning are related to each other and to performance over different perceptual and cognitive tasks. The meaning map approach provides a method for pursuing these questions. Relatedly, the meaning maps we have investigated so far based on context-free ratings may not be the type of meaning most associated with attention, and we may therefore be underestimating the relationship between semantic features and attention.

One advantage of saliency maps over meaning maps is that the former are image computable: They can be derived automatically and without human intervention from computational models. In comparison, meaning maps are not image computable and require human raters. For this reason, one might suggest that saliency models are to be preferred as an explanation of human attention. From an engineering perspective, this view has merit. However, from a vision and cognitive science perspective, it does not. In our view, the interesting psychological claim of the image salience hypothesis is that human attention in real-world scenes is driven by contrasts in basic image features. This claim has been supported by a large number of experiments showing a correlation between saliency maps and human fixations. The alternative hypothesis we are pursuing is that image salience effects are actually disguised meaning effects, because image features and semantic features in scenes are correlated [42]. Indeed, this is what we find, with very little if any unique variance accounted for by salience once the variance accounted for by semantic features is controlled. This comparison is not one of a computational model versus human ratings, but one of two competing psychological theories. That is, we are concerned with psychological principles here, not modeling. There is no logical requirement that testing this (or any) psychological hypothesis requires image computable semantic features. Furthermore, until we have a computational model of the entire semantic system (which is clearly a long way off), there is no other way to go about comparing salience to semantics. Saliency models have been influential because they have been the only game in town [4,10]. Meaning maps provide an alternative game, but their creation does require human judgment. From our perspective, this approach is similar to the approach that uses human labeling to parse and label objects in scenes as in the labelMe database [56] to produce object ground truth. Of course, it would be very interesting to use the ground truth represented by meaning maps to try to train a model to find meaningful regions (and indeed we are pursuing this idea), but that is not necessary for testing the theories at stake.

Relatedly, it has been argued that Koch and Ullman-inspired saliency models like GBVS are no longer state of the art, but instead have been replaced by a newer class of models based on deep neural networks (DNNs) that have recently been found to predict human attention quite well [57]. Given this, one might ask why we should take standard saliency models as the baseline for comparison to meaning. In our view, although these DNN models are impressive from the perspective of pushing the boundaries of deep learning and big data, it is not clear at this point how much they have to say about active biological vision. For example, DNN models are trained on fixations over one set of scenes and then predict fixations on another set of scenes. It is clear that humans do not learn where to fixate based on supervised learning from the fixations of others or by ingesting large amounts of external fixation data. It is also unclear whether the mechanisms used by DNNs to predict attention operate on the same principles that are used by the human brain. For these reasons, although we watch

developments in this field with great interest, we are not yet sure how their successes and failures should be interpreted from the perspective of human attentional processes.

Author Contributions: J.H. wrote the first draft and all authors edited and contributed to later drafts. All authors contributed to the conceptualization, implementation, and analysis of subsets of the experiments reviewed here.

Funding: 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.

