applied sciences

MDPI

Article

Eye Tracking of Visual Attention and Memory for Image and Text Regions in Complex Advertising Under Task-Directed Viewing

Jintao Wei [1], Chen Chen [1] and Chenyang Cai [1,2,*]

[1] College of Furnishings and Industrial Design, Nanjing Forestry University, Nanjing 210037, China; 8241421651@njfu.edu.cn (J.W.); cchen1990@njfu.edu.cn (C.C.)
[2] Jiangsu Co-Innovation Center of Efficient Processing and Utilization of Forest Resources, Nanjing Forestry University, Nanjing 210037, China
* Correspondence: chenyangc@njfu.edu.cn

Featured Application

Combining AOI-based eye tracking with memory testing can help designers assess whether patterns of gaze allocation are accompanied by corresponding differences in later memory.

Abstract

Eye tracking can reveal how visual attention is allocated within complex displays, but gaze allocation does not necessarily predict what is later remembered. This exploratory study examined image and text regions in three commercial poster identities, each presented in colour and black-and-white versions. Forty adults viewed the three posters in one presentation condition ($n = 20$ per condition). Each stimulus was shown for approximately 10 s, and eye movements were recorded with a Tobii Pro X-Series eye tracker at 30 Hz after five-point calibration. Predefined areas of interest (AOIs) separated image and text content. Participant-level Fixation Count (FC) and Total Fixation Duration (TFD), averaged across the three posters, were greater for text than image AOIs (FC: $t(39) = 19.46$, $p < 0.001$, $dz = 3.08$; TFD: $t(39) = 20.08$, $p < 0.001$, $dz = 3.17$). After viewing, participants completed the same 24-item questionnaire. Archived question-level summaries showed descriptively higher accuracy for image-related than text-related items in both conditions (black-and-white: 70.83% vs. 65.83%; colour: 76.67% vs. 70.42%). These findings show a discrepancy between gaze allocation patterns and descriptive memory performance under the present task conditions. They should be interpreted as exploratory associations between gaze behaviour and memory outcomes rather than evidence of a causal attention–memory mechanism.

Keywords: eye tracking; areas of interest; visual attention; advertising; memory; image–text integration; human factors

Check for updates

Academic Editor: Pawel Kasprowski

Received: 20 September 2026
Revised: 13 October 2026
Accepted: 20 October 2026
Published: 28 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. Visual Attention and Memory in Image–Text Displays

Advertisements combine images, typography, layout, and brand information within a single complex display. Eye tracking can reveal how overt attention is distributed across these elements, whereas a subsequent memory measure can test whether the attended information remains retrievable. This distinction matters because looking longer at a region does not necessarily imply that its content will be remembered more accurately.

Pictorial information often shows a memory advantage over verbal information. Recent work links image memorability to rapid perceptual encoding, intrinsic visual properties,

distinctive latent representations, and sustained activity across ventral visual regions [1–4]. These mechanisms suggest that an image can provide multiple perceptual and semantic retrieval cues even when it receives relatively little overt sampling.

Text requires orthographic recognition, lexical access, and semantic integration. These operations transform visual marks into increasingly abstract linguistic representations [5,6]. Reading can therefore demand repeated sampling and longer processing time, particularly when viewers expect to remember explicit message content.

The picture-superiority effect should not be treated as universal. Perceptual distinctiveness can reduce or eliminate the usual image advantage, and memory depends on the interaction among stimulus properties, encoding demands, and retrieval cues [7,8]. Authentic advertising is therefore a useful applied test because image and text content compete within a shared, visually structured display.

1.2. Task-Directed Attention and Eye-Tracking Measures

Task-directed attention allocates limited processing resources to information that appears relevant to the viewer's goal. Evidence from retrospective- and prospective-cue paradigms shows that prioritised representations can gain precision, whereas benefits vary with cue number, feature dimension, distraction, and working-memory load [9–15]. In an advertising memory task, viewers may consequently prioritise text because it appears to contain testable facts, even when images remain memorable.

Eye-tracking measures require an explicit theoretical interpretation. Fixation Count can index repeated visual sampling, and Total Fixation Duration within an area of interest (AOI) can index the accumulated time allocated to that region, but neither measure is a direct readout of comprehension or memory [16]. Prior work has linked oculomotor behaviour and pupil responses to task demand, cognitive load, encoding, and later retrieval [17–25]. These findings support the use of gaze measures as indicators of overt attention while also showing why an independent behavioural outcome is needed.

1.3. Research Gap and Study Objectives

Images and words follow partly different encoding pathways. Images can support rapid parallel extraction of perceptual and semantic information, whereas text generally requires more serial decoding and symbol-to-meaning mapping [26–30]. Format effects are not limited to pictures and words; even visually compact symbols can be remembered differently from their verbal labels [31]. These observations motivate an applied comparison of attentional input and memory output in image and text AOIs.

Many picture-text studies use isolated words, line drawings, or tightly controlled image sets. Such materials provide experimental control but do not reproduce the visual density, layout hierarchy, and semantic integration of commercial advertising. Applied eye-tracking research has examined attention to textual and pictorial regions in advertising images, banners, social-media branding, and visual-communication design [32–36]. Recent methodological reviews also emphasise that the selected gaze metric and unit of analysis must follow the research question rather than be treated as a general measure of advertising effectiveness [16,32,33]. The present study therefore explored the relationship between gaze allocation to image and text AOIs and subsequent descriptive memory performance during task-directed viewing of authentic advertising posters.

The study addressed two questions. First, under explicit memorisation instructions, how were Fixation Count and Total Fixation Duration allocated between image and text AOIs? Second, what descriptive accuracy pattern was observed for image- and text-related questionnaire items? Presentation condition was retained as a design factor and reporting

stratum, but colour was not treated as the principal theoretical contribution. The study did not test a direct participant-by-item coupling between gaze and memory.

2. Materials and Methods

2.1. Study Design and Stimuli

This exploratory, ecologically oriented study used three authentic commercial poster identities. Each underlying advertisement was available as a colour and a black-and-white version, yielding six stimulus versions: BP1 and CP1 represented Poster 1, BP2 and CP2 represented Poster 2, and BP3 and CP3 represented Poster 3. Presentation condition (colour or black-and-white) was a between-participant factor, whereas poster identity and AOI type (image or text) varied within participants. Each participant viewed the three poster identities in only one presentation condition. In this manuscript, poster identity denotes the underlying advertisement design, and stimulus version denotes its colour or black-and-white rendering.

The corresponding colour and black-and-white versions represented the same advertisement design, so poster content and layout were matched across presentation conditions at the identity level. The retained Track Map exports from ErgoLAB 3.0 reported a display/export resolution of 1920 × 1080 pixels. This value is not treated as the AOI coordinate system. Table 1 reports the stimulus version and AOI coding without reproducing the commercial advertisements. A schematic representation of the six stimulus versions and their AOI structure is shown in Figure 1.

Table 1. Poster identities, stimulus versions, and labels for areas of interest (AOIs).

