# Where the Eyes Wander: The Relationship Between Mind Wandering and Fixation Allocation to Visually Salient and Semantically Informative Static Scene Content

Kristina Krasich[1], Greg Huffman[1], Myrthe Faber[2], & James R. Brockmole[1]

[1] Department of Psychology, University of Notre Dame, USA
[2] Donders Institute for Brain, Cognition, and Behaviour, Radboud University, The Netherlands

Vision is crucial for many everyday activities, but the mind is not always focused on what the eyes see. Mind wandering occurs frequently and is associated with attenuated visual and cognitive processing of external information. Corresponding changes in gaze behavior—namely fewer, longer, and more dispersed fixations—suggest a shift in how the visual system samples external information. Using three computational models of visual salience and two innovative approaches for measuring semantic informativeness, the current work assessed whether these changes reflect how the visual system prioritizes visually salient and semantically informative scene content, two major determinates in most theoretical frameworks and computational models of gaze control. Findings showed that, in a static scene viewing task, fixations were allocated to scene content that was more visually salient 10 s prior to probe-caught, self-reported mind wandering compared to self-reported attentive viewing. The relationship between mind wandering and semantic content was more equivocal, with weaker evidence that fixations are more likely to fall on locally informative scene regions. This indicates that the visual system is still able to discriminate visually salient and semantically informative scene content during mind wandering and may fixate on such information more frequently than during attentive viewing. Theoretical implications are discussed in light of these findings.

Vision is crucial for many everyday activities, and an in-depth analysis of the visual world requires that the eyes move. This is because the visual system is subject to physical (i.e., the structure and organization of photoreceptors) and cognitive (i.e., attention and memory) constraints. For instance, visual input is acquired during *fixations*—periods when the eye remains relatively stable—but is perceptually (e.g., Matin, 1974; Zuber & Stark, 1966) and cognitively (e.g., Campbell & Wurtz, 1978; Irwin & Carlson-Radvansky, 1996; Irwin & Brockmole, 2004) suppressed during *saccades*—the ballistic eye movements that shift fixations from one location to another. Therefore, the timing and location of fixation allocation offers insight into the real-time information-processing priorities of the visual system (e.g., Just & Carpenter, 1976; Kowler et al., 1995).

There are a number of known factors that influence fixation allocation in static scene viewing, which is frequently used as a proxy for how the visual system samples information in the real-world. These factors include the low-level, visually salient features of stimuli, such as color, contrast, and edge orientation (e.g., Mannan et al., 1996, 1997; Parkhurst & Neibur, 2003; Reinagel & Zador, 1999; Tatler et al., 2005), as well as higher-order, observer-driven factors, such as semantic interest (e.g., Buswell, 1935; Loftus & Mackworth, 1978), momentary task goals (e.g., Land & Hayhoe, 2001; Land & Lee, 1994; Yarbus, 1967), and long-term schematic knowledge of

This work is based on a doctoral dissertation completed by Kristina Krasich at the University of Notre Dame. Kristina Krasich is now at Duke University. Greg Huffman is now at Leidos, Inc. The authors thank Sidney D'Mello, Bradley Gibson, Gabriel Radvansky, Benjamin Tatler, and two anonymous reviewers for helpful support, comments, and suggestions regarding this project and article. Address correspondence regarding this article to Kristina Krasich, Center for Cognitive Neuroscience, Duke Institute for Brain Sciences, Duke University via email to: kristina.krasich@duke.edu.

scene structure (e.g., Mandler & Johnson, 1977; Shinoda, et al., 2001; Võ & Henderson, 2009). Although this is not an exhaustive list, it shows that converging influences from both stimulus- and observer-based factors impact fixation allocation, and a number of frameworks and computational models have attempted to characterize gaze control in light of these factors (e.g., Garcia-Diaz et al., 2012; Harel et al., 2007; Henderson & Hayes, 2017, 2018; Itti & Koch, 2000, 2001; Itti et al., 1998; Riche et al., 2013; Tatler et al., 2017).

An implicit assumption of current frameworks and models of gaze control is that observers consistently and appropriately attend to their visual surroundings. In reality, however, people are not always avidly attentive, and instead they frequently mind wander—defined here as a shift in attentional priorities away from task-relevant goals to task-irrelevant internal thoughts (Smallwood & Schooler, 2006). In fact, laboratory and field-based research have shown that, when asked, people will report having been mind wandering 20-50% of the time (Killingsworth & Gilbert, 2010; Smallwood & Schooler, 2015; although see Seli et al., 2018a). Thus, current theories and models of gaze control likely fail to capture the full range of influences on the manner in which the mind and brain sample the visual world. Our goal in this report was to address this limitation by considering how visual and cognitive factors known to enter into gaze control decisions vary as a function of observers' level of attentiveness to their tasks and goals.

**Visual processing during mind wandering**

Although the cognitive origin and progression of mind wandering is currently debated (Christoff et al., 2016, 2018; Seli et al., 2018b,c), a common theoretical view construes mind wandering as an attentional state that is, to some degree, decoupled from the external world (Schooler et al., 2011; Murphy et al., 2019). A growing body of neurocognitive evidence supports this perceptual decoupling account (Kam & Handy, 2018). For example, mind wandering is associated with an attenuated P1 event-related potential (ERP) component (Baird et al., 2014; Kam et al., 2011; Smallwood et al., 2008)—the ERP component that reflects early low-level visual processing (Hillyard et al., 1973). Interestingly, Barron et al. (2011) showed that retrospective self-report measures of mind wandering were also associated with reduced P3a—the component that reflects the capture of attention by rare distractor stimuli (Escera et al., 2000; Knight, 1997)—and the central-parietal P3b—a component that reflects the maintenance of a task-relevant stimulus in working memory (Polich, 2003). This collective evidence indicates that visual information processing during mind wandering is also attenuated at multiple levels of cognitive analysis—perception, attention, and working memory.

The perceptual decoupling observed during mind wandering suggests a deprioritization of visual information processing that might correspond to changes in fixation allocation during scene viewing. Indeed, Krasich et al. (2018) found evidence in support of this idea. They asked participants to study pictures of urban scenes in preparation for a later memory test. Periodically, participants self-reported whether they were mind wandering or attentively viewing the scene at a given moment via pseudo-randomly distributed thought probes that occurred 45-75 s into scene viewing. Findings showed that probe-caught mind wandering was associated with fewer, longer, and more dispersed fixations (compared to reports of attentive viewing), with the most robust observations found 10 s prior to the onset of the thought probe. Findings were conceptually replicated using a paradigm where scenes were presented contiguously, and thought probes occurred at pseudorandom intervals over the course of the viewing task. Accordingly, Krasich et al. (2018) inferred that given the perceptual decoupling during mind wandering, the visual system

becomes less efficient and slower to extract and evaluate visual information, thus prolonging fixations. The co-occurring increase in fixation dispersion, the authors suggested, may reflect a systematic, rather than random, shift in how information is sampled.

What remained unclear from Krasich et al. (2018), however, was whether changes in fixation duration and dispersion reflected a shift in how external information is prioritized during mind wandering. That is, does the visual system systematically change *what* visual information is sampled during mind wandering or does it simply alter *how* information is sampled (i.e., more slowly and broadly) throughout the scene? Answering this question will identify what visual information the visual system detects and prioritizes during conditions of attenuated visual processing such as during mind wandering.

## The current study

The current study focuses on visual and cognitive factors that have been linked to gaze control and how these relationships vary over attentional states. Specifically, our goal was to assess how fixations are allocated to visually salient and semantically informative scene content prior to self-reported mind wandering. To do this, we re-analyzed the data reported by Krasich et al. (2018) with a new focus on fixation placement relative to scene content. Using this prior study as a basis for our investigation is advantageous because it has already been used to demonstrate that several parameters of gaze control vary as a function of attentiveness. Hence, it gives us a direct opportunity to compare the relationship between mind wandering and both content-independent (i.e., those considered by Krasich et al., 2018), and content-specific (i.e., those considered here) measures of visual sampling.

**Visual salience.** Some computational models of gaze control compute stimulus-based properties to predict and model fixation allocation (e.g., Garcia-Diaz et al., 2012; Harel et al., 2007; Itti & Koch, 2000, 2001; Itti, et al., 1998; Itti & Baldi, 2005; Riche et al., 2013). One popular approach to operationalizing stimulus-based properties assumes that visual input (such as from a static scene) can be represented with iconic topographic *feature maps* (e.g., color, contrast, edge orientation, etc.) that are first extracted and then computationally combined to create a single *saliency map* that denotes the visual distinctiveness of any given location relative to surrounding locations or the entire image. These saliency maps, therefore, incorporate stimulus-based properties with little regard to higher order scene structure, and—in the absence of any goal-based, volitional control—salience-based models predict that fixations should be allocated to the most salient location first before moving to areas of lower saliency.

Given the shift of attentional priorities away from task goals during mind wandering (Smallwood & Schooler, 2006), mind wandering may provide conditions of reduced goal-based, volitional control. Fixations may therefore be allocated to highly salient scene content more frequently during bouts of mind wandering compared to attentive viewing. That said, following the perceptual decoupling accounts of mind wandering showing attenuated visual processing, it might be that the visual system becomes less sensitive to low-level, stimulus-based properties. This possibility predicts that fixations would not be more likely—and perhaps even less likely—to occur in visually salient scene content. No observable change in how fixations are allocated to visually salient scene content during mind wandering might also indicate that the changes in content-independent measures of gaze behavior observed in Krasich et al. (2018) are not reflective of a shift in how fixations are allocated to visually salient scene content.

To assess these competing hypotheses, we characterized visual salience for each of the images from Krasich et al. (2018) using three different salience-based computational models, which are among the most effective models of gaze control (Riche et al., 2013; Tatler et al., 2017): the Graph-Based Visual Salience model (GBVS; Harel et al., 2007), the Adaptive Whitening Saliency Model (AWS; Garcia-Diaz et al., 2012), and a rarity-based saliency model called RARE2012 (RARE; Riche et al., 2013). The GBVS first computes multi-scale feature maps (i.e., intensity, color, and orientation) via linear center-surround computations that mimic human visual receptive fields. Graph algorithms are then used to build activation maps by defining random-walk Markov chains from these feature maps. Activation maps are then merged into a final salience map such that the saliency of a given region reflects its contrasts to the local surrounding regions. The GBVS also incorporates a "center bias" that promotes higher saliency values in the center of the image, which accounts for past research showing a greater tendency for fixations to be allocated toward the center of a static images (e.g., Bindemann, 2010; Busswell, 1935; Parkhurst et al., 2002; Parkhurst & Niebur, 2003; Tatler, 2007; Tatler et al., 2005).

The AWS model is biologically motivated by the idea that the nonlinear neural responses in the visual cortex should be considered as collective neuron populations rather than as single units (decorrelation of neural responses; e.g., Olshausen & Field, 2005). It also assumes that low-level information is carried by high-order statistical structures and adopts a hierarchical approach to statistically whiten low-level features and remove second-order information (i.e., decorrelation and contrast normalization). The AWS model uses L*a*b* color space, which reduces the correlation between color components. Then log-Gabor filters are used to transform luminance into multiscale and multioriented representations, which are then decorrelated using a principal component analysis (PCA). The final saliency map is then computed by taking the sum of the squared norm vectors in the final representation and normalizing it to the sum across all pixels of the image. Thus, visual salience in the AWS represents a global decorrelation of the entire image.

The RARE model first extracts several feature maps. Low-level feature maps are computed through a hierarchical color transformation (PCA decomposition) and medium-level feature maps (e.g., texture) are extracted using Gabor filters that are modeled after simple cell neuronal activity in the visual cortex. A multi-scale rarity mechanism—the unique feature of the RARE model— is then applied to each feature map to compute rarity maps that denote locally distinctive contrasts as well as regions that are rare throughout the entire image. Lastly, an intra-channel and inter-channel fusion combines rarity maps into a final saliency map. Visual salience in the RARE model thus represents local rarity relative to the entire image.

The GBVS, AWS, and RARE models therefore compute visual salience with unique approaches. For instance, visual salience computed by the GBVS model reflects contrasts of local regions that favor more central regions, whereas visual salience in the AWS and RARE models reflect either adaptive whitening or rarity contrasts relative to the entire image without incorporating a center-bias. Computing visual salience using these different models allowed for the relationship between mind wandering and fixation allocation to visually salient scene content to be explored in multiple ways according to the procedures by which salience was computed.

