Searching in the dark:
Cognitive relevance drives attention
in real-world scenes

JOHN M. HENDERSON AND GEORGE L. MALCOLM
University of Edinburgh, Edinburgh, Scotland

AND

CHARLES SCHANDL
McMaster University, Hamilton, Ontario, Canada

We investigated whether the deployment of attention in scenes is better explained by visual salience or by cognitive relevance. In two experiments, participants searched for target objects in scene photographs. The objects appeared in semantically appropriate locations but were not visually salient within their scenes. Search was fast and efficient, with participants much more likely to look to the targets than to the salient regions. This difference was apparent from the first fixation and held regardless of whether participants were familiar with the visual form of the search targets. In the majority of trials, salient regions were not fixated. The critical effects were observed for all 24 participants across the two experiments. We outline a cognitive relevance framework to account for the control of attention and fixation in scenes.

What we perceive, understand, and remember about a visual scene is influenced greatly by where we look (Findlay & Gilchrist, 2003; Henderson, 2008). Therefore, a fundamental goal in studying scene perception is to understand the processes that determine where we look (Buswell, 1935; Yarbus, 1967; see Henderson, 2003, 2007, for reviews). According to the *visual salience hypothesis*, we look at scene locations on the basis of image properties, such as intensity, color, and edge orientation, generated in a bottom-up manner from the scene (Henderson, Weeks, & Hollingworth, 1999; Itti & Koch, 2000; Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002; Torralba, 2003). Regions that are uniform along some image dimension are considered uninformative, whereas those that differ from neighboring regions across spatial scales are potentially informative and are worthy of attention. The visual salience hypothesis has had a large impact on research in scene perception, in part because it has been instantiated within a neurobiologically plausible computational model (Itti & Koch, 2000) that has been found to capture gaze behavior under some conditions (e.g., Parkhurst et al., 2002). According to the visual salience hypothesis, gaze control is a reaction to the visual properties of the stimulus confronting the viewer.

Alternatively, according to what we call the *cognitive relevance hypothesis*, the selection of fixation sites is based on the needs of the cognitive system in relation to an understanding of scene meaning (i.e., based on

cognitive knowledge structures in memory) interacting with the goal of the current task (Henderson, 2003, 2007; Henderson, Brockmole, Castelhano, & Mack, 2007; Henderson & Ferreira, 2004; Navalpakkam & Itti, 2005; Torralba, Oliva, Castelhano, & Henderson, 2006; see also Foulsham & Underwood, 2007, and Land & Hayhoe, 2001). According to this hypothesis, the visual stimulus is still relevant: The eyes are typically directed to objects rather than to background (Henderson, 2003; Henderson & Hollingworth, 1998), and the scene image serves as the key for accessing stored knowledge structures. However, the weight given to a particular object or visual feature for attentional control in the scene is determined by current cognitive information-gathering needs rather than by visual salience. Consistent with the cognitive relevance hypothesis, a recent study reported that low-salience search targets were found as quickly as were medium-salience targets during search in scenes (Foulsham & Underwood, 2007). However, the difference between low- and medium-salience targets may not have been extreme enough to produce a difference: Search was numerically facilitated for medium-salience as compared with low-salience targets in many of the reported eye movement measures, including the probability of fixating the target early in the trial.

Here, we investigated gaze control during scene search by contrasting the degree to which fixation location is related to visual salience versus cognitive relevance in the more extreme case of a nonsalient target. The stimuli

---

J. M. Henderson, john.m.henderson@ed.ac.uk

were full-color photographs of real-world scenes containing objects that required some effort to find (see Figure 1). The critical scenes (22 and 24 in Experiments 1 and 2) met three criteria: The search target was not salient according to the Itti and Koch (2000) algorithm, the scene contained highly salient regions at other locations, and the search target appeared in a semantically appropriate location. Search is clearly an important task in its own right (e.g., *Where did I leave my keys?*). More generally, we suggest that search is a ubiquitous activity in attentive vision. For example, a sandwich maker needs to search for the bread, the knife, the peanut butter, and so forth in the proper sequence to perform the task (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land & Hayhoe, 2001).

