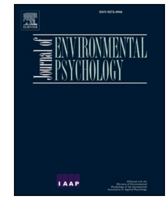
Journal of Environmental Psychology 114 (2026) 103144

Contents lists available at ScienceDirect

Journal of Environmental Psychology

journal homepage: www.sciencedirect.com/journal/journal-of-environmental-psychology



Recurrent visual behavior and viewpoint-specific spatial features in a heritage environment: immersive eye-tracking evidence from Jichang Garden

Yuyang Peng [a], Steffen Nijhuis [a], Yingwen Yu [b,*], Giorgio Agugiaro [a], Guanting Zhang [c]

[a] Department of Urbanism, Delft University of Technology, Julianalaan 134, 2628 BL, the Netherlands
[b] Department of Architectural Engineering and Technology, Delft University of Technology, Julianalaan 134, 2628 BL, the Netherlands
[c] School of Architecture, Nanjing Tech University, No. 30 Puzhu South Road, Jiangsu Province, 211816, China

A R T I C L E I N F O

Keywords:
Visual behavior
Eye-tracking
Heritage gardens
Spatial features
Scene perception
Viewpoint geometry

A B S T R A C T

How multiple dimensions of viewpoint-specific spatial structure relate directionally to recurrent visual behavior in heritage settings remains insufficiently specified at the level of matched scenes. This paper examines this relationship in Jichang Garden, China, by combining immersive eye-tracking with viewpoint-matched spatial features derived from three-dimensional visibility analysis. Gaze data from 51 participants across 22 useable viewpoints were used to identify four recurrent visual behavior patterns, exploratory scanning, hotspot attention, near-far alternation, and route-oriented attention, together with three broader exploration strategies. These behavioral patterns were then related to viewpoint-level spatial features, including openness, depth, complexity, orientation, viewpoint geometry, and scene composition. In the mixed-effects models, higher complexity was associated with more exploratory scanning and less hotspot attention, whereas greater openness was independently associated with less exploratory scanning. Viewpoint geometry showed differentiated behavioral profiles: greater upward visibility was associated with more exploratory scanning but less hotspot attention and near-far alternation, whereas greater eye-level visibility was associated with less route-oriented attention. Greater vegetation visibility was associated with more exploratory scanning, while greater water visibility was associated with more route-oriented attention. Random Forest diagnostics indicated moderate predictive performance and range-dependent relationships, particularly for openness. Brief post-session debriefs provided additional context for selected scenes in which specific focal elements, including a particular tree, koi, and inscriptions, were not represented as separate predictors in the spatial models. Overall, the study provides case-based, scene-level evidence from one heritage setting on how viewpoint-matched spatial features relate to recurrent visual behavior and where additional interpretive context remains useful.

1. Introduction

Research on the relationship between environmental characteristics and patterns of attention and perception has a long tradition in environmental psychology (Kaplan, 1995; Stamps, 2004). Eye-movement data provide a process-oriented means of observing how viewers scan, dwell, and shift across scenes, making it possible to relate visual behavior to environmental characteristics alongside, rather than solely through, post hoc evaluations (Berto et al., 2008; Bollen et al., 2025). Across natural, urban, and designed settings, spatial properties such as openness, depth, complexity, directional structure, visibility, and scene composition have been associated with broad scanning, focal attention, near–far shifts, and route-oriented viewing (Dupont et al., 2015; Frost et al., 2022; Gentin, 2011; Liu et al., 2021).

In heritage settings, these viewing processes occur in scenes where spatial form coexists with historically and culturally meaningful features (Ginzarly et al., 2019; Peng, Li et al., 2026; Taylor & Lennon, 2011). Yet recurrent visual behavior and multidimensional spatial structure are rarely examined within a common viewpoint-matched framework (Aalto & Steinert, 2025; Evans & Chamberlain, 2024). Consequently, it remains unclear how multiple dimensions of spatial structure correspond to recurrent viewing patterns and where spatial descriptions

* Corresponding author.
E-mail addresses: Y.Peng-1@tudelft.nl (Y. Peng), S.Nijhuis@tudelft.nl (S. Nijhuis), christinayu@tudelft.nl (Y. Yu), g.agugiaro@tudelft.nl (G. Agugiaro), gtzhang@njtech.edu.cn (G. Zhang).

https://doi.org/10.1016/j.jenvp.2026.103144
Received 13 September 2025; Received in revised form 15 July 2026; Accepted 17 July 2026
Available online 18 July 2026

become less informative for interpreting scene-specific attention.