**Semantic informativeness.** The success of salience maps in characterizing gaze control has fostered a robust empirical endeavor. That said, visual salience may not fully account for fixation allocation (e.g., Henderson et al., 2007), and the convenience of quantifying visual salience may disregard critical influences from semantic information (e.g., Henderson, 2017). For example, observers have a greater tendency to fixate locations that are rated as (e.g., Hayes & Henderson,

2019; Henderson & Hayes, 2017, 2018; Loftus & Mackworth, 1978; Mackworth & Morandi, 1967), or predicted to be (e.g., Bar, 2009; Clark, 2013; Friston, 2010; Lupyan & Clark, 2015), more semantically informative than surrounding regions even when those locations are less visually salient than surrounding regions (e.g., Henderson et al., 2009). Moreover, semantic information might be mischaracterized if it is also visually salient (e.g., Einhäuser et al., 2008; Elazary & Itti, 2008; Henderson, 2017). In fact, when compared directly, visual salience sometimes failed to explain fixation allocation above and beyond semantic information (e.g., Einhäuser, et al., 2008; Hayes & Henderson, 2019; Henderson, 2017; Henderson et al., 2007, Henderson & Hayes, 2017, 2018), although this is not universally true (Tatler et al., 2017). Therefore, the extent to which visual salience and semantic information contributes to gaze control is still debated; however, converging evidence indicates the importance of considering semantic information when investigating factors of gaze control.

Measuring semantic information poses somewhat of a challenge (in comparison to visual salience) given the subjective nature of observer evaluation. For instance, it is not always clear how objects should be defined, evaluated, and prioritized (e.g., Boriji et al., 2013a,b; Einhäuser, et al., 2008; Nuthmann & Henderson, 2010), and objects can be valued as important even before they are completely identified (Spain & Perona, 2009). Recent efforts, however, have attempted to map the variation in semantic information across an entire image in a conceptually similar way as a visual saliency map: semantic values are spatially distributed nonuniformly across an image with certain regions measured as more semantically informative than others.

Currently, there are two approaches that characterize semantic information in different ways. One approach, which we refer to as the *semantic interest map*, identifies regions within a scene that are judged to be the *most* semantically informative locations relative to the *entire global scene context* (Tatler et al., 2017). Specifically, third-party observers subjectively select the five most semantically informative regions of a scene while viewing the entire image, which allows observers to consider scene context but results in only a few areas that are indicated as being highly informative. The other method, referred to as the *meaning map* approach, gauges how *locally informative* or recognizable information is within small region of a scene (vignette) that is rated *independently of scene context* (Hayes & Henderson, 2019; Henderson & Hayes, 2017, 2018). For these maps, third-party observers rate how informative or recognizable information is within each vignette, and then vignettes are interpolated to produce a cohesive map so that each location within a scene contains a semantic value. The critical differences between these two approaches are 1) whether semantic information is evaluated in relation to or independently from the entire scene, and 2) whether values are based on *the most informative* (semantic interest map) or *locally informative* (meaning map) semantic information. Therefore, these two different approaches allow for the relationship between mind wandering and the prioritization of semantic information in fixation allocation to be explored in different ways.

In terms of hypotheses, the visual system may be less able to discriminate what scene content is semantically informative during mind wandering given perceptual decoupling, and, thus, gaze would less frequently fixate on highly informative scene content regardless of how semantic information is characterized. Alternatively, because visual and cognitive processing is only attenuated—not entirely eliminated—during mind wandering, it is possible that the visual system can still manage to detect and prioritize information that is the most semantically informative while neglecting less informative content. This idea predicts that fixations would be more likely—or at least equally likely—to occur within semantically informative regions during mind wandering especially when characterized by semantic interest maps. That said, if the meaning maps better

predict fixation allocation during mind wandering, findings would suggest that locally informative scene content remains detectable and perhaps becomes prioritized during mind wandering. No observable change in how gaze is allocated to semantically informative scene content during mind wandering would also suggest that the changes in the content-independent measures of gaze observed in Krasich et al. (2018) cannot be characterized as a systematic shift in how fixations are allocated to semantically informative scene content.

**Empirical approach.** Using the previously described models, visual salience and semantic informativeness scores were computed for each of the images used in Krasich et al. (2018). Then, scores for locations where fixations occurred in Krasich et al. (2018) were measured and compared across self-reports of mind wandering and attentive viewing, which were obtained via pseudo-randomly distributed thought probes.

## Methods

### Participants

Eye movement and mind wandering data were obtained from Krasich et al. (2018) which included 51 volunteers from the University of Notre Dame. All participants were compensated with course credit.

Semantic interest maps were generated with data collected from 31 college-aged students from the University of Notre Dame who did not participate in the Krasich et al. (2018) study. This sample size was established following a similar sample size ($n$ = 27) used in Tatler et al. (2017). Participants volunteered through the university psychology subject pool following procedures approved by the university IRB and received course credit for participation.

Meaning maps were generated with data collected from 150 volunteers from Amazon Mechanical Turk (MTurk) who had a hit approval rate of 95%, at least 100 hits approved, and were from the United States. This sample size was similar to that used in Henderson & Hayes (2017, 2018) ($n$ = 79; $n$ = 165, respectively). Participants volunteered through MTurk following procedures approved by the university IRB and were monetarily compensated for participation. 15 MTurk participants were removed for not properly completing the task (i.e., pressing the same response for all 300 patches). Therefore, ratings from 135 participants were used to generate the meaning maps.

### Stimuli and apparatus

The stimuli consisted of the twelve digitized color photographs of real-world urban scenes (800 pixels x 600 pixels) that were used in Krasich et al. (2018). Images were presented in 32-bit color on a 20-inch CRT monitor with a screen refresh rate of 85 Hz and a resolution of 1024 x 768. Examples images are shown in **Figure 1**.

Eye movements were sampled using an EyeLink 2K tower mounted eye tracking system (SR Research, Inc.) at a rate of 1,000 Hz. A viewing distance of 80 cm was maintained by a chin and forehead rest. Saccades were operationally defined as changes in recorded fixation position that exceeded 0.2° with either a velocity that exceeded 30°/s or an acceleration that exceeded 9,500°/s$^2$. The eye tracker was calibrated using a nine-point calibration at the beginning of the study and a

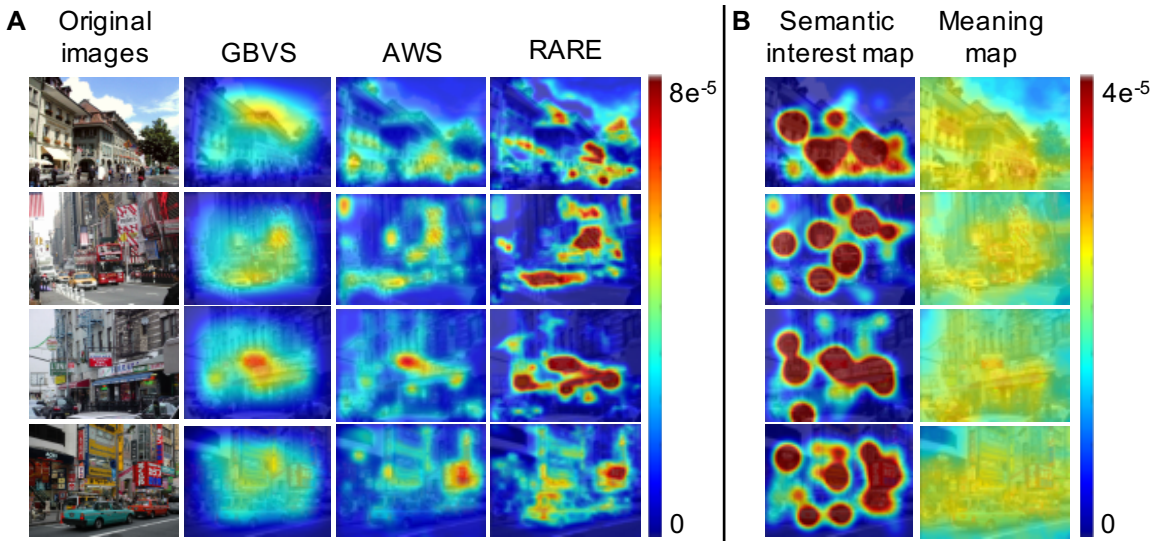



**Figure 1.** Example images and corresponding visual salience (A) and semantic maps (B). Note the difference in scale across the salience and semantic maps for visualization purposes.

one-point calibration before the presentation of each image to correct for any subtle drift in the eye tracker signal over time.

## Experimental Procedures

Participants from Krasich et al. (2018) were instructed to study twelve images for a later recognition test. They also received instructions for responding to the thought probes, and mind wandering (which was colloquially termed as "zoning out") was defined as the act of "looking at the picture but thinking of something else entirely" unrelated to task or the scene content. After receiving all instructions, participants viewed each image, presented in a different random order, sequentially for 45-75 s ($M$ = 60.0 s, $SD$ = 8.49 s). Eight thought probes were randomly presented at the end of eight trials. These probes asked, "In the moments right before this message, were you paying attention to the picture or were you zoning out?" (Schooler et al., 2004). Each participant received eight image-probe pairings, but the pairings differed across participants. Across participants, the resulting number of probes per image ranged from 28 to 41 ($Mdn$ = 34.5, $IQR$ = 29 – 36.5), with the number of reports of mind wandering per image ranging from 5 – 13 ($Mdn$ = 10, $IQR$ = 6 – 10.5). Therefore, mind wandering occurred with each of the images.

## Computing visually salience and semantically informative scene content

**Saliency maps.** The GBVS, AWS, and RARE models were used to generate saliency maps for each of the images used in Krasich et al. (2018). Example images and maps are illustrated in **Figure 1**. Each model computes an arbitrary "salience" value for each pixel, with greater values indicating greater salience. Following typical practice, values for each of these maps were then normalized so that the sum salience score of all pixels within an image was equal to 1.

*Table 1. Pearson Correlation coefficient matrix for salience and semantic scores for all images.*

| | AWS | RARE | Semantic map | Meaning map |
|---|---|---|---|---|
| GBVS | .55 | .52 | .47 | .54 |
| AWS | - | .75 | .53 | .61 |
| RARE | | - | .51 | .48 |
| Semantic interest map | | | - | .49 |

*Note*. Correlation coefficients represent a pixel by pixel comparison for all images

As reported in **Table 1**, salience scores computed by the GBVS were moderately correlated with scores from the AWS and RARE models, and scores from the AWS and the RARE models were strongly correlated. These findings indicate that although each model adopts different approaches for computing salience, a certain degree of similarity exists in how each map characterizes visually salient scene content. Measured coefficients, however, do indicate some variability across maps, which suggests an advantage to using multiple models to characterize visual salience.

**Semantic interest map.** The procedures for computing the semantic interest maps mirrored those used in Tatler et al. (2017). Participants viewed each full scene used in Krasich et al. (2018) and selected (with a mouse click) the five most semantically informative locations within the scene ignoring visual characteristics such as color, brightness, etc. (see **Appendix A** for full task instructions). "Semantically informative" locations were defined prior to the experiment as locations that are the most "informative about the meaning of the scene." Participants were able to reselect locations prior to moving on to the next trial but were not able to reselect locations after moving on to the next trial. The selections were used to create semantic interest maps in the same manner as Tatler et al. (2017)—centering Gaussians with full width at half maximum of two degrees around each selected location.[1] This approach computes an arbitrary value for each pixel of the image, with greater values indicating greater semantic informativeness. Values were normalized so that the sum score of all pixels was equal to 1 within each image.