To compare the influence of visual salience and cognitive relevance, we analyzed search eye movements with respect to two regions of interest (ROIs). One ROI surrounded the target object, and the other surrounded the most salient region. The visual salience hypothesis predicts that search for nonsalient targets should be slow and difficult. In the extreme, nonsalient targets should not be found at all, because attention and fixation should not be directed to them. The visual salience hypothesis also predicts that the most salient scene region should attract attention and so should be fixated more frequently and earlier than the nonsalient targets are. According to the cognitive relevance hypothesis, on the other hand, search for the target objects should be guided by knowledge of the task, the meaning and spatial layout of the scene, and memory for the relationship between a particular type of object and a particular type of scene. Fixation should not be directed to the most visually salient region but toward cognitively relevant regions, so the targets should be fixated more frequently and earlier than the salient regions are. In the extreme, salient regions should rarely be fixated at all.

The Present Experiments

In two experiments, participants searched through 22 (Experiment 1) and 24 (Experiment 2) critical scenes that contained nonsalient target objects. The critical scenes were embedded in a total set of 60 scenes in each experiment. In order to ensure that participants did not strategically ignore the salient regions, the extra scenes were included as fillers in which targets were visually salient.

It could be that, if the visual properties of the targets are known prior to search, visual salience can be overridden by feature similarity between a search template held in memory and a visual representation of the target generated from the scene (Rao, Zelinsky, Hayhoe, & Ballard, 2002; Vickery, King, & Jiang, 2005; Wolfe, 1994; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). On the
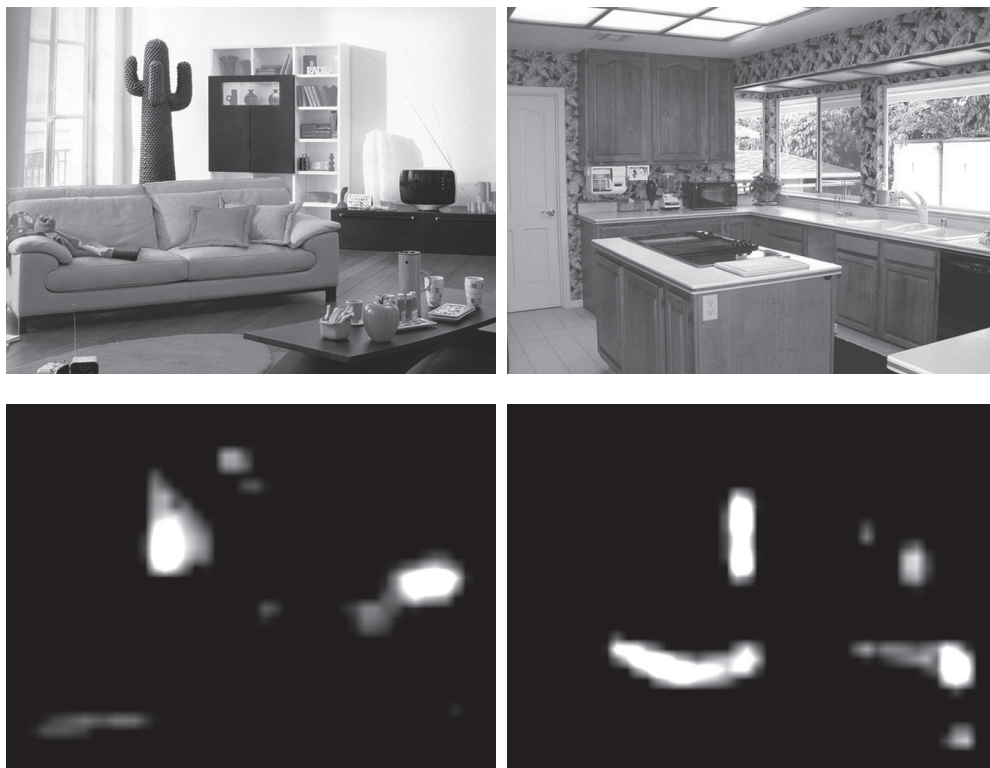

Figure 1. Example stimuli. The top row shows two scene photographs; the second row shows the winner-take-all salience maps for these photographs. The targets were an apple and a blender in the left and right pictures, respectively. (Original images were in color.)

other hand, with unfamiliar targets, search may rely on general salience. To investigate this issue, we manipulated target familiarity within and across experiments. In Experiment 1, participants were familiarized with the search targets prior to the experiment. In Experiment 2, no familiarization procedure was used. Within each experiment, we manipulated whether the search target was cued with a picture or with a word.