1.1. Beyond fixation summaries: recurrent visual behavior

Visual behavior provides a process-level account of environmental perception because fixations form temporally ordered sequences. Eye-tracking can reveal broad scanning, localized attention, near–far alternation, and route-oriented viewing, whereas aggregate measures such as fixation count, total dwell time, or hotspot concentration may obscure transitions among these forms of engagement (Berto et al., 2008; Bollen et al., 2025; Chen et al., 2022; Foulsham & Underwood, 2008; Williams & Castelhano, 2019). A pattern-based account provides a complementary level of analysis. It permits recurrent forms of exploratory, focal, depth-oriented, and route-oriented viewing to be identified within individual trials, while recognizing that several forms of behavior may occur during the same viewing period. The co-occurrence and relative prominence of these recurrent patterns can also characterize broader exploration strategies across participants and scenes. This two-level representation distinguishes the organization of gaze within a trial from higher-level tendencies in scene engagement.

Although immersive displays and eye-tracking have increasingly been applied to environmental perception (Peng et al., 2026; Qiu et al., 2023; Yuan et al., 2025), recurrent behavior patterns and broader exploration strategies remain insufficiently characterized across coherent sequences of viewpoint-matched heritage scenes. Establishing these behavioral structures is therefore necessary before their relationships with environmental characteristics can be examined.

1.2. From global 2D descriptors to viewpoint-specific 3D visibility

Classic research in environmental psychology and landscape perception has examined properties such as openness, depth, complexity, legibility, enclosure, and directional structure in relation to environmental appraisal and visual engagement (Herzog, 1992; Kaplan & Kaplan, 1989; Stamps, 2004, 2005). These properties represent different dimensions of spatial structure and should not be assumed to exhibit uniform relationships with visual behavior.

The behavioral relevance of spatial features may vary in direction, magnitude, and form. For example, complexity and openness represent conceptually different aspects of a visual environment. Complexity concerns the diversity and spatial variation of visible information, whereas openness concerns the extent of unobstructed visual space. The two properties may therefore show contrasting rather than equivalent relationships with exploratory and focused viewing. Relationships involving depth, orientation, and scene composition may likewise be selective or nonlinear rather than reducible to a common effect.

A further limitation concerns the dimensional representation of visibility. Much previous research has relied on photographs, horizontal measures, or global scene descriptors. Such representations provide limited differentiation among upward, eye-level, downward, and overhead visibility fields. Yet these vertical components correspond to different environmental conditions, including canopy and sky exposure, upper architectural structure, route continuity, near-field ground surfaces, and vertically layered views. Combining them into a single measure of overall visibility may therefore obscure behaviorally relevant variation.

Point-cloud-based three-dimensional visibility analysis permits these components to be examined separately while retaining their spatial relationship to the observer (Lonergan & Hedley, 2016; Wróżyński et al., 2024). When spatial features are calculated from the same physical viewpoints used in the behavioral experiment, the method also reduces mismatch between the scene from which visual behavior is recorded and the scene from which environmental indicators are derived.

These limitations raise the question of whether global spatial indicators and vertically differentiated visibility fields exhibit distinct relationships with recurrent visual behavior.

1.3. Spatial regularities and focal salience in heritage scenes

Heritage environments are both spatially organized and interpretively charged (Peng et al., 2024; Smith, 2006; Taylor & Lennon, 2011). In addition to enclosure, depth, visibility, and directional structure, these environments contain inscriptions, rockeries, symbolic plantings, architectural details, water reflections, animals, and other elements that may attract attention because of their distinctiveness, memorability, historical significance, or cultural meaning (Dewhurst et al., 2018; Di Pietro et al., 2018; Uzzell, 1996; Zou et al., 2023).

Visual-spatial and interpretive studies have generally examined these influences separately. Visual-spatial research tends to emphasize measurable environmental characteristics (Peng et al., 2025), whereas heritage interpretation research more often foregrounds meaning, identity, memory, and visitor understanding (Di Pietro et al., 2018; Poria et al., 2009). This separation limits understanding of how spatial structure and the attentional salience of individual scene elements operate together within the same viewing experience.