Poster Identity	Black-and-White Version	Colour Version	Image AOI	Text AOI
Poster 1	BP1	CP1	P1-I	P1-T
Poster 2	BP2	CP2	P2-I	P2-T
Poster 3	BP3	CP3	P3-I	P3-T

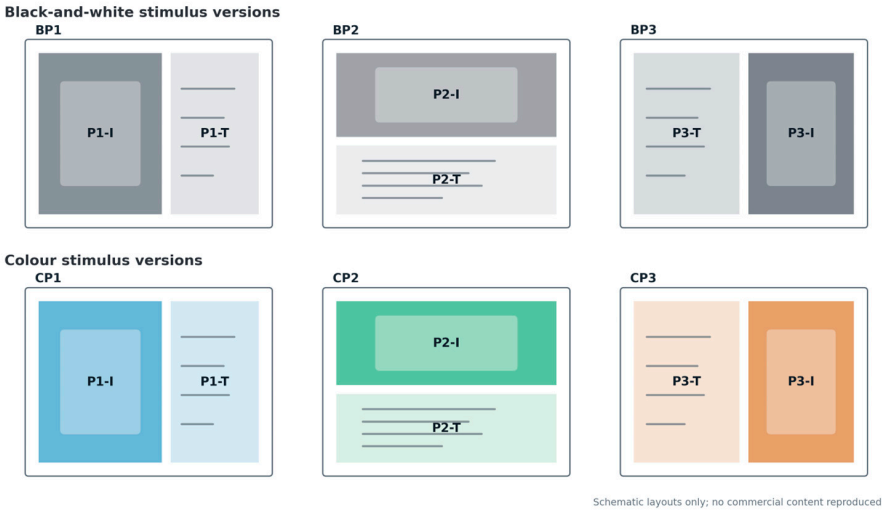


Figure 1. Schematic overview of the six stimulus versions and their AOI coding. The experiment used three poster identities, each presented in a black-and-white version and a colour version, yielding six stimulus versions (BP1–BP3 and CP1–CP3). For each poster identity, predefined image and text areas of interest (AOIs) were coded as P1-I/P1-T, P2-I/P2-T, and P3-I/P3-T. To avoid reproducing copyrighted commercial advertisements, this figure presents a schematic representation of the stimulus layouts rather than the original poster images.

2.2. Participants

Forty adults participated (20 men and 20 women; mean age = 23.45 years, SD = 2.35, range = 18–30). All participants reported normal or corrected-to-normal vision, provided written informed consent, and were screened for self-reported visual or neurological conditions that could interfere with normal viewing. The study was exploratory, and no a priori sample-size calculation was reported.

2.3. Eye-Tracking Apparatus

The archived experimental records confirmed the use of a Tobii Pro X-Series remote eye tracker (Tobii AB, Danderyd, Sweden) with a sampling rate of 30 Hz and five-point calibration. However, the specific X-Series submodel, validity thresholds, tracking-ratio criteria, and fixation-detection parameters were not available in the retained records. The experiment was conducted in the Human Factors and Ergonomics Laboratory, and the eye tracker was used as a remote device (Figure 2). Stimuli were displayed using a Dell E2423H monitor and a Dell OptiPlex 7000 compact computer (Dell Inc., Round Rock, TX, USA), with a display resolution of 1920 × 1080 pixels. ErgoLAB 3.0, including its Track Map export module, was used to process and export the eye-tracking records. Retained Track Map exports reported a resolution of 1920 × 1080 pixels.

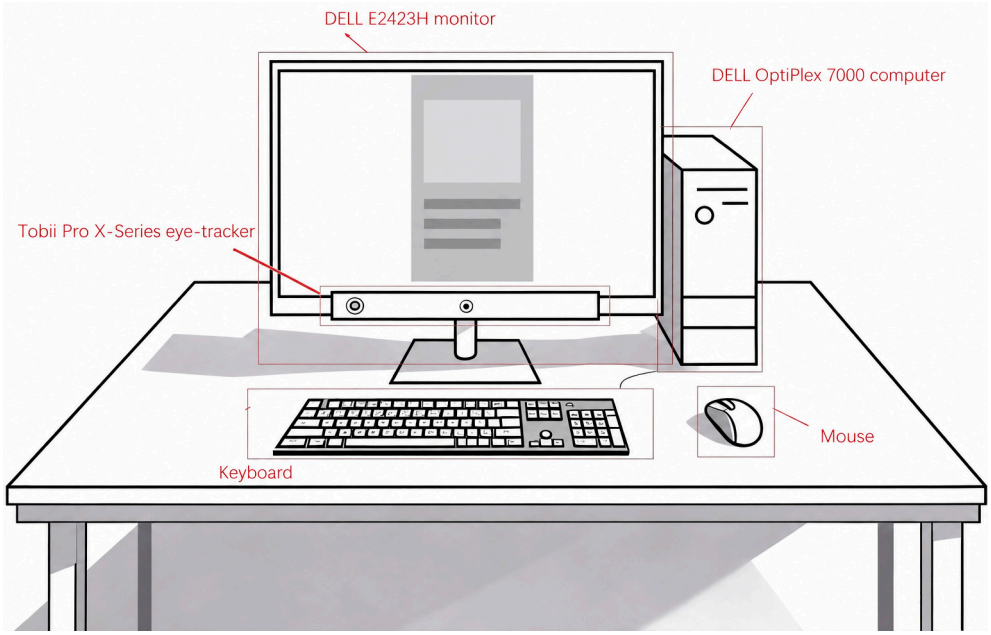


Figure 2. AI-assisted schematic illustration of the experimental setup. Commercial stimulus content was replaced by a neutral display; the figure does not reproduce an experimental stimulus.

2.4. Calibration and Experimental Procedure

Stimuli were presented at the centre of the monitor against a white background at a viewing distance of approximately 60 cm. The experimenter read standardised instructions before the formal trials. A five-point calibration procedure was performed before the experimental task. The surviving records do not preserve the numerical calibration acceptance criterion or error threshold, so no such value is reported. Each advertising stimulus was presented for approximately 10 s; retained ErgoLAB Track Map exports showed segment durations close to 10 s (for example, approximately 9.996 s).

Participants were assigned to one of two presentation conditions: 20 viewed CP1, CP2, and CP3, whereas 20 viewed BP1, BP2, and BP3. The available records do not establish that condition assignment was random, so no random-allocation claim is made. Participants

were instructed to memorise the advertisement content while viewing. After viewing all three posters, both groups completed the same 24-item questionnaire. Question wording, response options, item order, and questionnaire content were identical across presentation conditions. The questionnaire comprised 12 image-related and 12 text-related items; each of the three poster identities was associated with four image-related and four text-related questions. The exact response format, time limit, complete answer key, and detailed scoring documentation were not retained in the archived materials. Stimulus labels and operational definitions are provided in Tables 2 and 3. The overall experimental procedure is summarised in Figure 3.