**Meaning maps.** The procedures for computing the meaning maps were drawn from those used in Hayes & Henderson (2019). Each of the scenes was decomposed into partially overlapping circular patches with 3° ("fine" patches) and 7° ("coarse" patches) diameters. The full patch stimulus set consisted of 3,600 unique fine patches and 960 course patches for a total of 4,560 patches. Each subject rated 300 random patches (for a total of 40,500 ratings) without scene context with the instructions to assess the meaningfulness of each patch in terms of how informative or recognizable it was considered (see **Appendix B** for full task instructions). Specifically, participants rated the meaningfulness of the patch using a 6-point Likert scale (very low, low, somewhat low, somewhat high, high, very high). The ratings for each pixel at each scale (fine and coarse) were averaged to produce a fine and course rating map for each scene, which were then averaged together into a single map. This map was then smoothed with a Gaussian filter

[1] Exploratory analyses were also conducted using semantic interest maps that were generated centering Gaussians with full width at half maximum of three and seven degrees around each selected location, which were explored because of the similar patch sizes used to generate the meaning maps. The results, however, were consistent across maps generated with each Gaussian size, so only the results obtained using the procedures as described in Tatler et al. (2017) are reported and discussed.

(using the *imgaussfilt* function in MATLAB; Mathworks, Natick, MA). Lastly, values were normalized so that the sum meaning score of all pixels within each image was equal to 1.

Example semantic interest and meaning maps are illustrated in **Figure 1**. As reported in **Table 1**, scores across the two maps were moderately correlated, but the measured coefficients do indicate variability across the two maps, which suggests an advantage to investigating the spatial allocation of gaze using multiple models of semantic informativeness.

### Computing salience and semantic scores at fixated locations

The (x,y) coordinates were extracted for each fixation made by participants in Krasich et al., (2018). Then an area subtending 2° visual angle around each coordinate was established, and the mean and maximum salience and semantic informativeness scores among the pixels within each of these areas were calculated. This was done for several reasons. First, because of the physical characteristics of the human eye, visual acuity is best at central viewing—an area subtending roughly 2° visual angle—and a greater proportion of neurons within the primary visual cortex are devoted to processing central vision as opposed to the periphery. Accordingly, central vision is particularly apt for high-resolution visual analysis, and, therefore, measuring visual salience within a 2° area around a fixation provided insight into how the visual system prioritizes a high-resolution analysis of visual salience. Second, this approach also accommodated subtle instrument error (typically .15 deg) in saccade recording. Finally, computing the mean and maximum scores among the pixels surrounding fixated locations characterized the data in two different ways. Specifically, mean scores indicated the mass visual salience of a particular area while maximum scores indicated the highest pixel of salience within the fixated area. By computing both of these variables it could also be determined which measure might best predict gaze behavior during mind wandering. Average means and maximum values were then centered and scaled (*z*-scored) using the *scale* function in R (Becker et al., 1988).

Fixations that occurred outside of the scene boarders (3% of fixations), were shorter than 50 ms (2% of fixations), or longer than 10,000 ms (< .01% of fixations) were excluded. In total, 95% of all fixations were analyzed.

## Results

First, the frequency of reported mind wandering from Krasich et al. (2018) is provided for ease of interpreting the results. Then, the main research question regarding the relationship between mind wandering and fixation allocation to visually salient and semantically informative scene content is reported.

### Frequency and validation of mind wandering

Participants from Krasich et al. (2018) reported mind wandering on 27% of probes (*SD* = 22%). Eleven participants reported no instance of mind wandering and one participant reported mind wandering for all eight probes. This rate of mind wandering is within the range of rates typically reported in laboratory and field settings (Killingsworth & Gilbert, 2010; Seli et al., 2018; Smallwood & Schooler, 2015). Rates of mind wandering were correlated with worse performance on the memory tests, which further validated the self-reports. A more detailed description of

performance on the memory test and its relationship to mind wandering is found in Krasich et al. (2018).

**Fixation allocation prior to reported mind wandering**

Of the 408 trials (out of 612 total trials) that included a thought probe, only 2 trials were excluded due to tracking errors. Therefore, a total of 406 trials were analyzed. Fixation allocation for trials with reported mind wandering (107 trials) were compared to trials with reported attentive viewing (299 trials).

Analyses focused on those fixations that occurred 10 s prior to the thought probes. Krasich et al. (2018) observed the most robust changes in the spatial aspects of fixation allocation (i.e., increased fixation dispersion) associated with mind wandering 10 s prior to the self-report, and past research has suggested that, depending on the task, some gaze behaviors show stronger associations with mind wandering (Faber, Krasich et al., in press) and attentive viewing (Marsman et al., 2013; Unema et al., 2005) within smaller spans of time. Accordingly, mean and maximum salience and semantic scores of fixated locations were averaged across fixations that occurred 10 s prior to the thought probes.

Mixed-effect linear regression analyses were conducted using the *lme4* package in R (Bates et al., 2015) to model each dependent variable (average mean and maximum salience and semantic score as measured by the respective models) with *probe response* (two levels: paying attention [reference group] and mind wandering) and *image viewing time* (*z*-scored, Becker et al., 1988) as fixed-effects variables[2] and with *participant* and *image* as random effects. Standardized coefficients ($\beta$) were computed, which indicated the predicted change per unit (SD) in the dependent variable net of the other predictor variables. Wald chi-square ratios and *p*-values were also computed using the *Anova* function from the *car* package in R (Fox & Weisberg, 2011) with a Type II sum of squares to investigate the main effects of mind wandering controlling for covariates. Treatment contrasts were used for all comparisons across mind wandering and attentive viewing.

Because the three models of visual salience (GBVS, AWS, and RARE) were meant to measure the same construct, we elected to be conservative and correct for familywise error when analyzing each of the dependent variables (mean and maximum salience). Significance testing was, therefore, conducted using two-tailed tests with $\alpha$ set to .05, with Bonferroni correction. Thus, we rejected the null hypothesis when $p < .017$ (i.e., .05/3). We were similarly conservative in our analyses of mean and maximum semantic interest scores because we used two approaches to operationalize semantic content (semantic interest maps and meaning maps). We therefore rejected the null hypothesis in cases where $p < .025$ (i.e., .05/2).

Coefficients and test statistics for each predictor are reported in **Appendix C**. Test statistics most relevant to the effect mind wandering on fixations to visually salient and semantically informative scene content are reported in **Table 2**.

**Minimal detectable effect sizes.** We first assessed the sensitivity of our main study given its sample size by estimating the minimal detectable effect size (MDES) of mind wandering on each dependent variable net of the aforementioned covariates included in the mixed-effect linear regression analyses. The coefficient of the mind wandering fixed effect served as the effect size

[2] Across all analyses, there were no significant polynomial effects of image viewing time, thus the quadratic term was not included.

*Table 2. Standardized coefficients and test statistics assessing the main effect of mind wandering on the average mean and maximum salience and semantic scores of fixated locations.*

| | Main study | | | | | Joint Analyses | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *β* | *SE* | *χ2* | *p* | | *β* | *SE* | *χ2* | *p* | |
| **GBVS** | | | | | | | | | | |
| Avg mean | .076 | .105 | .529 | .467 | | .100 | .886 | 1.263 | .261 | |
| Avg maximum | .064 | .106 | .357 | .550 | | .093 | .089 | 1.086 | .297 | |
| **AWS** | | | | | | | | | | |
| Avg mean | .328 | .104 | 9.951 | .002 | * | .296 | .090 | 10.828 | .001 | * |
| Avg maximum | .343 | .102 | 11.377 | .001 | * | .299 | .087 | 11.789 | .001 | * |
| **RARE** | | | | | | | | | | |
| Avg mean | .281 | .108 | 6.763 | .009 | * | .292 | .093 | 9.890 | .002 | * |
| Avg maximum | .296 | .100 | 8.706 | .003 | * | .299 | .088 | 11.435 | .001 | * |
| **Semantic map** | | | | | | | | | | |
| Avg mean | .043 | .106 | .164 | .685 | | -.024 | .089 | .073 | .787 | |
| Avg maximum | .038 | .103 | .134 | .715 | | -.036 | .087 | .173 | .678 | |
| **Meaning map** | | | | | | | | | | |
| Avg mean | .174 | .091 | 3.69 | .055 | | .167 | .081 | 4.208 | .040 | |
| Avg maximum | .164 | .087 | 3.58 | .058 | | .143 | .077 | 3.419 | .064 | |

*Note: β* = standardized coefficients; *SE* = standard errors; *χ2* = chi-square ratios; degrees of freedom for all chi-square ratios = 1; * = statistically significant after Bonferroni adjustments (salience scores $p < .017$; semantic scores $p < .025$)

measure, and the magnitude of these coefficients reflected the estimated change in the dependent variables in SD units. Using the *simr* package in R (Green & MacLeod, 2016), a power analysis was conducted for each dependent variable to estimate the power associated with effect sizes ranging from .05 to .55 in increments of .05, α set to .05, and the number of simulations to 1,000. The lowest effect size that would on average yield a power of at least .80 was retained as the MDES. The average MDES was .305 and ranged from .250 - .350. These findings indicate .305 was on average the smallest effect detectable at power of .80. It further suggests that, at a power of .80, effect sizes smaller than .305 would on average require a larger sample size to achieve statistical significance. The specific MDES and associated power for each dependent variable are reported in **Appendix D.**

**The effect of mind wandering 10 s prior to thought probes.** When measuring visual salience with the GBVS, there was no effect of mind wandering on the average mean and maximum salience score of fixated locations. When measuring visual salience with the AWS, however, mind wandering was associated with greater average mean and maximum salience scores for fixated locations. The same observations held when measuring visual salience with the RARE. Together, these findings showed that, as measured by two (admittedly highly correlated) salience models, fixations made 10 s prior to self-reported mind wandering occurred in regions of higher visual salience than fixations made 10 s prior to reports of paying attention.

There was no effect of mind wandering on the average mean and maximum semantic informativeness score of fixated locations when measured by the semantic interest maps. This

*Table 3. Standardized coefficients and test statistics assessing the main effect of mind wandering on the average mean and maximum salience and semantic scores of fixated locations computed using randomly shuffled overlaid images.*

| | Main study: Shuffled Images | | | |
|---|---|---|---|---|
| | *β* | *SE* | *χ2* | *p* |
| **GBVS** | | | | |
| Avg mean | .119 | .111 | 1.147 | .284 |
| Avg maximum | .117 | .109 | 1.153 | .283 |
| **AWS** | | | | |
| Avg mean | .097 | .111 | .763 | .382 |
| Avg maximum | .055 | .109 | .251 | .616 |
| **RARE** | | | | |
| Avg mean | .129 | .108 | 1.423 | .233 |
| Avg maximum | .068 | .101 | .450 | .502 |
| **Semantic map** | | | | |
| Avg mean | .095 | .112 | .719 | .397 |
| Avg maximum | .052 | .110 | .221 | .638 |
| **Meaning map** | | | | |
| Avg mean | .139 | .111 | 1.565 | .211 |
| Avg maximum | .151 | .108 | 1.939 | .164 |

*Note. β* = standardized coefficients; *SE* = standard errors; *χ2* = chi-square ratios; degrees of freedom for all chi-square ratios = 1; The analysis assessing average mean salience score as measured by the RARE failed to converge. The *original image view time* variable was then removed, and the results from this revised analysis are reported here.

finding indicates a similar propensity during mind wandering and paying attention to look at scene content rated as the most semantically informative. When measuring semantic informativeness with the meaning maps, mind wandering tended to be associated with greater average mean and maximum scores, but this effect was not statistically significant. In light of our MDES analyses, though, a larger sample size could potentially yield statistically significant differences.