A distinction is therefore required between broad regularities in visual engagement and the selection of specific focal targets. Spatial geometry may be informative for whether viewing is exploratory, locally concentrated, organized through near–far alternation, or directed along a path or view axis. By contrast, the selection of a particular scene element may additionally depend on its perceived distinctiveness or meaning (Henderson & Hayes, 2018; Hwang et al., 2011).

Spatial geometry may therefore account more fully for broad engagement modes than for specific focal-target selection, particularly in scenes containing distinctive or meaning-laden elements. Post-session debriefs can provide complementary context for scenes that are less fully described by spatial variables.

1.4. Research question and directional predictions

The preceding sections identify three linked problems. First, recurrent visual behavior and broader exploration strategies remain insufficiently characterized across matched heritage scenes. Second, the directional and potentially nonlinear relationships between multidimensional spatial features and recurrent behavior remain incompletely specified, particularly for vertically differentiated visibility. Third, the relative relevance of spatial geometry and specific heritage elements to broad engagement modes and focal-target selection remains unclear. Accordingly, the study addresses one overarching research question: ***How do recurrent visual behavior patterns and broader exploration strategies vary across viewpoint-matched heritage scenes, and how do multiple dimensions of three-dimensional spatial geometry and scene composition relate to these behaviors?***

To specify the principal relationships examined within this broader question, the analysis is organized around three directional predictions:

Prediction 1: When depth, openness, orientation, and complexity are modeled jointly, greater complexity is expected to be associated with more exploratory scanning and less hotspot attention, whereas greater openness is expected to be associated with less exploratory scanning.

Prediction 2: Upward and eye-level visibility are expected to exhibit differentiated behavioral profiles rather than functioning as interchangeable measures of overall visibility. Greater upward visibility is expected to be associated with more exploratory scanning but less hotspot attention and near–far alternation, whereas greater eye-level visibility is expected to be associated with less route-oriented attention.

Prediction 3: Spatial models are expected to provide a more complete account of broad engagement modes than of specific focal-target selection. Accordingly, selected scenes showing comparatively larger discrepancies between predicted and observed behavior are expected to contain recurrently mentioned distinctive, memorable, or meaning-laden focal elements.

Predictions 1 and 2 are examined quantitatively, whereas **Prediction 3** is evaluated descriptively using model diagnostics and post-session debriefs. The empirical analysis is conducted in Jichang Garden, Wuxi, China, a classical Chinese Jiangnan private garden characterized by sequential spatial organization and substantial scene-to-scene variation in enclosure, layering, vertical visibility, and scene composition. Jichang Garden provides a coherent setting for matched within-case comparison, allowing recurrent visual behavior and multidimensional spatial features to be examined at the scene level without implying statistical representativeness. By combining immersive eye-tracking with viewpoint-matched three-dimensional visibility analysis, the study provides a scene-level examination of recurrent visual behavior, differentiated spatial associations, and the interpretive limits of spatial description within this heritage setting.

2. Methods

To address the study question, we adopted a case-based analytical design organized around three linked steps (Fig. 1a). First, immersive VR eye-tracking was used to record scanpaths at matched viewpoints, from which recurrent visual behavior patterns and broader exploration strategies were derived. Second, point-cloud-based visibility analysis was used to derive viewpoint-specific spatial features for the same scenes, including visibility structure and scene composition. Third, cross-classified mixed-effects models and descriptive diagnostics were used to examine associations between behavior and spatial features, while brief post-session debriefs were used to provide interpretive context for selected scenes that were less clearly described by spatial features alone.

2.1. Study setting and viewpoint selection

Jichang Garden, located inside Xihui Park at the foot of Huishan Mountain in Wuxi, Jiangsu, China, covers approximately 0.99 ha. As a classical Chinese Jiangnan private garden, it includes route-based spatial sequencing, marked shifts between enclosed and open scenes, layered near-far compositions, and a range of memorable scene features