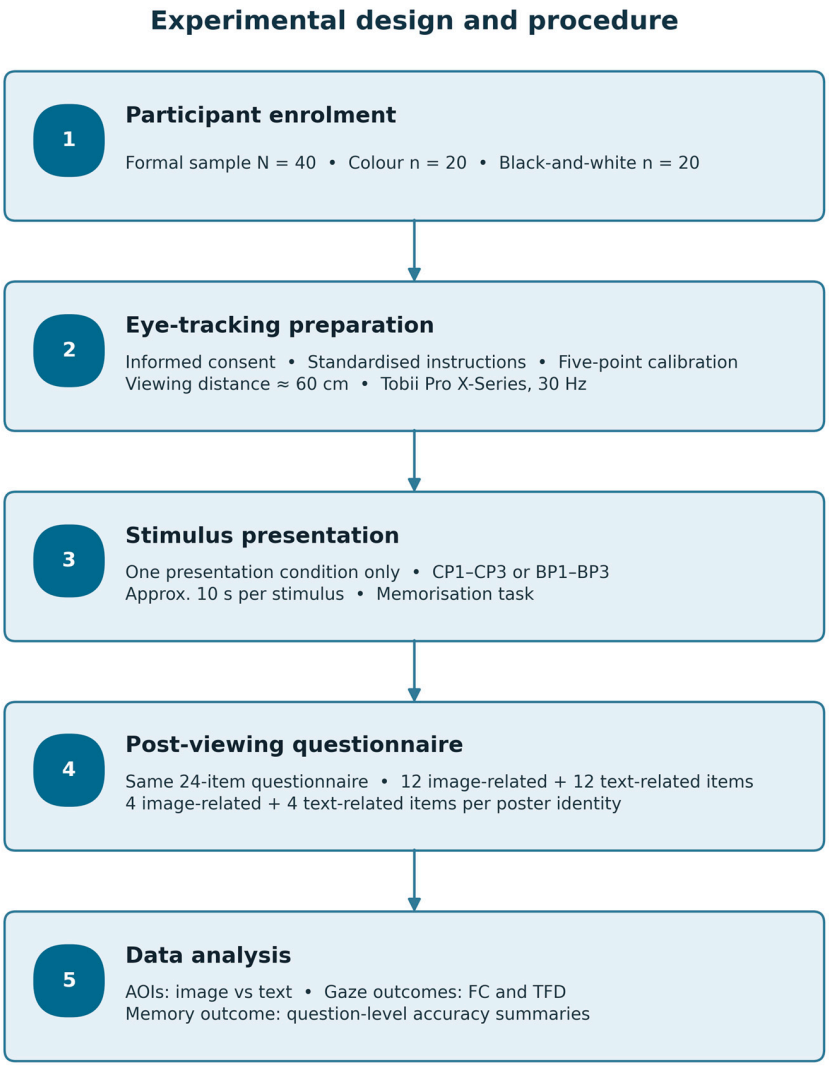


Figure 3. Experimental design and procedure flow. Forty participants were assigned to one of two presentation conditions (colour or black-and-white). After informed consent, task instructions, and five-point calibration, each participant viewed three poster identities within one presentation condition only (CP1–CP3 or BP1–BP3), with each stimulus presented for approximately 10 s at a viewing distance of about 60 cm. After viewing all three posters, participants completed the same 24-item questionnaire, comprising 12 image-related and 12 text-related items, with four items of each type corresponding to each poster identity. Eye-tracking data were extracted for predefined image and text AOIs.

Table 2. Stimulus and AOI labels.

Stimulus or AOI	Code	Members	Definition
Poster identity	P1–P3	Poster 1, Poster 2, Poster 3	Underlying advertisement design.
Stimulus version	BP/CP	BP1–BP3; CP1–CP3	Black-and-white or colour rendering of a poster identity.
Colour condition	C	CP1, CP2, CP3	Three colour stimulus versions viewed by 20 participants.
Black-and-white condition	B	BP1, BP2, BP3	Three black-and-white stimulus versions viewed by 20 participants.
Image AOI	I	P1-I, P2-I, P3-I	Pictorial content; fixed within a stimulus version across participants.
Text AOI	T	P1-T, P2-T, P3-T	Written content; fixed within a stimulus version across participants.

Table 3. Operational definitions of the study variables.

Variable	Abbreviation	Definition
Fixation Count	FC	Number of exported fixation events within an AOI for one participant and poster.
Total Fixation Duration	TFD	Summed duration of all exported fixation events within an AOI for one participant and poster (s).
Aggregate Fixation Count	-	Sum of participant-poster FC values for descriptive reporting.
Aggregate Total Fixation Duration	-	Sum of participant-poster TFD values for descriptive reporting (s).
Accuracy	ACC	Question-level accuracy: percentage of the 20 formal participants within a presentation condition who answered a given questionnaire item correctly.

2.5. AOI Definition and Data Processing

AOIs were defined according to semantic regions representing image and text content within each advertisement. AOIs were predefined for each stimulus version and applied identically across participants viewing the same stimulus version. The available records do not establish that the colour and black-and-white versions used numerically identical AOI geometry, although they represented the same poster identity. The 1920 × 1080 Track Map export resolution was not used as an AOI coordinate system. Image and text AOIs were not matched in physical area. Because the archived AOI coordinate files were unavailable, retrospective normalisation of fixation metrics by AOI surface area could not be performed. Differences in raw FC and TFD may therefore partly reflect unequal AOI geometry.

ErgoLAB 3.0 exported fixation-event records and AOI summaries that included Fixation Count, Average Fixation Duration, and Total Fixation Duration. The present analysis used Fixation Count (FC) and Total Fixation Duration (TFD). TFD was the summed duration of all exported fixation events within an AOI for one participant and poster. It was distinct from Average Fixation Duration. The retained exports did not preserve the specific fixation-detection algorithm or its velocity, dispersion, or minimum-duration settings. Accordingly, no I-VT, I-DT, or numerical event-detection threshold is attributed to this experiment.

The exported fixation-event durations showed an approximately 33.33 ms temporal granularity, which is consistent with the confirmed 30 Hz acquisition setting. This agreement was used only as a check on the temporal resolution of the exports; it was not treated as evidence of gaze-data validity. The archived materials do not document a prespecified

tracking-ratio threshold, a gaze-validity percentage, a calibration-error cut-off, or a validity-based participant-exclusion rule. No unverified value was reconstructed retrospectively, and TFD or fixation-time coverage was not interpreted as gaze-data validity.

2.6. Outcome Measures

Fixation Count (FC) was defined as the number of exported fixation events within an AOI for one participant and one poster identity. Total Fixation Duration (TFD) was defined as the summed duration of all exported fixation events within that AOI for the same participant and poster. One participant-poster observation therefore represented one participant's data for one poster identity and one AOI type within the assigned presentation condition. For the questionnaire, question-level accuracy was the percentage of the 20 formal participants within a presentation condition who answered a given item correctly. Mean item accuracy was calculated descriptively across the 12 image-related or 12 text-related items in each condition. FC and TFD represent raw overt visual sampling; they do not, by themselves, establish comprehension, memory, cognitive load, data validity, or an intrinsic attentional advantage independent of AOI area.