To increase confidence that the AWS- and RARE-based results were not spurious, we conducted an analysis in which we "shuffled" participants' fixation points (x, y coordinates) that were recorded while viewing each scene onto different randomly selected scenes. This process thereby broke the natural association between image content and fixated locations while retaining the given fixation pattern with its corresponding probe response (as well as continuing to incorporate natural biases in oculomotor behavior). Hence, in this analysis we would predict that no relationship between salience and mind wandering should be observed. We repeated the analyses conducted above after computing the average mean and maximum salience and semantic informativeness scores for each fixated location with respect to the overlaid shuffled images. We modeled each of these dependent variables using mixed-effect linear regression analyses with *probe response* (two levels: paying attention [reference group] and mind wandering) and *original image viewing time* (*z*-scored) as fixed-effects variables and with *participant* and *shuffled image* as random effects. The relevant coefficients and test statistics are reported in **Table 3**. Findings showed that across all models, the average mean and maximum salience and semantic scores did not differ across trials with reported mind wandering and attentive viewing. These analyses

*Table 4. Standardized coefficients and test statistics investigating the effect of mind wandering on visual salience within 10 s time windows with respect to probe onset in the main study*

| | **AWS: avg mean** | | | | **AWS: avg maximum** | | | |
|---|---|---|---|---|---|---|---|---|
| **Time before probe** | *β* | *SE* | *p* | | *β* | *SE* | *p* | |
| 0 s – 10 s | .292 | .104 | .005 | * | .316 | .103 | .002 | * |
| 10 s – 20 s | .080 | .104 | .443 | | -.002 | .102 | .987 | |
| 20 s – 30 s | .051 | .104 | .662 | | .049 | .102 | .633 | |
| 30 s – 40 s | -.160 | .104 | .123 | | -.132 | .102 | .196 | |
| | **RARE: avg mean** | | | | **RARE: avg maximum** | | | |
| **Time before probe** | *β* | *SE* | *p* | | *β* | *SE* | *p* | |
| 0 s – 10 s | .261 | .108 | .017 | | .266 | .100 | .008 | * |
| 10 s – 20 s | .129 | .108 | .232 | | .112 | .099 | .257 | |
| 20 s – 30 s | -.092 | .108 | .397 | | -.104 | .099 | .296 | |
| 30 s – 40 s | -.060 | .108 | .583 | | -.050 | .099 | .617 | |

*Note.* $\beta$ = standardized coefficients; *SE* = standard errors; * = statistically significant after Bonferroni adjustments for multiple comparisons ($p < .013$)

provide additional support for the conclusion that observers view scene content—as measured by two models of visual salience—differently when they are mind wandering versus paying attention.

**The effect of mind wandering across time.** The findings thus far have demonstrated that scene content fixated 10 s before reported mind wandering was more visually salient than the content that was fixated before reported attentive viewing. These findings corresponded to previously identified mind wandering-related changes in content-independent measures of gaze behavior within the same dataset (i.e., fewer, longer, and more dispersed fixations; Krasich et al., 2018) and thus suggest a shift in both what and how visual information was sampled. These analyses focused on the fixations 10 s prior to a thought probe because Krasich et al., (2018) had previously shown the most robust mind wandering-related changes in content-independent measures of gaze behaviors within this timeframe. Here, we report a secondary post-hoc analysis of the association between mind wandering and visual salience across viewing time, predicting that the relationship would dissipate further back in time before the mind wandering report.

We first created four 10 s time windows with respect to the onset of the thought probe (i.e., 40-30 s before probe, 30-20 s before probe etc.). Then within each window, we averaged the mean and maximum salience and semantic scores of fixated locations (*z*-scored). We then modeled each dependent variable as a *probe response* (two levels: paying attention [reference group] and mind wandering) by time window (four levels with 10-0 s before probe as the reference) interaction with *image viewing time* (*z*-scored, Becker et al., 1988) as fixed-effect variable and with *participant* and *image* as random effects. Significance testing was conducted using two-tailed tests with $\alpha$ set to .05, with Bonferroni corrections (i.e., visual salience: $p < .017$; semantic informativeness: $p < .025$).

We did not observe any significant mind wandering by time window interactions in scores from the GBVS model, semantic maps, or meaning maps (all *p*'s > .150). This indicates that the previously observed null effects of mind wandering with respect to these dependent variables in the 10 seconds prior to reports of MW were consistent across viewing time. There were, however, trending or significant interactions when visual salience measured by the AWS (mean: $\chi 2$ = 10.048, $p$ = .018; maximum: $\chi 2$ = 10.498, $p$ = .015) and RARE models (mean: $\chi 2$ = 7.392, $p$ =

.060; maximum: $\chi 2$ = 8.979, $p$ = .031). We followed up these interactions with pairwise comparisons within each time window using the *emmeans* package in R (Lenth, 2017) and again controlled for multiple comparisons with Bonferroni corrections ($p < .013$; i.e., .05/4). The effect of mind wandering could only be observed in the 10 s prior to the mind wandering report (see Table 4). Any timeframe more than 10 s prior showed no effect of mind wandering.

**Conceptual Replication and Joint-Experiment Analyses**

As we noted in the Introduction, Krasich et al., (2018) reported a separate successful conceptual replication of their main study that showed clear and robust mind wandering-related changes in content-independent measures of gaze behavior (i.e., fewer, longer, and more dispersed fixations). In this study, a different group of 41 participants completed a scene memorization task that was embedded within a larger task battery (Faber, Krasich et al., in press). Participants studied six contiguously presented images of urban scenes for 60 s each in preparation for a later memory test.[3] Thought probes were presented in pseudorandom time intervals of 90-120 s ($M$ = 33.03 s, $SD$ = 15.58 s) such that they occurred mid-viewing, and participants received a total of three thought probes each. Participants reported MW on average 47% of thought probes ($SD$ = 50%).

We endeavored to use this replication experiment to verify our findings related to visual salience and semantic informativeness described above. Visual salience maps for each image were generated using the GBVS, AWS, and RARE models. Semantic interest maps were generated using ratings from a new sample of 28 laboratory participants, and meaning maps were generated using a new sample of 150 MTurk participants following the same procedures used in the main study.[4] The average mean and maximum salience and semantic scores of fixated locations were computed across the fixations that occurred 10 s prior to probe onset.[5] Each of these dependent variables were modelled using linear mixed-effect regression models as a *probe response* (two levels: paying attention [reference group] and mind wandering) and *image viewing time* ($z$-scored, Becker et al., 1988) as fixed-effects variables and with *participant* and *image* as random effects. Because this task was randomly embedded within a larger task battery, we also included its *task order* as a categorical fixed-effect covariate.

As with the main study, we first assessed the sensitivity of this replication study with respect to the analyses of interest by estimating the MDES of mind wandering on each dependent variable following similar procedures as in the main study. The specific MDES and associated power for each dependent variable are reported in **Appendix D**. The average MDES was .50 and ranged from .45 - .50, which indicates that this study was on average only powerful enough to detect an average effect sizes of .50 and greater. The largest effect size observed in the analysis of the main study, however, was .343, indicating that the replication was not suitably powered to assess the effect of mind wandering on content-dependent behaviors (which were substantially smaller than the effects observed using content-independent measurements by Krasich et al., 2018). As a result, it would not be surprising to fail to replicate our previously observed findings related to mind wandering

[3] The images used in this study included six images from the main study that were first cropped then expanded (893 pixels x 1585 pixels) to standardize the viewing conditions across the entire task battery (see **Appendix E** for the images).

[4] 8 MTurk participants were removed for not properly completing the task (i.e., pressing the same response for all 300 patches).

[5] Fixations that occurred outside of the scene boarders (2% of fixations) and/or were shorter than 50 ms (2% of fixations) were excluded. No fixations were longer than 10,000 ms. Therefore, 96% of total fixations were analyzed.

and visual salience within this data set (indeed, analysis of the replication data returned universally null results; coefficients and test statistics for each predictor are reported in **Appendix F**).

We can, however, make use of the replication from Krasich et al. (2018) by combining it with the main study in a set of joint-experiment analyses.[6] In doing so, we can ensure the relationship between mind wandering and visual salience holds when additional data, collected from a different group of participants in a different experimental context, is also considered. We can also determine if effects of semantic informativeness in the main study emerge in a more powerful statistical analysis. For these analyses, we modeled each dependent variable (*z*-scored by experiment) using mixed-effect linear regressions with *probe response* (two levels: paying attention [reference group] and mind wandering), *image viewing time* (*z*-scored by experiment), and *experiment* (two levels: main study [reference group] and replication) as fixed-effect variables and with *participant* and *image* as random effects. Bonferroni corrections were again incorporated to account for familywise error.

**MDES.** The MDES for these joint-experiment analyses were estimated following similar procedures as the main study. The specific MDES and associated power for each dependent variable are reported in **Appendix D**. The average MDES was .27 and ranged from .25 - .30, which shows improved sensitivity over the main study.

**The effect of mind wandering on fixations 10 s prior to thought probes.** Coefficients and test statistics for each predictor are reported in **Appendix G**, and the most relevant test statistics are reported in **Table 2**. The findings showed that when measuring visual salience with the GBVS, there was still no effect of mind wandering on the average mean and maximum salience score of fixated locations. Mind wandering was, however, associated with greater average mean and maximum AWS and RARE salience scores. These effects survived Bonferroni corrections and were stronger than those observed in the main study. There was no effect of mind wandering on fixations to semantically informative scene content as measured by the semantic interest maps, but when measuring semantic informativeness with the meaning maps, mind wandering tended to be associated with greater scores. These effects were not statistically significant, however, especially after correcting for multiple comparisons.

We further quantified the observed effects of mind wandering by comparing the aforementioned regression analyses with baseline models that predicted salience and semantic scores without probe response as a predictor variable. That is, these baseline models included only *image viewing time* (*z*-scored by experiment) and *experiment* (two levels: main study [reference group] and replication) as fixed-effect variables and *participant* and *image* as random effects. The findings for these model comparisons are reported in **Table 5**. They showed that including probe response as a predictor variable significantly improved the quality of the baseline models predicting average mean and maximum AWS and RARE scores, as indicated by lower AIC/BIC values and significantly different deviances. Including probe response as a predictor variable did not significantly improve the models predicting average mean and maximum GBVS, semantic map, and meaning map scores. These findings further indicate a link between mind wandering at the propensity to fixate on visually salient scene content.

Considered collectively, findings from the joint-experiment analysis were consistent with those observed in the main study: mind wandering was most associated with fixations to scene content that was more visually salient than fixated content before reports of attentive viewing.

[6] We thank an anonymous reviewer for suggesting this analysis.

*Table 5. Test statistics exploring unique variance in mind wandering explained by visual salience and semantic informativeness*

| | | *df* | AIC | BIC | LLV | deviance | $\chi 2(1)$ | *p*-value | |
|---|---|---|---|---|---|---|---|---|---|
| **GBVS** | | | | | | | | | |
| *Avg mean* | | | | | | | | | |
| | Baseline | 6 | 1402 | 1427 | -695 | 1390 | | | |
| | w/ Probe response | 7 | 1402 | 1432 | -694 | 1388 | 1.258 | .262 | |
| *Avg maximum* | | | | | | | | | |
| | Baseline | 6 | 1405 | 1430 | -696 | 1393 | | | |
| | w/ Probe response | 7 | 1406 | 1435 | -696 | 1392 | 1.079 | .299 | |
| **AWS** | | | | | | | | | |
| *Avg mean* | | | | | | | | | |
| | Baseline | 6 | 1432 | 1458 | -710 | 1420 | | | |
| | w/ Probe response | 7 | 1423 | 1453 | -705 | 1409 | 10.783 | .001 | * |
| *Avg maximum* | | | | | | | | | |
| | Baseline | 6 | 1402 | 1428 | -695 | 1390 | | | |
| | w/ Probe response | 7 | 1392 | 1422 | -689 | 1378 | 11.695 | .001 | * |
| **RARE** | | | | | | | | | |
| *Avg mean* | | | | | | | | | |
| | Baseline | 6 | 1465 | 1491 | -726 | 1453 | | | |
| | w/ Probe response | 7 | 1458 | 1487 | -722 | 1444 | 9.810 | .002 | * |
| *Avg maximum* | | | | | | | | | |
| | Baseline | 6 | 1424 | 1450 | -706 | 1412 | | | |
| | w/ Probe response | 7 | 1415 | 1445 | -700 | 1401 | 11.259 | .001 | * |
| **Semantic map** | | | | | | | | | |
| *Avg mean* | | | | | | | | | |
| | Baseline | 6 | 1410 | 1436 | -699 | 1398 | | | |
| | w/ Probe response | 7 | 1412 | 1442 | -699 | 1398 | .066 | .797 | |
| *Avg maximum* | | | | | | | | | |
| | Baseline | 6 | 1393 | 1419 | -691 | 1381 | | | |
| | w/ Probe response | 7 | 1395 | 1425 | -691 | 1381 | .160 | .689 | |
| **Meaning map** | | | | | | | | | |
| *Avg mean* | | | | | | | | | |
| | Baseline | 6 | 1325 | 1350 | -656 | 1313 | | | |
| | w/ Probe response | 7 | 1323 | 1352 | -654 | 1309 | 4.217 | .040 | |
| *Avg maximum* | | | | | | | | | |
| | Baseline | 6 | 1275 | 1300 | -631 | 1263 | | | |
| | w/ Probe response | 7 | 1273 | 1303 | -630 | 1259 | 3.433 | .064 | |

*Note:* AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LLV = Log-Likelihood Value; $\chi 2$ = chi-square ratios; * = statistically significant after Bonferroni adjustments (salience scores $p < .017$; semantic scores = $p < .025$)

## General Discussion

The current work focused on visual and cognitive factors that have been linked to gaze control and how these relationships vary across attentional states. Previous work has shown that mind wandering is associated with fewer, longer, and more dispersed fixations in a scene memorization task, with the most robust effects occurring 10 s prior to reported mind wandering (Krasich et al., 2018). Our goal in this report was to determine if such shifts in gaze behavior were characterized by content-dependent adjustments to gaze control mechanisms. To do so, we assessed how the visual system samples visually salient and semantically informative scene content during mind wandering.