Acknowledgments: We thank the members of the UC Davis visual cognition group, and Carrick Williams and two anonymous reviewers, for their helpful feedback on this work. We also thank Brad Wyble, Tom Wallis, and others for a stimulating Twitter discussion of the issues raised here.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Land, M.F.; Hayhoe, M.M. In what ways do eye movements contribute to everyday activities? *Vis. Res.* **2001**, *41*, 3559–3565. [CrossRef]
2. Hayhoe, M.M.; Ballard, D. Eye movements in natural behavior. *Trends Cogn. Sci.* **2005**, *9*, 188–194. [CrossRef]
3. Henderson, J.M. Human gaze control during real-world scene perception. *Trends Cogn. Sci.* **2003**, *7*, 498–504. [CrossRef] [PubMed]
4. Henderson, J.M. Gaze Control as Prediction. *Trends Cogn. Sci.* **2017**, *21*, 15–23. [CrossRef]
5. Buswell, G.T. *How People Look at Pictures*; University of Chicago Press: Chicago, IL, USA, 1935.
6. Yarbus, A.L. *Eye Movements and Vision*; Plenum Press: New York, NY, USA, 1967; ISBN 0306302985.
7. Henderson, J.M.; Hollingworth, A. High-level scene perception. *Ann. Rev. Psychol.* **1999**, *50*, 243–271. [CrossRef] [PubMed]
8. Rayner, K. The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* **2009**, *62*, 1457–1506. [CrossRef]
9. Liversedge, S.P.; Findlay, J.M. Saccadic eye movements and cognition. *Trends Cogn. Sci.* **2000**, *4*, 6–14. [CrossRef]
10. Henderson, J.M. Regarding scenes. *Curr. Dir. Psychol. Sci.* **2007**, *16*, 219–222. [CrossRef]
11. Henderson, J.M. Eye movements and scene perception. In *The Oxford Handbook of Eye Movements*; Liversedge, S.P., Gilchrist, I.D., Everling, S., Eds.; Oxford University Press: New York, NY, USA, 2011; pp. 593–606.
12. Treisman, A.M.; Gelade, G. A Feature-Integration Theory of Attention. *Cogn. Psychol.* **1980**, *12*, 97–136. [CrossRef]
13. Wolfe, J.M. Guided Search 2.0. A revised model of visual search. *Psychon. Bull. Rev.* **1994**, *1*, 202–238. [CrossRef] [PubMed]
14. Wolfe, J.M.; Horowitz, T.S. Five factors that guide attention in visual search. *Nat. Hum. Behav.* **2017**, *1*, 1–8. [CrossRef]
15. Borji, A.; Parks, D.; Itti, L. Complementary effects of gaze direction and early saliency in guiding fixations during free viewing. *J. Vis.* **2014**, *14*, 3. [CrossRef]
16. Borji, A.; Sihite, D.N.; Itti, L. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Trans. Image Proc.* **2013**, *22*, 55–69. [CrossRef]
17. Harel, J.; Koch, C.; Perona, P. Graph-Based Visual Saliency. *Adv. Neural Inf. Proc. Syst.* **2006**, 1–8.
18. Itti, L.; Koch, C. Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2001**, *2*, 194–203. [CrossRef]
19. Koch, C.; Ullman, S. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. *Hum. Neurobiol.* **1985**, *4*, 219–227.
20. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]
21. Parkhurst, D.; Law, K.; Niebur, E. Modeling the role of salience in the allocation of overt visual attention. *Vis. Res.* **2002**, *42*, 107–123. [CrossRef]
22. Loftus, G.R.; Mackworth, N.H. Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol.* **1978**, *4*, 565–572. [CrossRef]
23. Antes, J.R. The time course of picture viewing. *J. Exp. Psychol.* **1974**, *103*, 62–70. [CrossRef]