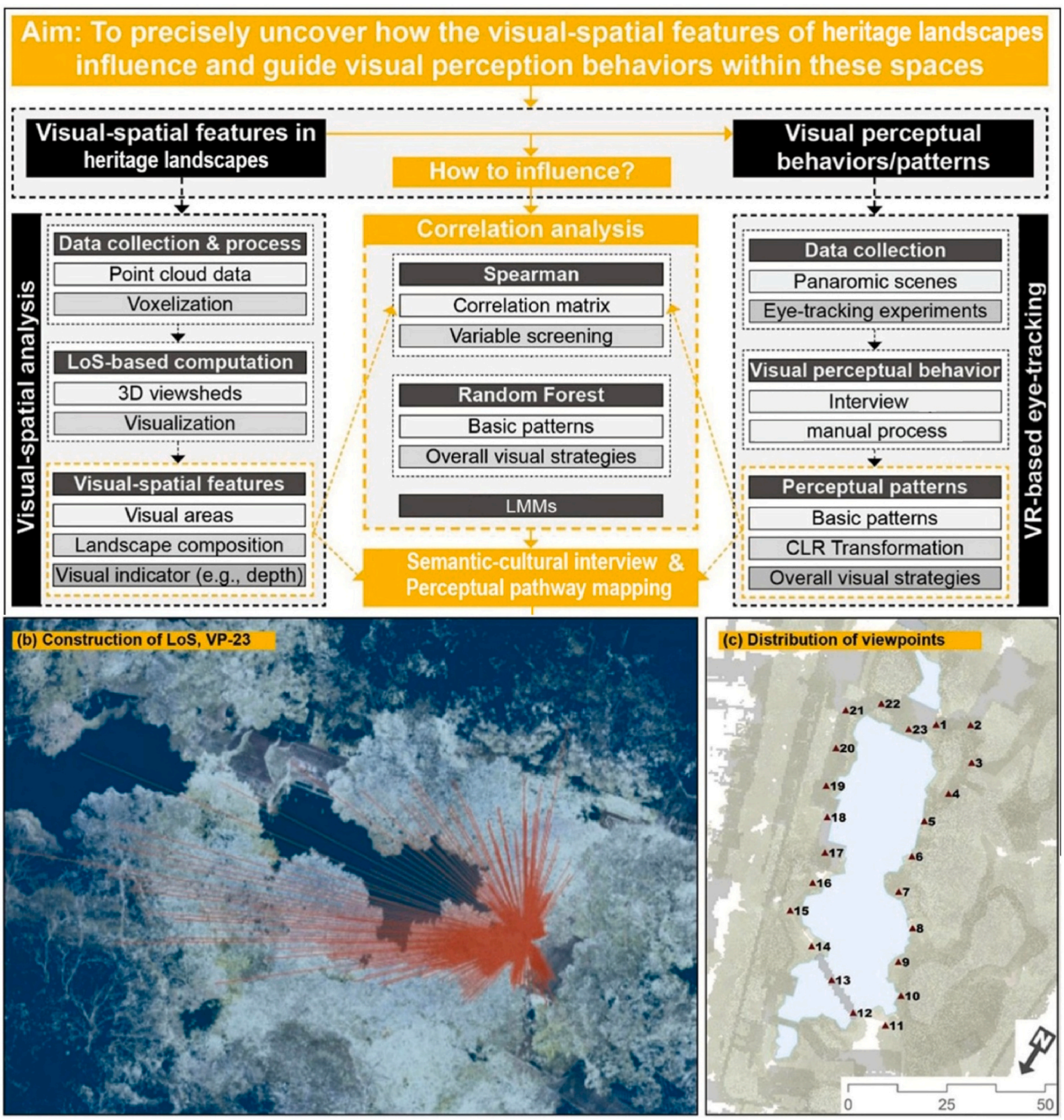


Fig. 1. Study workflow: (a) analytical overview of the study; (b) illustration of LoS (Line-of-Sight) construction in the point-cloud environment, taking viewpoint No. 23 as an example; (c) study area, Jichang Garden, and the 23 viewpoints used in this study.

within a compact setting (Keswick, 2003). The site is organized around a spring-fed pond and meandering waterways, with rockeries and Taihu stones forming near-field relief, while pavilions, bridges, framed openings, and borrowed views structure sightlines and visual transitions (Shu et al., 2018). Dense tree canopy and layered shrub planting coexist with scene features such as inscribed plaques, distinctive stones, and pavilion architecture (Yang, 2022). Together, these characteristics offered a suitable basis for matched viewpoint comparison.

Twenty-three viewpoints were selected along the main experiential sequence of the garden, especially along waterfront pathways, in order to capture variation in enclosure, openness, depth, route continuity, and scene composition (Fig. 1c). The same viewpoints were used in the point-cloud analysis and the VR panorama experiment so that spatial features and visual behavior could be compared scene by scene.