METHOD

Participants

Twenty-four participants (12 in each experiment) gave informed consent and were paid £6. All participants were naive about the purpose of the study.

Stimuli

Of the 60 scene photographs used in each experiment, 22 and 24 critical scenes (Experiments 1 and 2, respectively) included non-salient targets. The remaining filler scenes included targets that were visually salient. Each critical scene included a target object that appeared once, was not at the center, was not occluded, was smaller than 3° in diameter, and was easily identified once found. Target objects were not visually salient in their scenes (salience map black for values < 10 in the winner-take-all map of Itti & Koch, 2000). The nearest salient region was spatially distinct from the target, so fixations directed toward that region would be spatially separate from those directed toward the target. ROIs averaged 7.86° and 8.76° from the initial fixation point in the two experiments, respectively. Target ROIs were slightly closer than salient ROIs [8.29° vs. 7.43°, n.s., and 9.48° vs. 8.05°, respectively, in the two experiments; $t(23) = 2.335$, $p < .05$], which would tend to favor the salient ROIs since fixations tend to select closer targets (see Henderson & Hollingworth, 1998). The filler scenes were similar except that all of their targets were salient to some degree, and many were the most salient objects in their respective scenes. Images were presented at a resolution of 800 × 600 pixels with 24-bit color.

Apparatus

Eye movements were recorded using an SR Research EyeLink 2K eyetracker sampling at 1000 Hz. Saccades were defined with a 50°/sec velocity threshold using a nine-sample saccade detection model. Viewing was binocular, but only the right eye was tracked. Stimuli were presented on a 21-in. ViewSonic G225f monitor (ViewSonic, London) positioned 90 cm from the participant (18.72° × 24.28° field of view), with a refresh rate of 140 Hz.

Procedure

In Experiment 1, participants were familiarized with the targets before the search trials. Each target object was presented three times on a computer screen as a picture–word pair, and participants sequentially examined each pair at their own pace. In Experiment 2, no familiarization procedure was used.

In each trial, a central fixation cross appeared for 400 msec, followed by a target cue (picture or word) identifying the search target. Assignment of picture or word cue to scene was counterbalanced across participants, and picture cues were identical to targets. The scene appeared following the cue, and participants were asked to locate the search target as quickly and as accurately as possible and to press a response key as soon as the target was found. Participants were given nine practice trials prior to the experiment. Trial order was randomized for each participant.

ROI Definitions and Analyses

ROIs for scoring eye movements were rectangles surrounding the target objects and salient regions. Within a scene, the ROIs were identical in size and did not overlap spatially. ANOVAs were performed with ROI and cue as within-subjects factors.

RESULTS

Proportion of Trials in Which the Salient Region and Target Region Were Fixated

Consistent with the predictions of the cognitive relevance hypothesis, but contrary to those of the visual salience hypothesis, in Experiment 1, the target ROI was fixated on about 90% of trials, whereas the most salient ROI was fixated on fewer than 10% of trials [$F(1,11) = 522$, $MS_e = 0.014$, $p < .001$; see Table 1]. A similar pattern was observed for Experiment 2, where participants were not familiarized with the targets. The target ROI was fixated on about 95% of trials, whereas the most salient ROI was fixated on about 8% of trials [$F(1,11) = 525.868$, $MS_e = 0.050$, $p < .001$]. This pattern was true for all 24 participants across the two experiments. Thus, consistent with the cognitive relevance hypothesis but not the visual salience hypothesis, the relevant region was fixated with high probability, even though it was not visually salient, whereas the most salient region was rarely fixated.

Comparison of Experiments 1 and 2 suggests that familiarization with the targets did not contribute to the cognitive relevance effect. Similarly, there was no effect of cue type or interaction of cue type and ROI within either experiment ($Fs \leq 1$).