2.7. Statistical Analysis

Because the three 20-row poster blocks preserved the same participant order within each presentation condition, repeated poster observations were linked using anonymised reconstructed participant indices (B01–B20 and C01–C20). These indices were created for analysis and do not represent original participant identifiers. For each participant, FC and TFD were averaged across the three poster identities separately for image and text AOIs. FC and TFD were treated as complementary primary gaze outcomes and were analysed using two-sided paired-samples *t*-tests across all 40 participants, with Cohen's dz reported as the effect size. Family-wise type I error across these two primary gaze tests was controlled using a Bonferroni-adjusted alpha of 0.025. Normality of the participant-level Text-minus-Image difference scores was assessed using Shapiro–Wilk tests, and two-sided Wilcoxon signed-rank tests were conducted as sensitivity analyses without replacing the primary paired *t*-tests. Secondary presentation-condition differences were examined using Welch independent-samples *t*-tests on participant-level Text-minus-Image difference scores. Hedges' g was calculated using the pooled standard deviation with the small-sample correction; 95% confidence intervals for the raw Black-and-white-minus-Colour mean differences used the Welch standard error and degrees of freedom. Tables 4 and 5 and Figures 4 and 5 present the participant-level and stimulus-version summaries, respectively, for Fixation Count; Tables 6 and 7 and Figures 6 and 7 present the corresponding summaries for Total Fixation Duration. Statistical analyses were performed using IBM SPSS Statistics for Windows, version 27.0.1 (IBM Corp., Armonk, NY, USA).

Table 4. Participant-level Fixation Count summary by presentation condition and AOI type.

Presentation Condition	AOI	*n* Participants	Mean FC Across Posters ± SD
Black-and-white	Image	20	7.95 ± 1.87
Black-and-white	Text	20	21.92 ± 4.77
Colour	Image	20	7.60 ± 2.15
Colour	Text	20	21.55 ± 2.72

Note: $n = 20$ participants per presentation condition. Means and SDs were calculated from participant-level values averaged across the three poster identities.

Table 5. Descriptive Fixation Count results for each stimulus version and AOI type.

Stimulus Version	AOI	*n* Participants	Total FC	Mean ± SD
BP1	Image	20	151	7.55 ± 2.76
BP1	Text	20	450	22.50 ± 5.36
CP1	Image	20	144	7.20 ± 3.14
CP1	Text	20	371	18.55 ± 4.55
BP2	Image	20	154	7.70 ± 2.39
BP2	Text	20	417	20.85 ± 6.03
CP2	Image	20	147	7.35 ± 2.25
CP2	Text	20	490	24.50 ± 4.57
BP3	Image	20	172	8.60 ± 2.89
BP3	Text	20	448	22.40 ± 4.78
CP3	Image	20	165	8.25 ± 2.88
CP3	Text	20	432	21.60 ± 4.60

Note: Each stimulus version was viewed by 20 participants. The same participants contributed observations for three poster identities within their assigned condition.

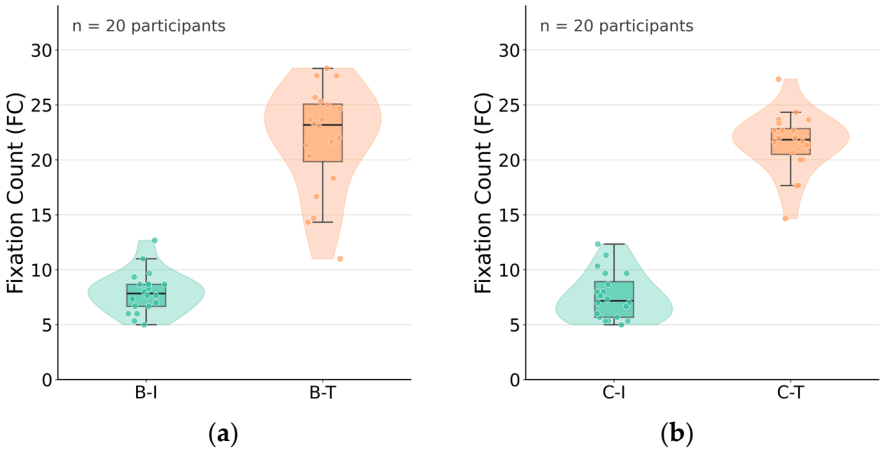


Figure 4. Participant-level Fixation Count distributions by presentation condition. Each plotted value is one participant's AOI-specific FC averaged across the three poster identities ($n = 20$ participants per condition): (**a**) black-and-white presentation condition; (**b**) colour presentation condition. Shading: density; boxplots: median, interquartile range, and whiskers; circles: individual participant observations.

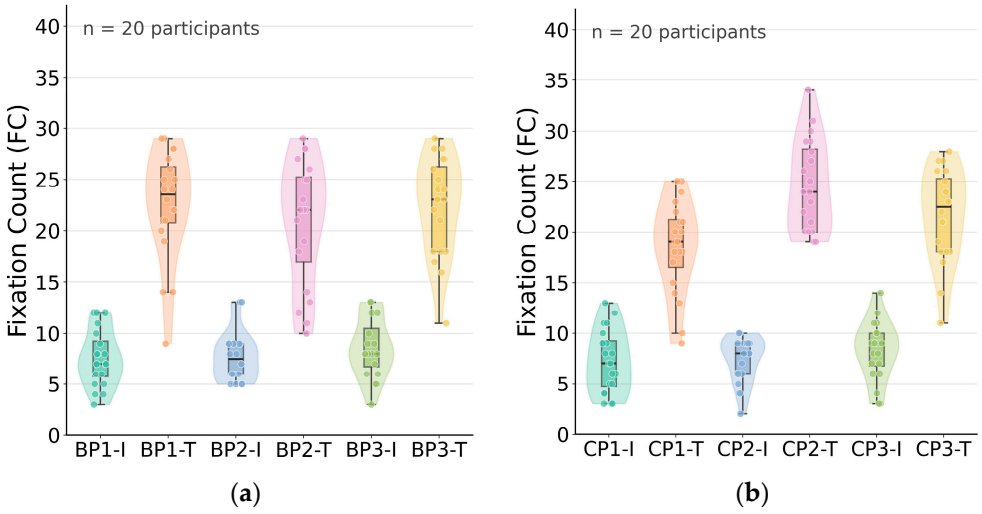


Figure 5. Stimulus-version Fixation Count distributions for the three poster identities: (**a**) black-and-white versions; (**b**) colour versions. Shading: density; boxplots: median, interquartile range, and whiskers; circles: individual participant observations.

Table 6. Participant-level Total Fixation Duration summary by presentation condition and AOI type.

Presentation Condition	AOI	*n* Participants	Mean TFD Across Posters ± SD (s)
Black-and-white	Image	20	1.90 ± 0.57
Black-and-white	Text	20	5.23 ± 0.50
Colour	Image	20	1.81 ± 0.64
Colour	Text	20	5.60 ± 0.75

Note: $n = 20$ participants per presentation condition. Means and SDs were calculated from participant-level values averaged across the three poster identities.

Table 7. Descriptive Total Fixation Duration results for each stimulus version and AOI type.