We operationalized mind wandering as moments directly prior to when participants reported not being focused on the scene memorization task, and thus to some degree, were perceptually decoupled from the processing of the external world (e.g., Schooler et al., 2011; Murphy et al., 2019). Mind wandering was self-reported following occasional thought probes that queried the focus of participants' attention. We then compared the associations between fixation location, visual salience, and semantically informative content within the scene prior to probes where participants indicated that they were paying attention to their scene memorization task and where participants admitted to mind wandering.

Our main study revealed an increased propensity for fixated scene content to be more visually salient (as measured by two of the three models used to operationalize salience) in the 10 s prior to reported mind wandering compared to reported attentive viewing. This time window corresponds to that in which changes in content-independent measures of gaze behavior were previously observed (Krasich et al., 2018). This fixated scene content also tended to be more semantically informative, when operationalized in terms of local identifiability, but the effect of mind wandering was not statistically significant. No differences were observed across attentional states when semantic content took into account the importance of local scene regions to the overall scene content. As such, findings from the main study indicate that gaze was directed to scene content that was more visually salient during mind wandering compared to attentive viewing. This suggests that changes in the spatial aspects of gaze during mind wandering reflect a content-dependent shift in what visual information is sampled.

Unfortunately, a conceptual replication provided by Krasich et al., (2018) was underpowered with respect to the effect sizes observed for visual salience, even though this same replication study was able to reveal strong mind wandering-related changes in content-independent measures of gaze behaviors (i.e., fewer, longer, and more dispersed fixations). That said, combining the data from the main study and the replication study yielded a more powerful statistical analysis in which all effects from the main study alone were maintained. Furthermore, we were able to show with this analysis that the ability to predict what content will be fixated by an observer is improved by knowing the observer's attentional state while viewing the scene.

Both the main study and the joint-experiment analyses also highlight a contrast in magnitude between the smaller association between mind wandering and what information is viewed (i.e., more salient regions) versus the much larger association between mind wandering and how information is viewed (e.g., more slowly). From a theoretical point of view, this suggests that the link between mind wandering and changes in local gaze behaviors are weaker, more fragile, and/or more sensitive to task-specific idiosyncrasies than content-independent measures of gaze behaviors. For example, the effects of mind wandering were only observed in two of the three models of visual salience, indicating that at least the idiosyncratic procedures for computing

salience characterizes the mind wandering-salience link. That is, the AWS and RARE models reflect contrasts relative to the entire image and do not incorporate a center-bias. The GBVS model characterizes salience in terms of difference across local regions and does favor regions centrally located. Although it is unclear which computational difference across these models best characterizes the mind wandering-visual salience link, our findings do indicate there is nuance in this relationship. This nuance requires further exploration, but it does suggest that content-independent gaze measures may provide a more efficacious set of parameters for identifying mind wandering across a range of contexts and tasks.

Future work is certainly needed to establish the link between gaze, scene content, and mind wandering, as well as whether, and to what extent, stimulus-specific or task-specific idiosyncrasies might influence these effects. The stimuli used in this study were admittedly few in number (12) and restricted in range (urban scenes) and beyond memorization tasks like the one used here, observers have many different goals when viewing or interacting with visual information. Thus, the extent to which the relationship between gaze and mind wandering may be modulated by exposure to different scenes, tasks, or intensions remain interesting questions. Despite these limitations, however, a clear message emerges from our data: contemporary frameworks of gaze control are incomplete, and explanatory models of gaze need to account for both shifts in sampling rate (i.e., longer fixations) and shifts in the kind of information that is sampled (i.e., higher salience, higher local semantics) during mind wandering.

While the exact nature of these mechanistic changes will require a great deal of additional work, our results give us an important first look into new ways of thinking about gaze and attention during mind wandering. Our data, for example, suggest that gaze control mechanisms may "rebalance" salient and semantically informative information during mind wandering. As attention shifts away from in-depth visual processing, gaze is more likely to be directed toward scene content that is visually distinct and "stands out," and less time is spent interrogating visually indistinct or difficult to interpret scene regions. Our findings are also consistent with the levels of inattention hypotheses derived from studies of mindless reading (Schad, Nuthmann, & Engbert, 2012). The levels of inattention hypothesis conceptualizes mind wandering as a matter of degree where "weak" and "deep" mind wandering have different effects on gaze. During deep mind wandering both low- and high-level processes are decoupled, whereas during weak mind wandering high-level processing is decoupled but low-level processing is intact. The shift in fixations toward salient information in our study may reflect weak mind wandering, where low-level properties become more important in the absence of higher-level cognition. Thus, the shift from weak to deep mind wandering may constitute the basis for our proposed rebalance of information that influences gaze control as mind wandering occurs.

An alternative account posits that the visual system may operate following similar principles across bouts of attentive viewing and mind wandering, but with an inefficiency that decreases sampling rate (i.e., fewer and longer fixations) and elicits a sort of exploration-exploitation tradeoff (e.g., Jepma & Nieuwenhuis, 2011) reflected by an increase in fixation dispersion (Faber, Krasich, in press, Krasich et al., 2018). Moreover, increased noise or variability in gaze control may inconsistently give rise to content-dependent changes or unspecified changes not directly captured by visual salience or semantic informativeness. Future work should further discern the relationship between mind wandering and content-dependent factors aside from visual salience and semantic informativeness to further assess this possibility.

In conclusion, everyday thoughts frequently consist of mind wandering, during which visual and cognitive processing is attenuated. Corresponding changes in gaze behaviors suggest a shift

in how the visual system samples information in light of perceptual decoupling thought to occur during mind wandering. This reflects a prioritization of visually salient scene content (and perhaps local semantics), although this effect may be sensitive to task-specific as well as mind wandering-specific idiosyncrasies. Theoretical frameworks and computational models of gaze control that consider fixation allocation should account for these changes in gaze associated with mind wandering for a comprehensive account of the visual processing priorities of the visual system across various attentive states. Doing so would inform applied efforts to predict and detect mind wandering in real-time (e.g., Hutt et al., 2019), disentangling how the specific content of a scene should be considered or whether focusing on content-independent measures of gaze behavior is optimal.

## References


Baird, B., Smallwood, J., & Schooler, J. W. (2011). Back to the future: autobiographical planning and the functionality of mind-wandering. *Consciousness and cognition*, *20*(4), 1604-1611.

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235-1243.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1– 48.

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision research*, *50*(23), 2577-2587.

Borji, A., Sihite, D. N., & Itti, L. (2013a). Objects do not predict fixations better than early saliency: A re-analysis of Einhäuser et al.'s data. *Journal of vision*, *13*(10), 18-18.

Borji, A., Sihite, D. N., & Itti, L. (2013b). What stands out in a scene? A study of human explicit saliency judgment. *Vision research*, *91*, 62-77.

Buswell, G. T. (1935). How people look at pictures: a study of the psychology and perception in art. Chicago: University of Chicago Press.

Campbell, F. W., & Wurtz, R. H. (1978). Saccadic omission: why we do not see a grey-out during a saccadic eye movement. *Vision research*, *18*(10), 1297-1303.

Christoff, K., Irving, Z. C., Fox, K. C., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nature Reviews Neuroscience*, *17*(11), 718.

Christoff, K., Mills, C., Andrews-Hanna, J. R., Irving, Z. C., Thompson, E., Fox, K. C., & Kam, J. W. (2018). Mind-wandering as a scientific concept: cutting through the definitional haze. *Trends in cognitive sciences*, *22*(11), 957-959.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18-18.

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of vision*, *8*(3), 3-3.

Escera, C., Alho, K., Schröger, E., & Winkler, I. W. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neurotology*, *5*(3-4), 151-166.

Faber, M., & D'Mello, S. K. (2018). How the stimulus influences mind wandering in semantically rich task contexts. *Cognitive Research: principles and implications*, *3*(1), 35.

Faber, M., Bixler, R., & D'Mello, S. K. (2018). An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*, *50*(1), 134-150.

Faber, M., Krasich, K., Bixler, R. E., Brockmole, J. R., & D'Mello, S. K. (in press). The eye-mind wandering link: Identifying gaze indices of mind wandering across tasks. *Journal of Experimental Psychology: Human Perception and Performance*.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. Thousand Oaks, CA: Sage Publications.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience*, *11*(2), 127.

Garcia-Diaz, A., Leboran, 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.

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. In *Advances in neural information processing systems* (pp. 545-552).

Hayes, T. R., & Henderson, J. M. (2019). Scene semantics involuntarily guide attention during visual search. *Psychonomic bulletin & review*, *26*(5), 1683-1689.

Henderson, J. M. (2017). Gaze control as prediction. *Trends in cognitive sciences*, *21*(1), 15-23.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In *Eye movements* (pp. 537-562).

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, *1*(10), 743.

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

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic bulletin & review*, *16*(5), 850-856.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. (1973). Electrical signs of selective attention in the human brain. *Science. 182*(4108), 177–180.

Hutt, S., Krasich, K., Mills, C., Bosch, N., White, S., Brockmole, J. R., & D'Mello, S. K. (2019). Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction*, *29*(4), 821-867.

Irwin, D. E., & Brockmole, J. R. (2004). Suppressing where but not what: The effect of saccades on dorsal-and ventral-stream visual processing. *Psychological Science*, *15*(7), 467-473.

Irwin, D. E., & Carlson-Radvansky, L. A. (1996). Cognitive suppression during saccadic eye movements. *Psychological Science*, *7*(2), 83-88.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10-12), 1489-1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, *2*(3), 194.

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

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, *8*(4), 441-480.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities?. *Vision Research*, *41*(25), 3559-3565.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*(6483), 742-744.

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.

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*(4), 279-284.

Kam, J. W., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., & Handy, T. C. (2011). Slow fluctuations in attentional control of sensory cortex. *Journal of Cognitive Neuroscience*, *23*(2), 460-470.

Kam, J. W., & Handy, T. C. (2018). Electrophysiological Evidence for Attentional Decoupling during Mind-Wandering. *The Oxford Handbook of Spontaneous Thought: Mind-Wandering, Creativity, and Dreaming*, 249.

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932-932.

Knight, R. T. (1997). Distributed cortical network for visual attention. *Journal of Cognitive Neuroscience*, *9*(1), 75-91.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision research*, *35*(13), 1897-1916.

Krasich, K., McManus, R., Hutt, S., Faber, M., D'Mello, S. K., & Brockmole, J. R. (2018). Gaze-based signatures of mind wandering during real-world scene processing. *Journal of Experimental Psychology: General*, *147*(8), 1111.

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

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*(1), 111-151.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*(3), 165-188.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception*, *26*(8), 1059-1072.

Marsman, J.-B. C., Renken, R., Haak, K. V., & Cornelissen, F. W. (2013). Linking cortical visual processing to viewing behaviour using fMRI. *Frontiers in System Neuroscience*, *7*, 109.

Matin, E. (1974). Saccadic suppression: a review and an analysis. *Psychological bulletin*, *81*(12), 899.

Mills, C., Raffaelli, Q., Irving, Z. C., Stan, D., & Christoff, K. (2018). Is an off-task mind a freely-moving mind? Examining the relationship between different dimensions of thought. *Consciousness and cognition*, *58*, 20-33.