Proportion of Trials in Which Target ROI Was Fixated First

We also examined the proportion of trials in which the target or salient ROI was fixated first (Table 1). For this analysis, if the eyes entered the target ROI before entering the salient ROI (regardless of whether the salient ROI was eventually fixated), it was scored as a hit for the target ROI. Alternatively, if the eyes entered the salient ROI before entering the target ROI (regardless of whether the target ROI was eventually fixated), it was scored as a hit for the salient ROI. Since at least one of these two regions was fixated on every trial, these two proportions sum to 1.0. This measure provides temporal information about

Table 1
Proportion of Trials on Which the Target and Salient Regions Were Fixated, Proportion of Trials on Which the Target Region Was Fixated First, and Search Time to Target Region (in Milliseconds), for Word and Picture Cues in Experiments 1 and 2

	Experiment 1		Experiment 2	
	Word	Picture	Word	Picture
Proportion of Trials Target and Salient Regions Fixated				
Target ROI	.92	.87	.96	.94
Salient ROI	.12	.06	.08	.11
Proportion of trials target regions fixated first	.87	.94	.92	.90
Search time to target regions	1,058	836	1,150	1,067

Note—ROI, region of interest.

ROI fixation order. In Experiment 1, summing across cue conditions, participants fixated the target region first on about 90% of the trials, compared with about 10% for the salient region. The same pattern was observed in Experiment 2: Participants fixated the target region first on 91% of the trials, compared with 9% for the salient region. This pattern held for all 24 participants across the two experiments.

In Experiment 1, cue type (picture or word) interacted with region [$F(1,11) = 8.163$, $MS_e = 0.008$, $p = .016$] with a greater bias to fixate the target region first following a picture than following a word (94% vs. 87%). However, this effect was not observed in Experiment 2 (90% vs. 92%), with no effect of cue type and no interaction of cue type and ROI ($Fs < 1$).

Region Fixation by Ordinal Fixation Number

It could be that visual salience drives initial eye movements, but that cognitive relevance increases over time as more knowledge is acquired about the scene (Henderson et al., 1999; Parkhurst et al., 2002). Alternatively, it may be that, in complex meaningful scenes, the first saccade is under cognitive control because the global scene gist acquired from the first fixation provides sufficient information about where that object is likely to be found (Antes, 1974; Brockmole & Henderson, 2006; Castelhano & Henderson, 2007; Eckstein, Drescher, & Shimozaki, 2006; Henderson et al., 1999; Mackworth & Morandi, 1967; Torralba et al., 2006). To investigate this issue, we examined the proportion of fixations to the target and salient regions as a function of ordinal fixation number. Figure 2 plots the fixations on the two regions as a proportion of all fixations and as a function of ordinal fixation number. As can be seen in Figure 2, by the first few viewer-determined fixations, the target regions were already receiving proportionally more fixations than were salient regions. This effect held for both experiments. Therefore, once again we found no evidence that early fixations were directed toward the salient regions; instead, initial fixations were more likely to be directed toward the target regions.