Stimulus Version	AOI	*n* Participants	Aggregate TFD (s)	Mean TFD ± SD (s)
BP1	Image	20	35.53	1.78 ± 0.85
BP1	Text	20	110.97	5.55 ± 0.81
CP1	Image	20	35.49	1.77 ± 1.14
CP1	Text	20	121.32	6.07 ± 1.26
BP2	Image	20	44.63	2.23 ± 0.98
BP2	Text	20	95.66	4.78 ± 0.83
CP2	Image	20	29.94	1.50 ± 0.60
CP2	Text	20	110.83	5.54 ± 0.68
BP3	Image	20	33.55	1.68 ± 0.47
BP3	Text	20	107.45	5.37 ± 0.67
CP3	Image	20	43.17	2.16 ± 1.04
CP3	Text	20	103.68	5.18 ± 0.90

Note: Each stimulus version was viewed by 20 participants. Values are descriptive and were not adjusted for AOI area.

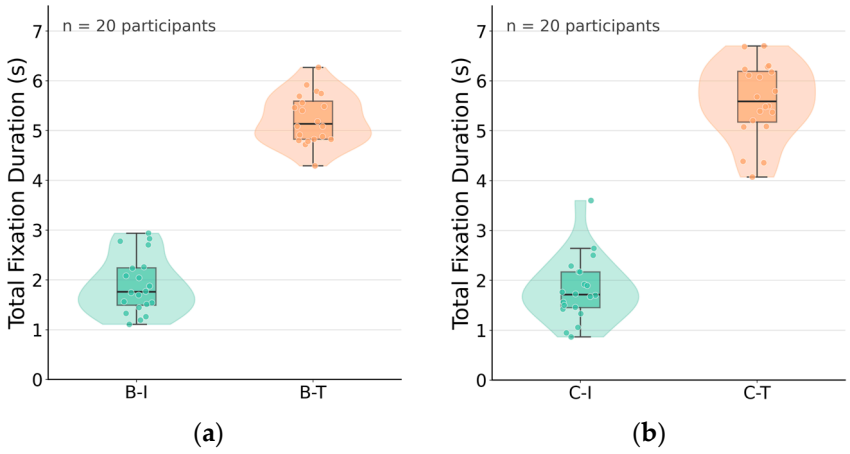


Figure 6. Participant-level Total Fixation Duration distributions by presentation condition. Each plotted value is one participant's AOI-specific TFD averaged across the three poster identities ($n = 20$ participants per condition): (**a**) black-and-white presentation condition; (**b**) colour presentation condition. Shading: density; boxplots: median, interquartile range, and whiskers; circles: individual participant observations.

For memory outcomes, only archived question-level accuracy percentages were available. Mean item accuracy was summarised separately for the 12 image-related and 12 text-related questions within each presentation condition. Because the original participant-by-item response matrix was unavailable, participant-level variance, paired image–text differences, and mixed-effects analyses could not be reconstructed. Memory results were therefore treated as descriptive.

For gaze outcomes, presentation-condition comparisons of participant-level Text-minus-Image difference scores were treated as secondary analyses; non-significant results were not interpreted as evidence of equivalence.

The archived questionnaire materials did not retain the original participant-by-item response matrix, individual response records, exact response-format documentation, time-limit information, complete answer key, or detailed scoring documentation.

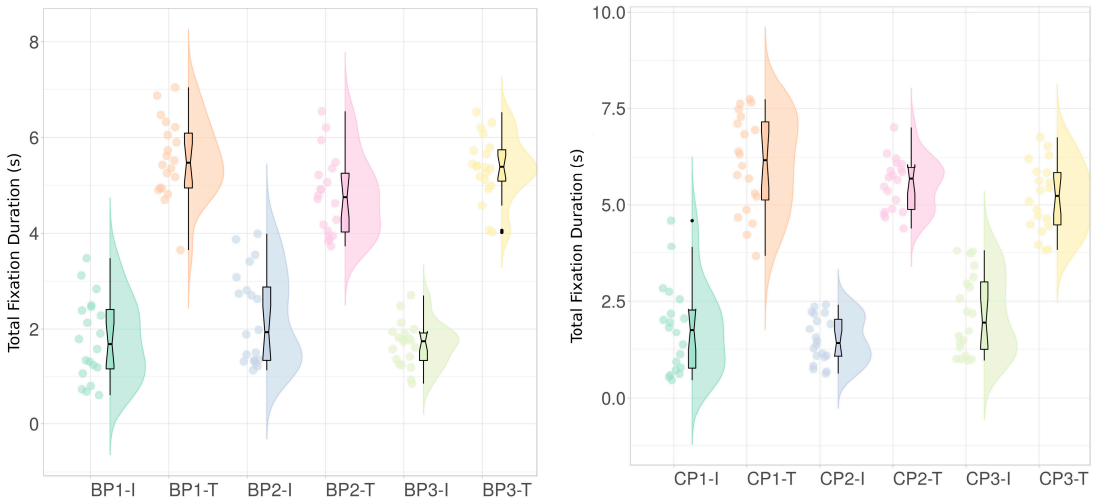


Figure 7. Stimulus-version Total Fixation Duration distributions for the three poster identities: (**a**) black-and-white versions; (**b**) colour versions. Shading: density; boxplots: median, interquartile range, and whiskers; circles: individual participant observations.

2.8. Use of Generative Artificial Intelligence During Manuscript Revision

During manuscript revision, ChatGPT-6 (OpenAI) was used to assist with language refinement, structural reorganisation, methodological consistency checking, and the development and verification of statistical analysis procedures based on author-provided data. The tool was not used to generate, alter, or fabricate experimental data. The inferential gaze analyses reported in the manuscript were independently checked and verified by the authors using IBM SPSS Statistics for Windows, version 27.0.1. The descriptive questionnaire-accuracy summaries were cross-checked against the archived Wenjuanxing question-level accuracy summaries. ChatGPT was also used to assist in preparing the non-data-bearing schematic Figures 1–3. These figures were reviewed by the authors for consistency with the documented experimental design and do not represent generated experimental data.

3. Results

3.1. Analysis of Fixation Count

Across all 40 participants, text AOIs received substantially more fixations than image AOIs, paired $t(39) = 19.46$, $p < 0.001$, Cohen's $dz = 3.08$; this result met the Bonferroni-adjusted alpha of 0.025. The participant-level Text-minus-Image FC difference scores showed no statistically detectable departure from normality, Shapiro–Wilk $W = 0.971$, $p = 0.388$. A Wilcoxon signed-rank sensitivity analysis supported the same substantive conclusion, $W = 0$, $p < 0.001$. The analyses used participant-level FC values obtained by averaging across the three poster identities within each participant. These results show higher fixation counts for text regions under the present viewing conditions.

In the black-and-white condition ($n = 20$), participant-level mean FC was 7.95 ± 1.87 for image AOIs and 21.92 ± 4.77 for text AOIs. In the colour condition ($n = 20$), the corresponding values were 7.60 ± 2.15 and 21.55 ± 2.72 (Table 4). The mean Text-minus-Image difference was 13.97 in the black-and-white condition and 13.95 in the colour condition. The secondary between-condition comparison was Welch's $t(34.74) = 0.01$, $p = 0.991$, Hedges' $g = 0.004$; the raw Black-and-white-minus-Colour mean difference was 0.02 fixations, 95% CI [−2.93, 2.97]. There was no statistically detectable between-condition difference in the magnitude of the image–text FC contrast. Figure 4 shows the participant-level distributions.