Murphy, C., Jefferies, E., Rueschemeyer, S. A., Sormaz, M., Wang, H. T., Margulies, D. S., & Smallwood, J. (2018). Distant from input: Evidence of regions within the default mode network supporting perceptually-decoupled and conceptually-guided cognition. *Neuroimage*, *171*, 393-401.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of vision*, *10*(8), 20-20.

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding V1?. *Neural computation*, *17*(8), 1665-1699.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107-123.

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial vision*, *16*(2), 125-154.

Polich, J. (2003). Theoretical overview of P3a and P3b. In J. Polich (Ed.), *Detection of change: Event-related potential and fMRI findings* (pp. 83–98). Boston, MA: Kluwer Academic Press.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, *10*(4), 341-350.

Riche, N., Mancas, M., Duvinage, M., Mibulumukini, M., Gosselin, B., & Dutoit, T. (2013). Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, *28*(6), 642-658.

Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your Mind Wanders Weakly, Your Mind Wanders Deeply: Objective Measures Reveal Mindless Reading at Different Levels. *Cognition*, *125*(2), 179–194.

Schooler, J. W., Reichle, E. D., & Halpern, D. V. (2004). Zoning-out during reading: Evidence for dissociations between experience and meta-consciousness. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 204–226). Cambridge, MA: MIT Press.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A.(2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, *15*(7), 319-326.

Seli, P., Beaty, R. E., Cheyne, J. A., Smilek, D., Oakman, J., & Schacter, D. L. (2018). How pervasive is mind wandering, really?. *Consciousness and cognition*, *66*, 74-78.

Seli, P., Kane, M. J., Metzinger, T., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., & Smilek, D. (2018a). The family-resemblances framework for mind-wandering remains well clad. *Trends in cognitive sciences*, *22*(11), 959-961.

Seli, P., Kane, M. J., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., & Smilek, D. (2018b). Mind-wandering as a natural kind: A family-resemblances view. *Trends in cognitive sciences*, *22*(6), 479-490.

Seli, P., Risko, E. F., & Smilek, D. (2016). On the necessity of distinguishing between unintentional and intentional mind wandering. *Psychological science*, *27*(5), 685-691.

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments?. *Vision Research*, *41*(25), 3535-3545.

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *Journal of Cognitive Neuroscience*, *20*(3), 458-469.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*, 487-518.

Spain, M., & Perona, P. (2011). Measuring and predicting object importance. *International Journal of Computer Vision*, *91*(1), 59-76.

Steindorf, L., & Rummel, J. (2019). Do your eyes give you away? A validation study of eye-movement measures used as indicators for mindless reading. *Behavior research methods*, 1-15.

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., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision research*, *45*(5), 643-659.

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A Model of Saccadic Decisions in Space and Time. *Psychological Review*, *124,* 267-300.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494.

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 Vision*, *9*(3), 24-24.

Yarbus, A. L. (1967). *Eye movements during perception of complex objects*. New York: Plenum Press.

Zuber, B. L., & Stark, L. (1966). Saccadic suppression: elevation of visual threshold associated with saccadic eye movements. *Experimental neurology*, *16*(1), 65-79.

## Appendix A

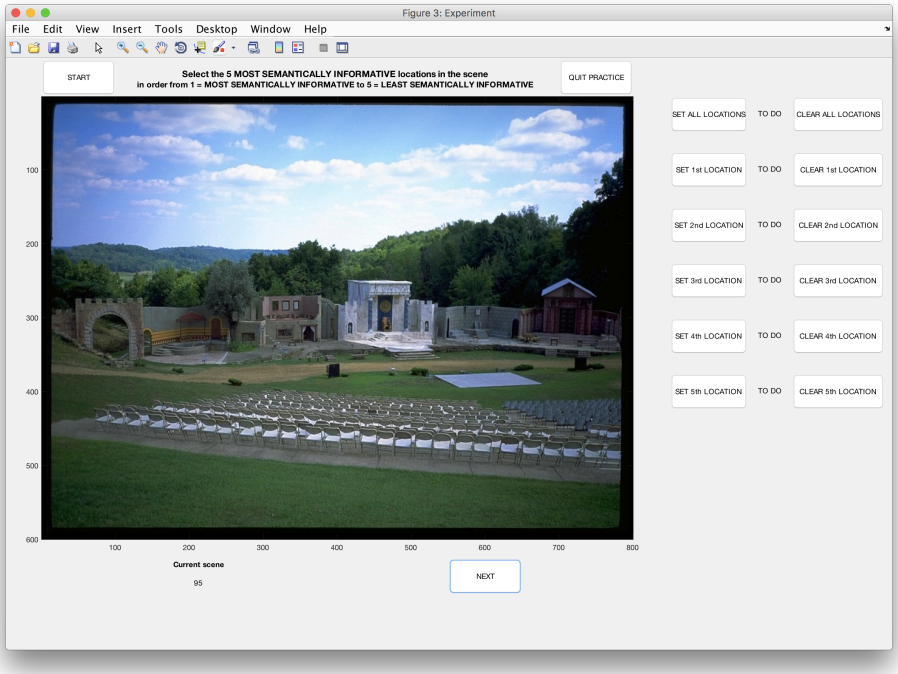
*Instructions for semantic informative ratings used to generate the semantic interest maps.*

In the following experiment you will be asked to select the locations that you feel are the most SEMANTICALLY INFORMATIVE in each scene you see. That is, you should select the locations that are the most informative about the meaning of the scene you are viewing.

Please try to ignore visual characteristics like brightness, colour, size etc, and base your selections on the importance of each location for the meaning of the scene.

For each scene you will be asked to select FIVE locations and to select these in the order of the MOST SEMANTICALLY INFORMATIVE of the five, to the LEAST SEMANTICALLY INFORMATIVE of the five you select.

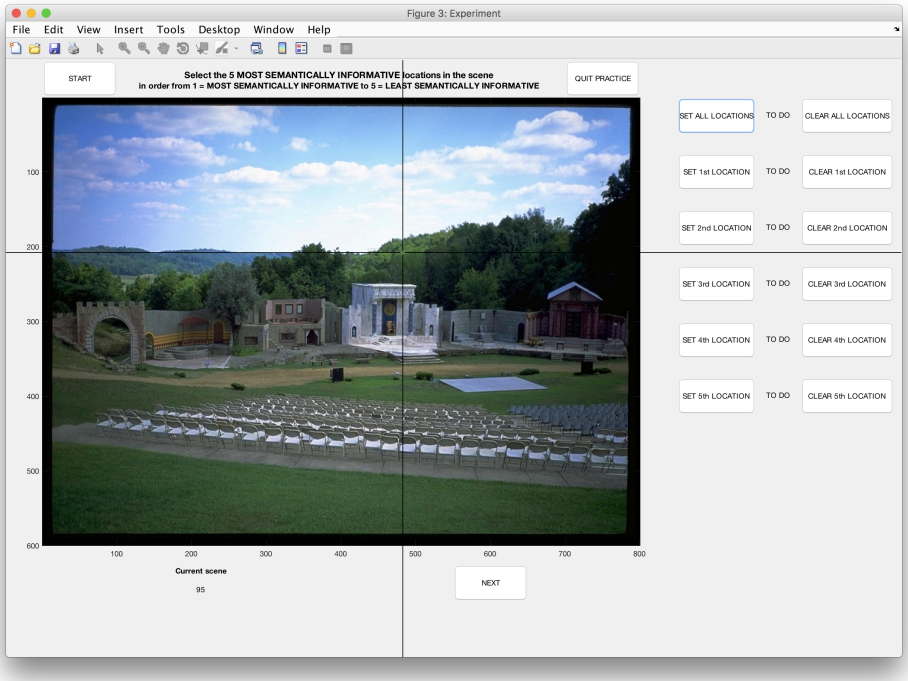
In the experiment you will see the following screen:




To display the first scene click START.

To select the locations, click on the SET ALL LOCATIONS button.

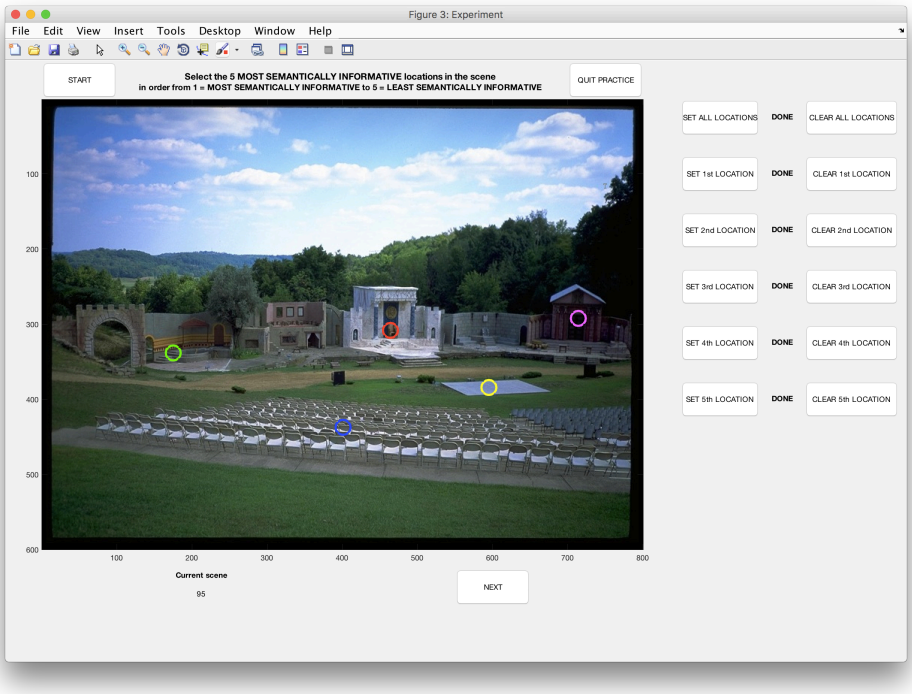
You will then see a cross hair as shown below:



Click on the location that you feel is the MOST SEMANTICALLY INFORMATIVE. Your selection will be shown with a red circle.

Then click in turn on the locations you feel are the 2nd, 3rd, 4th and 5th MOST SEMANTICALLY INFORMATIVE.

Each selection is shown by a different coloured circle once you have selected it.




If you are unhappy with any of your selections you can use the other buttons on the right to clear and re-set individual selections.

When you are happy with all 5 selections click the NEXT button to display the next scene.

You will be given a chance to practice this procedure by the experimenter. If you have any questions then please ask the experimenter before beginning the main experiment

## Appendix B

*Instructions for semantic informative ratings used to generate the meaning maps.*

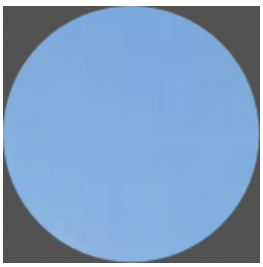
The purpose of this study is to gain a better understanding of how people perceive real-world visual scenes like this:

Real-world visual scene example

**Procedures**

If you agree to take part in this study, you will be presented with a series of images which are small patches of larger real-world scenes like the one above.
Your task will be to rate how "meaningful" you think each scene patch is.
What do we mean by "meaningful"? We want you to assess how "meaningful" an image is based on how informative or recognizable you think it is. For example, here are two scene patches taken from the example scene above that would be very low meaning.

**Edge of sidewalk**

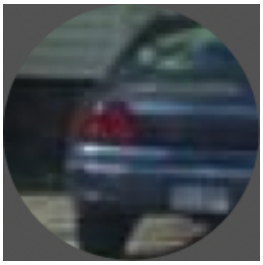


**Sky above house**

Without the example scene it would be difficult to recognize what either of these image patches are.

And here are two example patches that would be very high meaning.

**Window and roof**



**Back of car**

Both of these patches contain information that is easily recognized even without the example scene.

You will be asked to rate how "meaningful" you think each scene patch is using a 6- point scale. A rating of 1 means you think the scene patch is very low meaning like the sky example. A rating of 6 means you think the scene patch is very high meaning like the car example. The 6-point scale will look like this:

| | Very Low 1 | Low 2 | Somewhat Low 3 | Somewhat High 4 | High 5 | Very High 6 |
|---|---|---|---|---|---|---|
| Meaningfulness Rating | ○ | ○ | ○ | ○ | ○ | ○ |

You will select your answer by using the mouse to click on the bubble below the rating you wish to select. The task consists of 300 scene patches and will take approximately 20 minutes or less. This study will be conducted with an online Qualtrics-created survey.