Time to First Fixate Target Object

Search in crowded displays of unrelated objects is typically difficult and time consuming. On the other hand, search for meaningful objects in real-world scenes can be extraordinarily fast, particularly given the number of objects that typically occupy a complex scene (Castelhano & Henderson, 2005; Henderson, Chanceaux, & Smith, 2009; Henderson et al., 1999; Torralba et al., 2006). To examine how quickly the targets were found in the present study, we computed the time from scene onset until the target was fixated the first time (Table 1). Search was very fast and efficient, with targets found in about 947 and 1,108 msec, on average, in Experiments 1 and 2, respectively. Targets cued by pictures were found faster than those cued by words in Experiment 1 [$F(1,11) = 10.59$, $MS_e = 55,870$, $p < .01$], and there was a numerical but nonsignificant advantage for targets cued by pictures over words in Experiment 2 (1,067 vs. 1,150 msec;
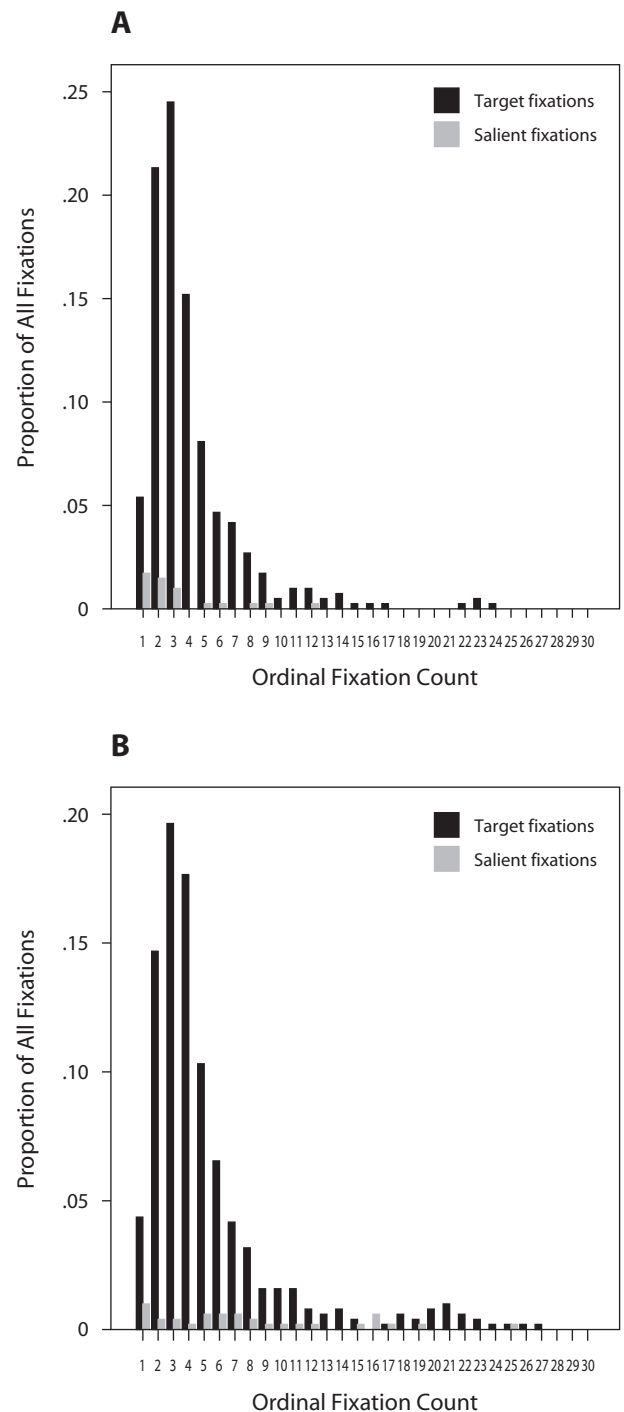

Figure 2. Proportion of all fixations that landed in the target region of interest (ROI, black bars) and the salient ROI (gray bars) as a function of ordinal fixation number, Experiment 1 (A) and Experiment 2 (B). Fixation 1 is the first viewer-determined fixation.

$F < 1$). But it is clear from Table 1 that even targets cued only by words were found quickly in both experiments. These data suggest that search for objects in scenes is remarkably fast and efficient, even if those objects are

not visually salient. More generally, because participants were not familiarized with the visual properties of the search targets in Experiment 2, the results with the word cues in that experiment suggest that search for targets cued only by words is fast and efficient, even when the specific visual features of the targets are unknown.

DISCUSSION

The control of attention and eye movements is a central issue in scene perception. By measuring eye movements during search through photographs, we investigated the roles of visual salience and cognitive relevance in controlling gaze. We reasoned that, if visual salience plays a dominant role in guiding attention and eye movements, viewers should tend to look at highly salient regions. In the extreme, a nonsalient target should be very difficult to find, because the eyes should never be directed toward it. If, on the other hand, the eyes are guided by cognitive relevance, highly salient regions that do not accord with likely target locations should be ignored, and nonsalient targets in expected locations should be relatively easy to find.

The results of two search experiments were clear: All 24 participants were very likely to look at a nonsalient scene region containing a search target and very unlikely to look at the most salient scene region. In the vast majority of trials, salient regions were not fixated at all. Furthermore, looks to the target regions tended to occur before looks to the salient regions in those few cases where the latter occurred. Fixation preference for the target region over the most salient region was observed from the very beginning of the search trial, and it held regardless of whether participants were familiar with the visual form of the search target. Search for nonsalient targets was fast and efficient: On average, the first fixation on a target required only three to four saccades. Overall, these results strongly call into question the adequacy of a theoretical approach based on visual salience to account for human eye movements in real-world scenes.

Toward a Cognitive Relevance Theory

We propose an alternative to the visual salience framework that involves doing away with the concept of an image-based salience map as a representation over which attention is directed and that instead places primary emphasis on knowledge-based control operating on an alternative visuospatial representation. In this view, which we refer to as the *cognitive relevance framework*, objects are prioritized for attention and fixation primarily on the basis of cognitive knowledge structures interacting with task goals. The scene image in this model still plays a critical role in at least two important ways. First, the scene image is needed to generate a representation over which to direct the eyes. Second, the image serves as a source of input for activating relevant cognitive knowledge structures. However, in contrast to the key assumption of the visual salience view, the potential saccade targets generated from the visual stimulus are not ranked for priority according

to visual salience, but rather are ranked on the basis of relevance to the task and current scene understanding.