Within the black-and-white condition, image and text totals were 151 and 450 for BP1, 154 and 417 for BP2, and 172 and 448 for BP3, respectively.

Within the colour condition, image and text totals were 144 and 371 for CP1, 147 and 490 for CP2, and 165 and 432 for CP3, respectively.

Poster-level descriptive results showed the same direction for all six stimulus versions: text AOIs received more raw fixations than image AOIs (Table 5). This pattern was observed for each poster identity in both presentation conditions, but it was not area normalised.

Figure 5 visualises the poster-level Fixation Count distributions.

3.2. Analysis of Total Fixation Duration

Across all 40 participants, text AOIs accumulated substantially greater Total Fixation Duration than image AOIs, paired $t(39) = 20.08$, $p < 0.001$, Cohen's $dz = 3.17$; this result met the Bonferroni-adjusted alpha of 0.025. The participant-level Text-minus-Image TFD difference scores showed no statistically detectable departure from normality, Shapiro–Wilk $W = 0.976$, $p = 0.550$. A Wilcoxon signed-rank sensitivity analysis supported the same substantive conclusion, $W = 0$, $p < 0.001$. The analyses used participant-level TFD values obtained by averaging across the three poster identities within each participant. These results show longer accumulated fixation duration for text regions under the present viewing conditions.

In the black-and-white condition ($n = 20$), participant-level mean TFD was 1.90 ± 0.57 s for image AOIs and 5.23 ± 0.50 s for text AOIs. In the colour condition ($n = 20$), the corresponding values were 1.81 ± 0.64 s and 5.60 ± 0.75 s (Table 6). The mean Text-minus-Image difference was 3.34 s in the black-and-white condition and 3.79 s in the colour condition. The secondary between-condition comparison was Welch's $t(31.38) = -1.27$, $p = 0.213$, Hedges' $g = -0.39$; the raw Black-and-white-minus-Colour mean difference was −0.45 s, 95% CI [−1.17, 0.27]. There was no statistically detectable between-condition difference in the magnitude of the image–text TFD contrast. Figure 6 shows the participant-level distributions.

Within the black-and-white presentation condition, aggregate image and text TFD values were 35.53 s and 110.97 s for BP1, 44.63 s and 95.66 s for BP2, and 33.55 s and 107.45 s for BP3, respectively.

Within the colour presentation condition, aggregate image and text TFD values were 35.49 s and 121.32 s for CP1, 29.94 s and 110.83 s for CP2, and 43.17 s and 103.68 s for CP3, respectively.

Poster-level descriptive results showed the same direction for all six stimulus versions: text AOIs accumulated greater raw TFD than image AOIs (Table 7). This pattern was not adjusted for AOI area.

Figure 7 visualises the poster-level Total Fixation Duration distributions.

3.3. Descriptive Question-Level Memory Accuracy

Each presentation condition contained 20 formal participants. After viewing all three posters, participants in both conditions completed the same 24-item questionnaire, with

identical question wording, response options, item order, and content. The questionnaire comprised 12 image-related and 12 text-related items, with four questions of each type associated with each poster identity. The recovered questionnaire records contained question-level accuracy percentages rather than the original participant-by-item response matrix.

In the black-and-white condition, mean question-level accuracy was 70.83% for image-related items (170 correct responses out of 240 response opportunities) and 65.83% for text-related items (158/240), a descriptive difference of 5.00 percentage points. In the colour condition, mean question-level accuracy was 76.67% for image-related items (184/240) and 70.42% for text-related items (169/240), a descriptive difference of 6.25 percentage points. Across the two presentation conditions, descriptively, image-related accuracy was 73.75% (354/480) and text-related accuracy was 68.13% (327/480), a difference of 5.63 percentage points. No inferential test was conducted because the archived summaries did not preserve the participant-by-item dependency structure. Relative to the gaze allocation measures, these question-level memory summaries showed a different descriptive tendency; they do not establish participant-level fixation–memory relationships.

These descriptive question-level results are summarised in Table 8.

Table 8. Descriptive question-level memory accuracy by presentation condition and question type.

Presentation Condition	Question Type	Number of Items	Mean Item Accuracy (%)	Correct Responses/ Total Responses
Black-and-white	Image	12	70.83	170/240
Black-and-white	Text	12	65.83	158/240
Colour	Image	12	76.67	184/240
Colour	Text	12	70.42	169/240

Note: Each item was answered by 20 participants within each presentation condition. Because the participant-by-item response matrix was not retained, these values are descriptive question-level summaries and were not subjected to participant-level inferential testing.

4. Discussion

Within the current task and three-poster stimulus set, the findings show a discrepancy between gaze allocation patterns and descriptive memory performance. Participant-level gaze analyses showed substantially greater FC and TFD for text AOIs than for image AOIs under explicit memorisation instructions, whereas archived aggregate question-level memory summaries showed descriptively higher accuracy for image-related than text-related questionnaire items in both presentation conditions. Because the gaze evidence is participant-level and inferential while the memory evidence is aggregate and descriptive, this pattern is not interpreted as evidence of a causal attention–memory mechanism or as a direct test of encoding processes.

Participant-level gaze analyses consistently showed greater raw overt sampling of text AOIs. They do not establish that text required deeper processing, imposed greater cognitive load, or held an intrinsic attentional priority. Reading may involve serial sampling, but the present raw FC and TFD values were also sensitive to AOI geometry and information density [16,30,32]. The most defensible interpretation is therefore that text AOIs received greater raw gaze allocation under the present task conditions.

The descriptive memory pattern is consistent with work showing that perceptual and semantic properties can make pictorial information memorable [1–4,26,29], but it does not demonstrate a universal picture-superiority effect. The image–text differences were modest, questionnaire item difficulty was not experimentally equated, and the original participant-by-item response matrix was unavailable. The observed descriptive pattern is

therefore limited to the current participants, task, and stimulus set and does not support participant-level inferential claims.

Although AOI-based fixation metrics provide useful information regarding gaze allocation, differences in AOI spatial extent may influence raw fixation measures. Because AOI-area normalisation could not be performed in the present study, comparisons between image and text regions should be interpreted cautiously. Future studies should incorporate AOI-area normalisation to improve comparability between image and text regions.

The 30 Hz sampling rate used in this study supported the reported fixation-based analyses; however, it may provide limited sensitivity for fine-grained temporal events, such as rapid fixation transitions or reading-related micro-events. Future studies using higher-frequency eye trackers may provide additional temporal information.

Because participant-level item-response linkage between individual fixation behaviour and specific memory outcomes was unavailable in the archived dataset, the present study does not establish direct fixation-to-memory relationships. Instead, the descriptive memory results are interpreted as complementary evidence at a different analytical level.

Limitations

The study used only three underlying poster identities, each presented in two versions. Using the same identities across colour and black-and-white conditions matched the underlying layout, text, depicted objects, and general composition more closely than using different advertisements. Nevertheless, three identities provide limited stimulus-level generalisability. Condition assignment was not documented as random, so colour-condition comparisons do not support strict causal inference.