2.2. Participants

Because the study materials depicted a Jiangnan heritage garden, recruitment was restricted to Chinese participants to reduce broad cross-cultural heterogeneity in scene interpretation. Approximate gender balance was targeted during recruitment. Participants were recruited mainly through university instructors near the study site who circulated the study invitation to students; a small proportion were additionally recruited through the researchers' acquaintances. All participants received RMB 30 in cash per session. The study protocol was approved by the institutional ethics committee of blind-for-review, and all participants provided informed consent.

To reduce expertise-related variation in spatial viewing, participants from architecture, landscape architecture, urban planning, environmental design, and related art and design fields were not recruited. All included participants had received their formal education in China and had no formal training in spatial or visual design. All had previously visited one or more classical gardens in Suzhou, indicating general experiential familiarity with Jiangnan classical gardens. However, none had previously visited Jichang Garden or reported prior awareness of the site. The sample therefore had general familiarity with this type of heritage environment but no site-specific familiarity or professional design expertise.

Participants reported normal or corrected-to-normal vision, little or no prior head-mounted VR experience, and the ability to complete an HMD-VR viewing task. To manage VR tolerance, participants completed a brief pre-session screening checklist addressing dizziness or motion-sickness susceptibility. During the experiment, discomfort symptoms such as nausea, dizziness, and eye strain were checked verbally after calibration and between blocks, and participants could pause, take breaks, or withdraw at any time.

Following eligibility screening and initial data-quality assessment, 60 volunteers were enrolled. Seven did not yield analyzable experimental datasets because stable eye-tracker calibration could not be obtained after repeated attempts, gaze validity remained insufficient for analysis, or the experimental procedure was not completed because of discomfort or early termination. The remaining 53 participants completed the VR procedure and the post-session debrief. During subsequent participant-level quality review, two additional eye-tracking datasets were excluded because persistent recording irregularities across viewpoints prevented reliable fixation detection and trajectory coding. The final eye-tracking analytic sample therefore comprised 51 participants, whereas debrief data were available for all 53 participants who completed the procedure.

Demographic characteristics of the 53 participants who completed the VR procedure and post-session debrief are summarized in Table 1. Participants were aged around 18–35 and were approximately balanced by gender. Age and years of formal education are reported as estimates because exact individual values were not retained. The final eye-tracking sample of 51 participants remained within the range commonly reported in previous HMD-VR eye-tracking studies, including samples of approximately 46 (Ke et al., 2023), 50 (Jeong et al., 2022), and 51 useable records (Pfaller et al., 2021). A post hoc sensitivity analysis (two-tailed paired-samples *t*-test, $\alpha = 0.05$, $1 - \beta = 0.80$) indicated that n = 51 provides 80% power to detect within-participant effects of approximately $dz \approx 0.40$.

Table 1
Demographic characteristics of the participants.

Characteristic	Male (n = 28)	Female (n = 25)	Overall (N = 53)
Age group, n (%)			
18–25 years	13 (46.4)	12 (48.0)	25 (47.2)
26–30 years	13 (46.4)	12 (48.0)	25 (47.2)
31–35 years	2 (7.1)	1 (4.0)	3 (5.7)
Age, estimated M ± SD, years	**25.3 ± 3.9**	**25.1 ± 3.6**	**25.2 ± 3.7**
Academic status, n (%)			
Undergraduate-level participants[a]	15 (53.6)	14 (56.0)	29 (54.7)
Postgraduate-level participants[b]	11 (39.3)	11 (44.0)	22 (41.5)
PhD holders	2 (7.1)	0 (0.0)	2 (3.8)
Formal education, estimated M ± SD, years	**16.8 ± 2.6**	**16.5 ± 2.3**	**16.7 ± 2.5**

Note: Demographic characteristics refer to the 53 participants who completed the VR procedure and post-session debrief. Eye-tracking analyses were based on 51 participants after participant-level recording-quality review. Values are reported as estimated M ± SD for continuous variables and n (%) for categorical variables.

[a] Undergraduate-level participants included current undergraduate students and bachelor's-degree holders who had not continued to postgraduate study.

[b] Postgraduate-level participants included current master's students and PhD candidates who had completed a master's degree.