To make the cognitive relevance proposal concrete, we provide the following outline. First, we assume that scene locations are chosen for fixation on the basis of their potential cognitive relevance, rather than on the basis of their visual salience. Cognitive relevance is based on knowledge of the task, semantic knowledge about the type of scene in view, episodic knowledge about the particular scene, and current scene interpretation and understanding (Henderson & Ferreira, 2004). This information will be available, both from initially generated semantic and structural representations (i.e., *gist*) and from more detailed representations generated during prior fixations. Second, we assume that the visual scene is parsed to generate a visuospatial representation that includes a coding of potential object locations and their spatial layout. We might think of this representation as a flat landscape in contrast to the peaked salience map, in that the regions are not yet ranked for attentional priority (Henderson et al., 2007). Figure 3 provides a cartoon of an initial scene parse.

As soon as a figure–ground representation coding object regions has been generated, saccade target ranking based on cognitive relevance can be applied to it. For example, in the scene depicted in Figure 3, if the viewer's task is to find a misplaced cell phone, potential objects in regions of space that are likely to contain the phone will be ranked more highly than other regions. If instead the task is to determine whether an awaited plane has arrived at its gate, potential objects occupying different spatial regions will be ranked more highly (Figure 4). Attention is then directed to regions in rank order on the basis of their cognitive relevance.

Visual salience is related to differences in image properties, such as luminance, color contrast, edge orientation, and motion over multiple spatial scales. We assume that these image properties are computed early in the visual system and that they contribute to the figure–ground parse and to the generation of the visuospatial representation of scene regions over which eye movements are planned. To

Flat Landscape

Scene Input

Image parsed into potential object regions (mostly unidentified) and not ranked by visual saliency
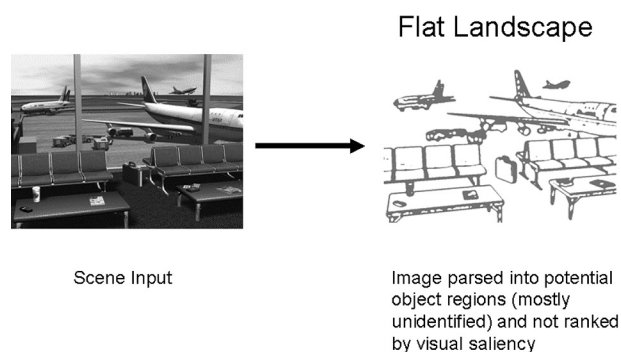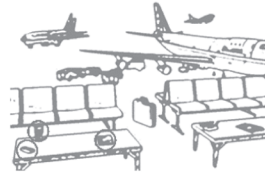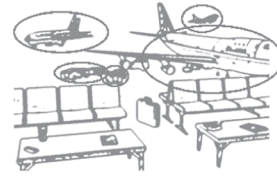
Figure 3. Schematic illustration of an initial scene parse and generation of a flat landscape representation. Although the objects in this representation are clearly identifiable as depicted, they are meant to represent preidentification proto-objects.

"Where is my phone?"   "Is my plane here yet?"

Saccade target ranking based on semantic and episodic knowledge and task (Where are phones likely to be? Where did I leave my phone? Where was I sitting?)

Figure 4. Schematic illustration of assigning weight to potential saccade targets on the basis of cognitive relevance. Given the same scene, different objects are relevant depending on the nature of the task and a general understanding of the scene and its spatial layout.

the extent that the boundaries between regions are easily identified, they are more likely to be included in the resulting visuospatial representation. And such regions are much more likely to be identified if they are more different from their surround (i.e., the region boundaries are more salient). However, the critical point is that ranking for the purposes of directing attention is not based on these visual salience values, but rather is based on cognitive relevance (Figure 5).[1]