24. Mackworth, N.H.; Morandi, A.J. The gaze selects informative details within pictures. *Percept. Psychophy.* **1967**, *2*, 547–552. [CrossRef]
25. Wu, C.C.; Wick, F.A.; Pomplun, M. Guidance of visual attention by semantic information in real-world scenes. *Front. Psychol.* **2014**, *5*, 1–13. [CrossRef] [PubMed]
26. Tatler, B.W.; Hayhoe, M.M.; Land, M.F.; Ballard, D.H. Eye guidance in natural vision: Reinterpreting salience. *J. Vis.* **2011**, *11*, 5. [CrossRef] [PubMed]
27. Rothkopf, C.A.; Ballard, D.H.; Hayhoe, M.M. Task and context determine where you look. *J. Vis.* **2007**, *7*, 16.1-20. [CrossRef] [PubMed]
28. Hayhoe, M.M.; Ballard, D. Modeling Task Control of Eye Movements Minireview. *Curr. Biol.* **2014**, *24*, R622–R628. [CrossRef]
29. Einhäuser, W.; Rutishauser, U.; Koch, C. Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *J. Vis.* **2008**, *8*, 2.1-19. [CrossRef] [PubMed]
30. Torralba, A.; Oliva, A.; Castelhano, M.S.; Henderson, J.M. Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychol. Rev.* **2006**, *113*, 766–786. [CrossRef] [PubMed]
31. Castelhano, M.S.; Mack, M.L.; Henderson, J.M. Viewing task influences eye movement control during active scene perception. *J. Vis.* **2009**, *9*, 6.1-15. [CrossRef] [PubMed]
32. Neider, M.B.; Zelinsky, G. Scene context guides eye movements during visual search. *Vis. Res.* **2006**, *46*, 614–621. [CrossRef] [PubMed]
33. Turano, K.A.; Geruschat, D.R.; Baker, F.H. Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vis. Res.* **2003**, *43*, 333–346. [CrossRef]
34. Foulsham, T.; Underwood, G. How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception* **2007**, *36*, 1123–1138. [CrossRef]
35. Henderson, J.M.; Brockmole, J.R.; Castelhano, M.S.; Mack, M. Visual saliency does not account for eye movements during visual search in real-world scenes. In *Eye Movements: A Window on Mind and Brain*; Van Gompel, R.P.G., Fischer, M.H., Murray, W.S., Hill, R.L., Eds.; Elsevier Ltd.: Oxford, UK, 2007; pp. 537–562, ISBN 9780080449807.
36. Henderson, J.M.; Malcolm, G.L.; Schandl, C. Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychon. Bull. Rev.* **2009**, *16*, 850–856. [CrossRef]
37. Borji, A.; Itti, L. State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 185–207. [CrossRef]
38. De Graef, P.; Christiaens, D.; d'Ydewalle, G. Perceptual effects of scene context on object identification. *Psychol. Res.* **1990**, *52*, 317–329. [CrossRef]
39. Henderson, J.M.; Weeks, P.A., Jr.; Hollingworth, A. The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol.* **1999**, *25*, 210–228. [CrossRef]
40. Võ, M.L.H.; Henderson, J.M. Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *J. Vis.* **2009**, *9*, 1–15. [CrossRef]
41. Brockmole, J.R.; Henderson, J.M. Prioritizing new objects for eye fixation in real-world scenes: Effects of object-scene consistency. *Vis. Cogn.* **2008**, *16*, 375–390. [CrossRef]
42. Henderson, J.M.; Hayes, T.R. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat. Hum. Behav.* **2017**, *1*, 743–747. [CrossRef]
43. Henderson, J.M.; Pierce, G.L. Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychon. Bull. Rev.* **2008**, *15*, 566–573. [CrossRef]
44. Nuthmann, A.; Smith, T.J.; Engbert, R.; Henderson, J.M. CRISP: A computational model of fixation durations in scene viewing. *Psychol. Rev.* **2010**, *117*, 382–405. [CrossRef]
45. Henderson, J.M.; Smith, T.J. How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Vis. Cogn.* **2009**, *17*, 1055–1082. [CrossRef]
46. Glaholt, M.G.; Reingold, E.M. Direct control of fixation times in scene viewing: Evidence from analysis of the distribution of first fixation duration. *Vis. Cogn.* **2012**, *20*, 605–626. [CrossRef]
47. Henderson, J.M.; Nuthmann, A.; Luke, S.G. Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations. *J. Exp. Psychol.* **2013**, *39*, 318–322. [CrossRef]
48. Van Diepen, P.; Ruelens, L.; d'Ydewalle, G. Brief foveal masking during scene perception. *Acta Psychol.* **1999**, *101*, 91–103. [CrossRef]

49. Luke, S.G.; Nuthmann, A.; Henderson, J.M. Eye movement control in scene viewing and reading: Evidence from the stimulus onset delay paradigm. *J. Exp. Psychol.* **2013**, *39*, 10–15. [CrossRef] [PubMed]
50. Henderson, J.M.; Hayes, T.R. Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *J. Vis.* **2018**, *18*. [CrossRef] [PubMed]
51. Henderson, J.M.; Hayes, T.R.; Rehrig, G.; Ferreira, F. Meaning Guides Attention during Real-World Scene Description. *Sci. Rep.* **2018**, *8*, 1–9. [CrossRef]
52. Ferreira, F.; Swets, B. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* **2002**, *46*, 57–84. [CrossRef]
53. Peacock, C.E.; Hayes, T.R.; Henderson, J.M. Meaning guides attention during scene viewing, even when it is irrelevant. *Atten. Percept. Psychophy.* **2019**, *81*, 20–34. [CrossRef]
54. Huettig, F.; McQueen, J.M. The tug of war between phonological, semantic and shape information in language-mediated visual search. *J. Mem. Lang.* **2007**, *57*, 460–482. [CrossRef]
55. Shomstein, S.; Malcolm, G.L.; Nah, J.C. Intrusive Effects of Task-Irrelevant Information on Visual Selective Attention: Semantics and Size. *Curr. Opin. Psychol.* **2019**. [CrossRef] [PubMed]
56. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. 2008 LabelMe. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
57. Kummerer, M.; Wallis, T.S.A.; Gatys, L.A.; Bethge, M. Understanding Low- and High-Level Contributions to Fixation Prediction. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4799–4808. [CrossRef]