Image and text AOIs were not matched in physical area, information density, or semantic complexity. Because the archived AOI coordinate files and participant-level gaze coordinates were not preserved, fixation count and duration could not be normalised by AOI surface area, and individual heatmaps or gaze plots could not be reconstructed. Raw image–text differences may therefore partly reflect unequal opportunities to receive a fixation. Questionnaire item difficulty was also not experimentally matched. Only question-level accuracy summaries were retained for the memory questionnaire; the absence of the original participant-by-item response matrix prevented reconstruction of participant-level variability and dependency-aware inferential analysis. The exact response-format documentation, complete answer key, and detailed scoring documentation were also unavailable.

The exact Tobii Pro X-Series (Tobii AB, Danderyd, Sweden) submodel could not be recovered. Although the confirmed 30 Hz sampling rate supported the reported fixation-based analyses, its temporal resolution may limit sensitivity to rapid fixation transitions and other fine-grained gaze dynamics, including reading-related micro-events. The historical records did not retain a numerical calibration acceptance criterion, calibration error, tracking-ratio threshold, gaze-validity percentage, or validity-based exclusion rule. The specific ErgoLAB 3.0 fixation-detection algorithm and its velocity, dispersion, and minimum-duration parameters were also not preserved. These omissions limit complete methodological replication, and the exported fixation metrics cannot retrospectively establish gaze-data validity.

Participant linkage across the three poster blocks was reconstructed from the preserved row order, which the authors confirmed was identical across poster blocks within each presentation condition. Analyses therefore used participant-level averages across the three poster identities. Poster-level observations were retained for descriptive breakdowns, and multilevel modelling at the participant-by-poster level was not performed. Gaze and questionnaire data were not linked at the participant-by-item level, so fixation on a particular poster element cannot be coupled directly to memory for an individual item. Consequently, the memory findings provide complementary question-level descriptive evidence rather

than a direct test of fixation-to-memory relationships. These constraints support exploratory and associative interpretation.

Despite these limitations, the study retains applied value. It combines task-directed viewing, authentic advertising stimuli, AOI-based gaze measures, and an independent memory outcome in an ecologically meaningful visual-communication setting. The matched poster identities reduce content differences between presentation conditions, and the observed discrepancy between gaze allocation and descriptive memory patterns provides a basis for designing larger studies with area-controlled AOIs, preserved trial-level data, and preregistered multilevel analyses.

Future research could further investigate causal relationships by manipulating encoding instructions, retention intervals, and modality-specific viewing tasks. Higher-frequency eye tracking and preserved raw gaze coordinates would also enable more detailed spatial and temporal analyses.

5. Theoretical and Practical Implications

For applied eye-tracking research, the main implication is methodological: raw gaze measures should be interpreted at the correct statistical level and paired with an outcome that represents the purpose of the display. In the present stimulus versions, participant-level gaze analyses showed greater raw visual sampling of text AOIs, while archived question-level memory summaries showed descriptively higher accuracy for image-related than text-related questionnaire items. Designers should not infer communication effectiveness, cognitive effort, or memorability from FC or TFD alone, particularly when AOI areas differ. The pattern can guide hypotheses about image–text hierarchy, but it requires replication with more poster identities and preservation of participant-by-item data for dependency-aware analysis.

6. Conclusions

Across three commercial poster identities presented in colour or black-and-white form, participant-level analyses showed a large and consistent difference in raw gaze allocation, with text AOIs receiving more fixations and greater Total Fixation Duration than image AOIs under explicit memorisation instructions. Archived aggregate question-level summaries showed descriptively higher accuracy for image-related than text-related questionnaire items in both presentation conditions.

The present findings suggest that visual attention allocation and descriptive memory performance may exhibit partially divergent patterns during complex advertising viewing under the current task and stimulus conditions. The participant-level gaze differences do not establish an intrinsic attentional advantage because AOI area may influence raw FC and TFD; gaze was also not linked to individual memory items. Combining eye tracking with an independent memory measure nevertheless provided a more informative account than either measure alone. Future work should preserve complete acquisition and event-detection settings, document randomisation, match or model AOI geometry, and analyse participant-by-poster and item-level data with appropriate multilevel models.

Author Contributions: Conceptualization, C.C. (Chen Chen); methodology, C.C. (Chen Chen) and J.W.; validation, J.W.; formal analysis, J.W.; investigation, J.W. and C.C. (Chenyang Cai); data curation, J.W.; writing—original draft preparation, J.W. and C.C. (Chenyang Cai); writing—review and editing, C.C. (Chen Chen) and J.W.; visualization, J.W.; supervision, C.C. (Chen Chen); project administration, C.C. (Chen Chen). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethics Committee of the College of Furnishings and Industrial Design, Nanjing Forestry University (protocol code 2025023; approval date: 25 December 2025).

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are not publicly available. Requests concerning the archived materials may be directed to the corresponding author and are subject to verification of the retained records, participant-privacy requirements, and third-party copyright restrictions.

Acknowledgments: The authors thank the Laboratory of Human Factors and Ergonomics of Nanjing Forestry University for supporting the experiment. The authors used ChatGPT-6 (OpenAI) during manuscript revision as described in Section 2.8. All AI-assisted outputs were critically reviewed by the authors, who take full responsibility for the final manuscript.