If the cognitive relevance hypothesis is correct, why is evidence supporting visual salience sometimes reported? The main sources of support—correlation of fixation positions with (model-determined) visual salience and differences in scene statistics at fixated and nonfixated locations—are both problematic (Henderson, 2003; Henderson et al., 2007). In the case of correlations between fixations and visual salience, there is very little evidence that salience does a good job of predicting fixation locations for active viewing tasks in the world itself (Turano, Geruschat, & Baker, 2003). The present results similarly demonstrate that salience does a poor job of predicting

fixation locations during active visual search in static scene images when targets are in expected locations (see also Henderson et al., 2007; Torralba et al., 2006). In the case of differences in image statistics at fixated and nonfixated locations, recent results suggest that previously reported effects are as well explained by differences in region meaning as by differences in image statistics (Henderson et al., 2007). This confound is probably unavoidable: Objects differ from scene background in their image properties, and objects are entities to which meaning can be assigned. In other words, salience serves as a (weak) proxy for objects, and viewers prefer to look at objects over background (Buswell, 1935; Yarbus, 1967).

In summary, on a cognitive relevance theory of gaze control, the scene input is used to generate a flat (unranked) visuospatial landscape of potential object regions. This representation is fundamentally object based rather than image based. Critically, ranking for attention and fixation is based on cognitive knowledge structures and task requirements interacting with this visuospatial representation rather than directly on image properties.
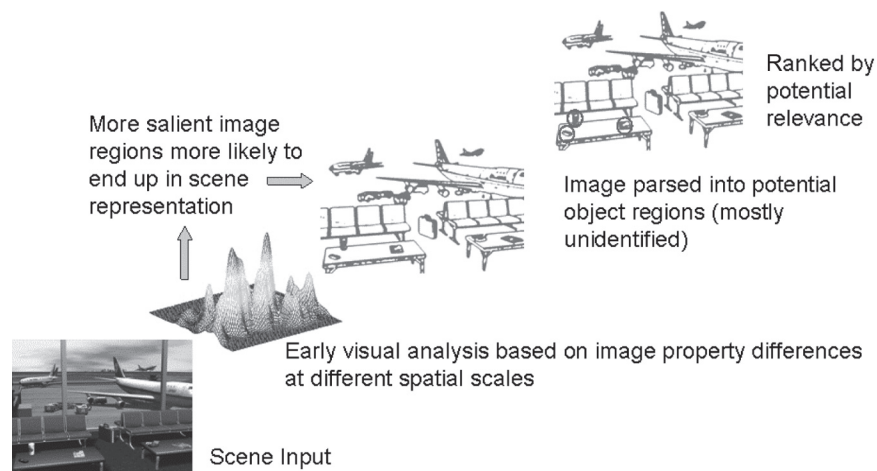

More salient image regions more likely to end up in scene representation

Ranked by potential relevance

Image parsed into potential object regions (mostly unidentified)

Early visual analysis based on image property differences at different spatial scales

Scene Input

Figure 5. What role does visual salience play in the control of attention? Salience influences the initial scene parse.

REFERENCES

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

Brockmole, J. R., & Henderson, J. M. (2005a). Object appearance, disappearance, and attention prioritization in real-world scenes. *Psychonomic Bulletin & Review*, **12**, 1061-1067.

Brockmole, J. R., & Henderson, J. M. (2005b). Prioritization of new objects in real-world scenes: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception & Performance*, **31**, 857-868.

Brockmole, J. R., & Henderson, J. M. (2006). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, **59**, 1177-1187.

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, **12**, 1017-1040.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **33**, 753-763.

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, **17**, 973-980.

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

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception*, **36**, 1123-1138.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, **3**, 49-63.

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

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

Henderson, J. M. (2008). Eye movements and scene memory. In S. Luck & A. Hollingworth (Eds.), *Visual memory* (pp. 87-121). Oxford: Oxford University Press.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Image salience versus cognitive control of eye movements in real-world scenes: Evidence from visual search. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movement research: Insights into mind and brain* (pp. 537-562). Oxford: Elsevier.

Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision*, **9**, 1-8. doi:10.1167/9.1.32

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1-58). New York: Psychology Press.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269-293). Oxford: Elsevier.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). Effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 210-228.

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

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

Land, M., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, **41**, 3559-3565.

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

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

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

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, **42**, 1447-1463.

Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America A*, **20**, 1407-1418.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, **113**, 766-786.

Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research*, **43**, 333-346.

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, **5**, 81-92.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, **1**, 202-238.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, **44**, 1411-1426.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum.

NOTE

1. The possible exception may be sudden-onset transients (Brockmole & Henderson, 2005a, 2005b; see Henderson et al., 2007, for a discussion).