# Appendix C

*Table C1. Coefficients for all variables in the regression models assessing the mean visual salience of fixated scene content prior to self-reported mind wandering for the main study.*

| | GBVS | | | | AWS | | | | RARE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | | ***β*** | ***95% CI*** | ***p*** | | ***β*** | ***95% CI*** | ***p*** | |
| (Intercept) | -.028 | -.301 – .246 | .843 | | .096 | -.354 – .163 | .469 | | -.072 | -.290 – .145 | .515 | |
| Probe Response [MW] | .076 | -.129 – .282 | .467 | | .328 | .124 – .532 | .002 | * | .281 | .069 – .493 | .009 | * |
| View Time | .046 | -.042 – .134 | .304 | | .016 | -.074 – .105 | .731 | | -.035 | -.128 – .058 | .465 | |
| ***Random Effects*** | | | | | | | | | | | | |
| $\sigma^2$ | .73 | | | | .81 | | | | .89 | | | |
| $\tau_{00}$ | .11 $_{ID}$ | | | | .01 $_{ID}$ | | | | .00 $_{ID}$ | | | |
| | .18 $_{Image}$ | | | | .17 $_{Image}$ | | | | .11 $_{Image}$ | | | |
| ICC | .28 | | | | .19 | | | | .11 | | | |
| N | 12 $_{Image}$ | | | | 12 $_{Image}$ | | | | 12 $_{Image}$ | | | |
| | 51 $_{ID}$ | | | | 51 $_{ID}$ | | | | 51 $_{ID}$ | | | |
| Observations | 406 | | | | 406 | | | | 406 | | | |
| Marginal $R^2$ / Conditional $R^2$ | .003 / .284 | | | | .021 / .206 | | | | .016 / .127 | | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance; ICC = intraclass correlation coefficient; * = statistically significant after Bonferroni adjustments (salience scores $p < .017$; semantic scores $p < .025$)

*Table C2. Coefficients for all variables in the regression models assessing the mean semantic informativeness of fixated scene content prior to self-reported mind wandering for the main study.*

| | Semantic map | | | Meaning map | | |
|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** |
| (Intercept) | .008 | -.266 – .283 | .952 | -.058 | -.419 – .304 | .754 |
| Probe Response [MW] | .043 | -.164 – .250 | .685 | .174 | -.004 – .352 | .055 |
| View Time | -.025 | -.114 – .064 | .584 | .017 | -.059 – .092 | .666 |
| ***Random Effects*** | | | | | | |
| $\sigma^2$ | .76 | | | .53 | | |
| $\tau_{00}$ | .09 $_{ID}$ | | | .15 $_{ID}$ | | |
| | .18 $_{Image}$ | | | .35 $_{Image}$ | | |
| ICC | .26 | | | .49 | | |
| N | 12 $_{Image}$ | | | 12 $_{Image}$ | | |
| | 51 $_{ID}$ | | | 51 $_{ID}$ | | |
| Observations | 406 | | | 406 | | |
| Marginal $R^2$ / Conditional $R^2$ | .001 / .264 | | | .006 / .490 | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance; ICC = intraclass correlation coefficient

*Table C3. Coefficients for all variables in the regression models assessing the maximum visual salience of fixated scene content prior to self-reported mind wandering for the main study*

| | GBVS | | | | AWS | | | | RARE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | | ***β*** | ***95% CI*** | ***p*** | | ***β*** | ***95% CI*** | ***p*** | |
| (Intercept) | -.027 | -.292 – .239 | .843 | | -.096 | -.376 – .190 | .469 | | -.069 | -.376 – .237 | .657 | |
| Probe Response [MW] | .064 | -.145 – .272 | .550 | | .343 | .144 – .543 | .001 | * | .296 | .100 – .493 | .003 | * |
| View Time | .045 | -.044 – .135 | .320 | | .008 | -.079 – .096 | .854 | | -.032 | -.118 – .055 | .472 | |
| ***Random Effects*** | | | | | | | | | | | | |
| $\sigma^2$ | .77 | | | | .77 | | | | .77 | | | |
| $\tau_{00}$ | .09 $_{ID}$ | | | | .01 $_{ID}$ | | | | .00 $_{ID}$ | | | |
| | .17 $_{Image}$ | | | | .22 $_{Image}$ | | | | .26 $_{Image}$ | | | |
| ICC | .25 | | | | .23 | | | | .11 | | | |
| N | 12 $_{Image}$ | | | | 12 $_{Image}$ | | | | 12 $_{Image}$ | | | |
| | 51 $_{ID}$ | | | | 51 $_{ID}$ | | | | 51 $_{ID}$ | | | |
| Observations | 406 | | | | 406 | | | | 406 | | | |
| Marginal $R^2$ / Conditional $R^2$ | .003 / .250 | | | | .023 / .245 | | | | .023 / NA | | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance; ICC = intraclass correlation coefficient; * = statistically significant after Bonferroni adjustments (salience scores $p < .017$; semantic scores $p < .025$)

*Table C4. Coefficients for all variables in the regression models assessing the maximum semantic informativeness of fixated scene content prior to self-reported mind wandering for the main study.*

| | Semantic map | | | Meaning map | | |
|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** |
| (Intercept) | .018 | -.266 – .283 | .952 | -.062 | -.454 – .329 | .755 |
| Probe Response [MW] | .038 | -.164 – .250 | .685 | .164 | -.006 – .335 | .059 |
| View Time | -.023 | -.114 – .064 | .584 | -.004 | -.076 – .069 | .922 |
| ***Random Effects*** | | | | | | |
| $\sigma^2$ | .75 | | | .48 | | |
| $\tau_{00}$ | .05 $_{ID}$ | | | .12 $_{ID}$ | | |
| | .22 $_{Image}$ | | | .43 $_{Image}$ | | |
| ICC | .27 | | | .53 | | |
| N | 12 $_{Image}$ | | | 12 $_{Image}$ | | |
| | 51 $_{ID}$ | | | 51 $_{ID}$ | | |
| Observations | 406 | | | 406 | | |
| Marginal $R^2$ / Conditional $R^2$ | .001 / .270 | | | .005 / .536 | | |

*Note. β* = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance; ICC = intraclass correlation coefficient

# Appendix D

*Table D. The minimum detectable effect sizes and associated power for the main study, replication, and joint-experiment analyses.*

| | Main Study | | | Replication | | | Joint-Experiment | | |
|---|---|---|---|---|---|---|---|---|---|
| | *MDES* | *(1- β)* | *95% CI* | *MDES* | *(1- β)* | *95% CI* | *MDES* | *(1- β)* | *95% CI* |
| **GBVS** | | | | | | | | | |
| Avg mean | .30 | .80 | .78 – .83 | .50 | .87 | .85 – .89 | .25 | .82 | .79 - .84 |
| Avg maximum | .30 | .80 | .77 – .82 | .45 | .85 | .83 – .87 | .25 | .81 | .79 - .84 |
| **AWS** | | | | | | | | | |
| Avg mean | .30 | .82 | .79 – .84 | .50 | .82 | .79 – .84 | .30 | .92 | .91 - .94 |
| Avg maximum | .30 | .84 | .82 – .86 | .45 | .84 | .81 – .86 | .25 | .81 | .79 - .84 |
| **RARE** | | | | | | | | | |
| Avg mean | .30 | .85 | .83 – .87 | .55 | .82 | .79 – .84 | .30 | .90 | .88 - .92 |
| Avg maximum | .35 | .89 | .87 – .91 | .50 | .80 | .78 – .83 | .30 | .92 | .90 - .93 |
| **Semantic map** | | | | | | | | | |
| Avg mean | .30 | .84 | .82 – .86 | .50 | .82 | .79 – .84 | .30 | .92 | .90 - .94 |
| Avg maximum | .30 | .81 | .79 – .84 | .50 | .87 | .85 – .89 | .25 | .82 | .80 - .84 |
| **Meaning map** | | | | | | | | | |
| Avg mean | .30 | .90 | .88 – .92 | .55 | .85 | .82 – .87 | .25 | .87 | .85 - .89 |
| Avg maximum | .25 | .80 | .78 – .83 | .50 | .87 | .85 – .89 | .25 | .90 | .88 - .92 |

*Note. MDES* = the estimated minimum detectable effect size that retained a power of at least .80; *(1- β)* = the average power associated with the MDES; 95% *CI* = the 95% confidence interval for the estimated average power

## Appendix E


**Figure E.** Example figures from the main study and the corresponding probed images in the replication study. Images in the replication were first cropped from the top and then expanded to achieve a standard viewing condition across the larger task battery.

# Appendix F

*Table F1. Coefficients for all variables in the regression models assessing the mean visual salience of fixated scene content prior to self-reported mind wandering for the replication*

| | GBVS | | | AWS | | | RARE | | |
|---|---|---|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** |
| (Intercept) | -.078 | -1.049 – .892 | .875 | -.046 | -.844 – .752 | .910 | -.247 | -.833 – .339 | .409 |
| Probe Response [MW] | .106 | -.216 – .428 | .518 | .233 | -.108 – .575 | .180 | .345 | -.022 – .711 | .066 |
| View Time | .258 | .022 – .494 | .032 | .176 | -.075 – .427 | .168 | -.094 | -.338 – .149 | .448 |
| Trial Position 2 | -.410 | -1.105 – .285 | .248 | -.622 | -1.407 – .163 | .121 | -.291 | -1.075 – .494 | .468 |
| Trial Position 3 | .467 | -.174 – 1.108 | .153 | .671 | -.054 – 1.396 | .070 | .573 | -.150 – 1.296 | .120 |
| Trial Position 4 | .182 | -.380 – .743 | .526 | .072 | -.562 – .705 | .824 | .231 | -.402 – .864 | .475 |
| Trial Position 5 | -.235 | -.826 – .356 | .435 | -.395 | -1.062 – .272 | .246 | -.075 | -.740 – .589 | .824 |
| Trial Position 6 | .207 | -.425 – .839 | .521 | -.064 | -.780 – .653 | .862 | -.083 | -.795 – .630 | .820 |
| Trial Position 7 | .369 | -.405 – 1.143 | .350 | .060 | -.814 – .934 | .892 | .352 | -.517 – 1.222 | .427 |
| ***Random Effects*** | | | | | | | | | |
| $\sigma^2$ | .57 | | | .62 | | | .77 | | |
| $\tau_{00}$ | $.12_{ID}$ | | | $.20_{ID}$ | | | $.14_{ID}$ | | |
| | $.60_{Image}$ | | | $.32_{Image}$ | | | $.10_{Image}$ | | |
| ICC | .56 | | | .46 | | | .24 | | |
| N | $3_{Image}$ | | | $3_{Image}$ | | | $3_{Image}$ | | |
| | $41_{ID}$ | | | $41_{ID}$ | | | $41_{ID}$ | | |
| Observations | 116 | | | 116 | | | 116 | | |
| Marginal $R^2$ / Conditional $R^2$ | .096 / .600 | | | .118 / .523 | | | .090 / .305 | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance; ICC = intraclass correlation coefficient

*Table F2. Coefficients for all variables in the regression models assessing the mean semantic informativeness of fixated scene content prior to self-reported mind wandering for the replication*

| | Semantic map | | | Meaning map | | |
|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** |
| (Intercept) | -.117 | -.950 – .717 | .784 | .071 | -.577 – .718 | .831 |
| Probe Response [MW] | -.208 | -.549 – .134 | .233 | .126 | -.238 – .491 | .497 |
| View Time | -.030 | -.270 – .211 | .808 | .256 | .006 – .506 | .045 |
| Trial Position 2 | .070 | -.610 – .751 | .839 | -.461 | -1.220 – .297 | .233 |
| Trial Position 3 | .263 | -.364 – .889 | .411 | .165 | -.534 – .864 | .644 |
| Trial Position 4 | .181 | -.369 – .731 | .519 | -.100 | -.713 – .513 | .749 |
| Trial Position 5 | .467 | -.110 – 1.045 | .113 | -.373 | -1.016 – .271 | .256 |
| Trial Position 6 | .406 | -.209 – 1.022 | .196 | -.035 | -.723 – .653 | .921 |
| Trial Position 7 | .304 | -.451 – 1.059 | .430 | -.002 | -.843 – .840 | .997 |
| ***Random Effects*** | | | | | | |
| $\sigma^2$ | .70 | | | .77 | | |
| $\tau_{00}$ | $.06_{ID}$ | | | $.11_{ID}$ | | |
| | $.41_{Image}$ | | | $.16_{Image}$ | | |
| ICC | .40 | | | .26 | | |
| N | $3_{Image}$ | | | $3_{Image}$ | | |
| | $41_{ID}$ | | | $41_{ID}$ | | |
| Observations | 116 | | | 116 | | |
| Marginal $R^2$ / Conditional $R^2$ | .032 / .423 | | | .084 / .326 | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance;