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

References

1. Ye, C.; Guo, L.; Wang, N.; Liu, Q.; Xie, W. Perceptual encoding benefit of visual memorability on visual memory formation. *Cognition* **2024**, *248*, 105810. [CrossRef] [PubMed]
2. Wakeland-Hart, C.D.; Cao, S.A.; DeBettencourt, M.T.; Bainbridge, W.A.; Rosenberg, M.D. Predicting visual memory across images and within individuals. *Cognition* **2022**, *227*, 105201. [CrossRef] [PubMed]
3. Bagheri, E.; Mohsenzadeh, Y. Modeling visual memorability assessment with autoencoders reveals characteristics of memorable images. *Sci. Rep.* **2026**, *16*, 27725. [CrossRef] [PubMed]
4. Lahner, B.; Mohsenzadeh, Y.; Mullin, C.; Oliva, A. Visual perception of highly memorable images is mediated by a distributed network of ventral visual regions that enable a late memorability response. *PLoS Biol.* **2024**, *22*, e3002564. [CrossRef] [PubMed]
5. Grainger, J. Letters, words, sentences, and reading. *J. Cogn.* **2024**, *7*, 66. [CrossRef] [PubMed]
6. Bechtold, L.; Cosper, S.H.; Malyshevskaya, A.; Montefinese, M.; Morucci, P.; Niccolai, V.; Repetto, C.; Zappa, A.; Shtyrov, Y. Brain signatures of embodied semantics and language: A consensus paper. *J. Cogn.* **2023**, *6*, 61. [CrossRef] [PubMed]
7. Higdon, K.F.; Neath, I.; Surprenant, A.M.; Ensor, T.M. Distinctiveness, not dual coding, explains the picture-superiority effect. *Q. J. Exp. Psychol.* **2025**, *78*, 180–191. [CrossRef] [PubMed]
8. Nieznański, M.; Ford, D.; Obidziński, M. Representation of shared surface information and false memory for abstract versus concrete pictures in the conjoint recognition paradigm. *Psychol. Res.* **2024**, *88*, 950–973. [CrossRef] [PubMed]
9. Li, D.; Hu, Y.; Qi, M.; Zhao, C.; Jensen, O.; Huang, J.; Song, Y. Prioritizing flexible working memory representations through retrospective attentional strengthening. *NeuroImage* **2023**, *269*, 119902. [CrossRef] [PubMed]
10. DiPuma, A.; Lockhart, H.A.; Emrich, S.M.; Ester, E.F. Retrospective cue benefits in visual working memory are limited to a single location at a time. *Atten. Percept. Psychophys.* **2023**, *85*, 1474–1485. [CrossRef] [PubMed]
11. Serin, F.; Günseli, E. Internal attention is the only retroactive mechanism for controlling precision in working memory. *Atten. Percept. Psychophys.* **2023**, *85*, 1375–1386. [CrossRef] [PubMed]
12. Zheng, W.; Jia, L.; Sun, N.; Liu, Y.; Geng, J.; Zhang, D. Effects of attention direction and perceptual distraction within visual working memory. *Front. Psychol.* **2022**, *13*, 801252. [CrossRef] [PubMed]
13. Macedo-Pascual, J.; Capilla, A.; Campo, P.; Hinojosa, J.A.; Poch, C. Selection within working memory impairs perceptual detection. *Psychon. Bull. Rev.* **2023**, *30*, 1442–1451. [CrossRef] [PubMed]
14. Liu, R.; Guo, L.; Lin, X.; Nie, D.; Astikainen, P.; Ye, C. Dimension-based retro-cue benefit in working memory does not require unfocused dimension removal. *Front. Psychol.* **2024**, *15*, 1433405. [CrossRef] [PubMed]
15. Yuan, Y.; He, X.; Yue, Z. Working memory load modulates the processing of audiovisual distractors: A behavioral and event-related potentials study. *Front. Integr. Neurosci.* **2023**, *17*, 1120668. [CrossRef] [PubMed]
16. Hessels, R.S.; Nuthmann, A.; Nyström, M.; Andersson, R.; Niehorster, D.C.; Hooge, I.T.C. The fundamentals of eye tracking part 1: The link between theory and research question. *Behav. Res. Methods* **2025**, *57*, 16. [CrossRef] [PubMed]
17. Walter, K.; Bex, P. Cognitive load influences oculomotor behavior in natural scenes. *Sci. Rep.* **2021**, *11*, 12405. [CrossRef] [PubMed]
18. Fadardi, M.S.; Fadardi, J.S.; Mahjoob, M.; Doosti, H. Post-saccadic eye movement indices under cognitive load: A path analysis to determine visual performance. *J. Ophthalmic Vis. Res.* **2022**, *17*, 397–404. [CrossRef] [PubMed]
19. Rodemer, M.; Karch, J.; Bernholt, S. Pupil dilation as cognitive load measure in instructional videos on complex chemical representations. *Front. Educ.* **2023**, *8*, 1062053. [CrossRef]

20. Sáiz-Manzanares, M.C.; Marticorena-Sánchez, R.; Martín-Antón, L.J.; González-Díez, I.; Carbonero-Martín, M.Á. Using eye tracking technology to analyse cognitive load in multichannel activities in university students. *Int. J. Hum.-Comput. Interact.* **2024**, *40*, 3263–3281. [CrossRef]
21. Kafkas, A. Eyes on memory: Pupillometry in encoding and retrieval. *Vision* **2024**, *8*, 37. [CrossRef] [PubMed]
22. Brooks, P.P.; Guzman, B.A.; Kensinger, E.A.; Norman, K.A.; Ritchey, M. Eye tracking evidence for the reinstatement of emotionally negative and neutral memories. *PLoS ONE* **2024**, *19*, e0303755. [CrossRef] [PubMed]
23. Wynn, J.S.; Schacter, D.L. Eye movements reinstate remembered locations during episodic simulation. *Cognition* **2024**, *248*, 105807. [CrossRef] [PubMed]
24. Mitre-Hernandez, H.; Covarrubias-Carrillo, R.; Lara-Alvarez, C. Pupillary responses for cognitive load measurement to classify difficulty levels in an educational video game: Empirical study. *JMIR Serious Games* **2021**, *9*, e21620. [CrossRef] [PubMed]
25. Chakravarthula, P.N.; Suffridge, J.E.; Wang, S. Gaze dynamics during natural scene memorization and recognition. *Cognition* **2025**, *259*, 106098. [CrossRef] [PubMed]
26. Deng, W.; Beck, D.M.; Federmeier, K.D. Image memorability is linked to facilitated perceptual and semantic processing. *Imaging Neurosci.* **2024**, *2*, 1–13. [CrossRef] [PubMed]
27. Huang, J.; Velarde, I.; Ma, W.J.; Baldassano, C. Schema-based predictive eye movements support sequential memory encoding. *eLife* **2023**, *12*, e82599. [CrossRef] [PubMed]
28. Navarro-Báez, S.; Undorf, M.; Bröder, A. Predicting the memorability of scene pictures: Improved accuracy through one's own experience. *Q. J. Exp. Psychol.* **2025**, *78*, 546–565. [CrossRef] [PubMed]
29. Lin, Q.; Li, Z.; Lafferty, J.; Yildirim, I. Images with harder-to-reconstruct visual representations leave stronger memory traces. *Nat. Hum. Behav.* **2024**, *8*, 1309–1320. [CrossRef] [PubMed]
30. Loh, Z.; Hall, E.H.; Cronin, D.; Henderson, J.M. Working memory control predicts fixation duration in scene-viewing. *Psychol. Res.* **2023**, *87*, 1143–1154. [CrossRef] [PubMed]
31. Roberts, B.R.T.; MacLeod, C.M.; Fernandes, M.A. Symbol superiority: Why $ is better remembered than 'dollar'. *Cognition* **2023**, *238*, 105435. [CrossRef] [PubMed]
32. Zhou, L.; Xue, F. Show products or show people: An eye-tracking study of visual branding strategy on Instagram. *J. Res. Interact. Mark.* **2021**, *15*, 729–749. [CrossRef]
33. Simonetti, A.; Bigné, E. Does banner advertising still capture attention? An eye-tracking study. *Span. J. Mark.-ESIC* **2024**, *28*, 3–20. [CrossRef]
34. Liang, S.; Liu, R.; Qian, J. Fixation prediction for advertising images: Dataset and benchmark. *J. Vis. Commun. Image Represent.* **2021**, *81*, 103356. [CrossRef]
35. Guo, R.; Kim, N.; Lee, J. Empirical insights into eye-tracking for design evaluation: Applications in visual communication and new media design. *Behav. Sci.* **2024**, *14*, 1231. [CrossRef] [PubMed]
36. Segijn, C.M. Eye-tracking research in advertising scholarship: A review and practical guide. *J. Advert.* **2025**, *54*, 637–654. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.