ICC = intraclass correlation coefficient

*Table F3. Coefficients for all variables in the regression models assessing the maximum visual salience of fixated scene content prior to self-reported mind wandering for the replication*

| | GBVS | | | AWS | | | RARE | | |
|---|---|---|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** |
| (Intercept) | -.056 | -1.106 – .994 | .917 | .037 | -.933 – 1.007 | .941 | -.164 | -.730 – .402 | .570 |
| Probe Response [MW] | .103 | -.202 – .408 | .507 | .207 | -.103 – .517 | .191 | .355 | -.016 – .726 | .060 |
| View Time | .269 | .045 – .492 | .018 | .192 | -.040 – .424 | .104 | -.048 | -.286 – .190 | .691 |
| Trial Position 2 | -.406 | -1.055 – .243 | .220 | -.674 | -1.383 – .035 | .062 | -.398 | -1.224 – .428 | .344 |
| Trial Position 3 | .439 | -.160 – 1.037 | .151 | .562 | -.092 – 1.216 | .092 | .471 | -.291 – 1.233 | .226 |
| Trial Position 4 | .160 | -.364 – .684 | .549 | -.051 | -.623 – .521 | .861 | .090 | -.576 – .756 | .791 |
| Trial Position 5 | -.286 | -.838 – .266 | .310 | -.506 | -1.108 – .097 | .100 | -.281 | -.980 – .418 | .431 |
| Trial Position 6 | .202 | -.387 – .792 | .502 | -.028 | -.674 – .618 | .933 | -.045 | -.797 – .707 | .907 |
| Trial Position 7 | .331 | -.391 – 1.053 | .369 | -.029 | -.819 – .760 | .942 | .276 | -.638 – 1.191 | .554 |
| ***Random Effects*** | | | | | | | | | |
| $\sigma^2$ | .52 | | | .51 | | | .77 | | |
| $\tau_{00}$ | $.10_{ID}$ | | | $.16_{ID}$ | | | $.19_{ID}$ | | |
| | $.74_{Image}$ | | | $.59_{Image}$ | | | $.06_{Image}$ | | |
| ICC | 0.62 | | | .60 | | | .25 | | |
| N | $3_{Image}$ | | | $3_{Image}$ | | | $3_{Image}$ | | |
| | $41_{ID}$ | | | $41_{ID}$ | | | $41_{ID}$ | | |
| Observations | 116 | | | 116 | | | 116 | | |
| Marginal $R^2$ / Conditional $R^2$ | .094 / .653 | | | .106 / .639 | | | .088 / .313 | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance; ICC = intraclass correlation coefficient

*Table F4. Coefficients for all variables in the regression models assessing the maximum semantic informativeness of fixated scene content prior to self-reported mind wandering for the replication*

| | Semantic map | | | Meaning map | | |
|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** |
| (Intercept) | -.141 | -.809 – .527 | .678 | .104 | -.702 – .911 | .800 |
| Probe Response [MW] | -.163 | -.490 – .164 | .329 | .056 | -.279 – .391 | .742 |
| View Time | -.245 | -.473 – -.018 | **.035** | .285 | .050 – .520 | **.017** |
| Trial Position 2 | -.036 | -.693 – .622 | .915 | -.417 | -1.077 – .243 | .216 |
| Trial Position 3 | .651 | .046 – 1.257 | .035 | .210 | -.398 – .817 | .499 |
| Trial Position 4 | .122 | -.409 – .654 | .651 | -.153 | -.687 – .381 | .574 |
| Trial Position 5 | .417 | -.141 – .975 | .143 | -.397 | -.957 – .164 | .165 |
| Trial Position 6 | .202 | -.393 – .797 | .506 | .006 | -.591 – .603 | .984 |
| Trial Position 7 | .401 | -.329 – 1.130 | .281 | .042 | -.690 – .775 | .910 |
| ***Random Effects*** | | | | | | |
| $\sigma^2$ | .64 | | | .68 | | |
| $\tau_{00}$ | $.07_{ID}$ | | | $.05_{ID}$ | | |
| | $.22_{Image}$ | | | $.38_{Image}$ | | |
| ICC | .31 | | | .39 | | |
| N | $3_{Image}$ | | | $41_{ID}$ | | |
| | $41_{ID}$ | | | $3_{Image}$ | | |
| Observations | 116 | | | 116 | | |
| Marginal $R^2$ / Conditional $R^2$ | .105 / .384 | | | .091 / .446 | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance;

ICC = intraclass correlation coefficient

# Appendix G

*Table G1. Coefficients for all variables in the regression models assessing the mean visual salience of fixated scene content prior to self-reported mind wandering for the joint-experiment analyses*

| | GBVS | | | | AWS | | | | RARE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | | ***β*** | ***95% CI*** | ***p*** | | ***β*** | ***95% CI*** | ***p*** | |
| (Intercept) | -.034 | -.329 – .260 | .820 | | -.087 | -.348 – .174 | .513 | | -.075 | -.291 – .140 | .493 | |
| Probe Response [MW] | .100 | -.074 – .273 | .261 | | .296 | .120 – .473 | .001 | * | .292 | .110 – .474 | .002 | * |
| View Time | .066 | -.015 – .147 | .112 | | .030 | -.053 – .113 | .481 | | -.039 | -.125 – .047 | .373 | |
| Experiment [Replication] | .018 | -.613 – .649 | .956 | | -.032 | -.597 – .533 | .911 | | -.059 | -.523 – .405 | .803 | |
| ***Random Effects*** | | | | | | | | | | | | |
| $\sigma^2$ | .71 | | | | .79 | | | | .88 | | | |
| $\tau_{00}$ | .11 $_{ID}$ | | | | .04 $_{ID}$ | | | | .01 $_{ID}$ | | | |
| | .22 $_{Image}$ | | | | .17 $_{Image}$ | | | | .11 $_{Image}$ | | | |
| ICC | .32 | | | | .21 | | | | .12 | | | |
| N | 15 $_{Image}$ | | | | 15 $_{Image}$ | | | | 15 $_{Image}$ | | | |
| | 92 $_{ID}$ | | | | 92 $_{ID}$ | | | | 92 $_{ID}$ | | | |
| Observations | .71 | | | | .79 | | | | .88 | | | |
| Marginal $R^2$ / Conditional $R^2$ | .11 $_{ID}$ | | | | .04 $_{ID}$ | | | | .01 $_{ID}$ | | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance; ICC = intraclass correlation coefficient;

* = statistically significant after Bonferroni adjustments (salience scores $p < .017$; semantic scores $p < .025$)

*Table G2. Coefficients for all variables in the regression models assessing the mean semantic informativeness of fixated scene content prior to self-reported mind wandering for the joint-experiment analysis*

| | Semantic map | | | Meaning map | | |
|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** |
| (Intercept) | .027 | -.263 – .316 | .857 | -.056 | -.404 – .292 | .753 |
| Probe Response [MW] | -.024 | -.199 – .151 | .787 | .167 | .007 – .327 | .040 |
| View Time | .004 | -.622 – .629 | .991 | -.010 | -.762 – .741 | .978 |
| Experiment [Replication] | -.028 | -.110 – .055 | .510 | .040 | -.035 – .114 | .295 |
| ***Random Effects*** | | | | | | |
| $\sigma^2$ | .74 | | | .57 | | |
| $\tau_{00}$ | .08 $_{ID}$ | | | .15 $_{ID}$ | | |
| | .22 $_{Image}$ | | | .32 $_{ImageE}$ | | |
| ICC | .28 | | | .45 | | |
| N | 15 $_{Image}$ | | | 15 $_{ImageE}$ | | |
| | 92 $_{ID}$ | | | 92 $_{ID}$ | | |
| Observations | 522 | | | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | .001 / .282 | | | .007 / .455 | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance;

ICC = intraclass correlation coefficient

*Table G3. Coefficients for all variables in the regression models assessing the maximum visual salience of fixated scene content prior to self-reported mind wandering for the joint-experiment analyses*

| | GBVS | | | | AWS | | | | RARE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | | ***β*** | ***95% CI*** | ***p*** | | ***β*** | ***95% CI*** | ***p*** | |
| (Intercept) | -.035 | -.332 – .262 | .818 | | -.082 | -.380 – .217 | .592 | | -.070 | -.359 – .219 | .634 | |
| Probe Response [MW] | .093 | -.082 – .267 | .297 | | .299 | .128 – .469 | **.001** | * | .299 | .126 – .472 | **.001** | * |
| View Time | .066 | -.016 – .147 | .114 | | .024 | -.056 – .105 | .552 | | -.028 | -.111 – .054 | .503 | |
| Experiment [Replication] | .023 | -.618 – .664 | .944 | | -.032 | -.683 – .620 | .924 | | -.068 | -.701 – .566 | .835 | |
| ***Random Effects*** | | | | | | | | | | | | |
| $\sigma^2$ | .73 | | | | .74 | | | | .80 | | | |
| $\tau_{00}$ | .08 $_{ID}$ | | | | .03 $_{ID}$ | | | | .01 $_{ID}$ | | | |
| | .23 $_{Image}$ | | | | .24 $_{Image}$ | | | | .23 $_{Image}$ | | | |
| ICC | .30 | | | | .27 | | | | .23 | | | |
| N | 15 $_{Image}$ | | | | 15 $_{Image}$ | | | | 15 $_{Image}$ | | | |
| | 92 $_{ID}$ | | | | 92 $_{ID}$ | | | | 92 $_{ID}$ | | | |
| Observations | 522 | | | | 522 | | | | 522 | | | |
| Marginal $R^2$ / Conditional $R^2$ | .006 / .304 | | | | .019 / .283 | | | | .018 / .243 | | | |

*Note.* $\beta$ = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance; ICC = intraclass correlation coefficient;

* = statistically significant after Bonferroni adjustments (salience scores $p < .017$; semantic scores $p < .025$)

*Table G4. Coefficients for all variables in the regression models assessing the maximum semantic informativeness of fixated scene content prior to self-reported mind wandering for the joint-experiment analyses*

| | Semantic map | | | Meaning map | | |
|---|---|---|---|---|---|---|
| ***Predictors*** | ***β*** | ***95% CI*** | ***p*** | ***β*** | ***95% CI*** | ***p*** |
| (Intercept) | .038 | -.268 – .344 | .808 | -.057 | -.444 – .330 | .774 |
| Probe Response [MW] | -.036 | -.208 – .135 | .678 | .143 | -.009 – .295 | .064 |
| View Time | -.048 | -.129 – .033 | .245 | .024 | -.047 – .094 | .511 |
| Experiment [Replication] | -.002 | -.670 – .666 | .995 | .005 | -.840 – .850 | .991 |
| ***Random Effects*** | | | | | | |
| $\sigma^2$ | .73 | | | .52 | | |
| $\tau_{00}$ | $.05_{ID}$ | | | $.12_{ID}$ | | |
| | $.25_{Image}$ | | | $.42_{Image}$ | | |
| ICC | .29 | | | .51 | | |
| N | $15_{Image}$ | | | $15_{Image}$ | | |
| | $92_{ID}$ | | | $92_{ID}$ | | |
| Observations | 522 | | | 522 | | |
| Marginal $R^2$ / Conditional $R^2$ | .003 / .293 | | | .005 / .510 | | |

*Note.* *β* = standardized coefficients; *CI* = confidence interval; $\sigma^2$ = within-group variance; $\tau_{00}$ = between-group variance;

ICC = intraclass correlation coefficient