2.3. Procedure and data acquisition

To capture visual behavior under matched viewing conditions, gaze data were recorded in VR panoramas using an HTC VIVE Pro Eye headset with integrated Tobii eye-tracking (120 Hz; approximately 0.5°–1.1° accuracy). Each visual stimulus consisted of a static 360° panoramic image presented in the HMD as an immersive surrounding scene rather than as a conventional flat image or video. The participant's virtual viewing position was fixed at the center of the corresponding panorama, while natural head and eye movements allowed exploration of the full scene. The integrated eye-tracking system recorded gaze direction relative to the displayed panorama, and Tobii Pro Lab registered fixation locations on the corresponding panoramic stimulus, from which scanpaths and fixation heatmaps were generated. The eye-tracking data were not projected fixation by fixation onto the point cloud. Instead, correspondence between the behavioral and spatial data was established at the viewpoint level: each panoramic stimulus and the corresponding point-cloud-based visibility analysis represented the same physical observation location and shared the same viewpoint identifier. Gaze data were processed in Tobii Pro Lab using the default I-VT Attention Filter (velocity threshold = 100°/s; minimum fixation duration = 60 ms; other parameters at software defaults). The same preprocessing and event-detection settings were applied across all participants and viewpoints.

Participants completed the experiment indoors under a standardized procedure (Fig. 2). They first received instructions for free viewing and general VR safety. The eye tracker was then calibrated in Tobii Pro Lab and recalibrated when needed. Participants completed a short acclimation session using practice panoramas that were not included in the formal analyses.

During the formal trials, participants viewed 23 viewpoint-matched panoramas for 30 s each. Each scene began from a standardized starting orientation toward the lake so that participants shared the same initial viewing direction at a given viewpoint. Viewpoint order was fixed from VP-1 to VP-23 to preserve the sequential route experience of the garden.

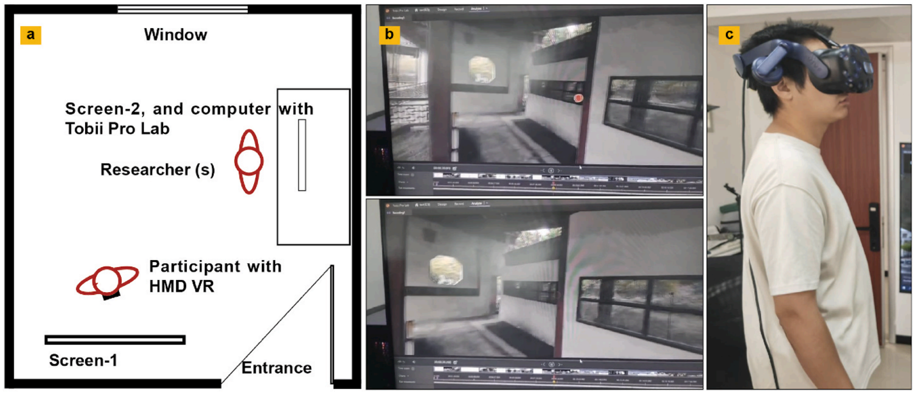


Fig. 2. Experimental setup for the VR eye-tracking study: (a) laboratory layout and equipment placement; (b) photos of the Tobii Pro Lab setup; (c) participant wearing the HMD with integrated eye-tracking.

No explicit search or memory task was assigned; participants were asked simply to explore each scene freely.

Trials were excluded if calibration failed or if Tobii Pro Lab's built-in quality checks indicated insufficient validity for analysis (e.g., excessive tracking loss). One viewpoint (VP3) was removed from the eye-tracking analyses because its recordings were consistently of low quality.

2.4. *Characterizing visual behavior patterns and broader exploration strategies*

This subsection derives two complementary behavioral summaries from the eye-tracking data. At the trajectory level, each participant-by-viewpoint trial was coded as a composition of recurrent visual behavior patterns rather than assigned to a single category. At a broader level, these trial-level compositions were clustered into exploration strategies. The pattern proportions served as the primary behavioral outcomes, while the clusters provided a higher-level summary of scene engagement.

Step 1. Recurrent visual behavior patterns

Valid gaze trials (N $=$ 1122; 22 viewpoints $\times$ 51 participants, excluding VP3) were manually coded separately for each participant and viewpoint. Two trained raters independently inspected the ordered fixations, scanpath diagrams, fixation heatmaps, and supportive metrics including total dwell time, scanpath length, and fixation concentration.

Following fixation order, each trial was segmented when the viewing behavior changed. Each segment was coded as broad exploratory scanning, sustained hotspot attention, near–far alternation, or route-oriented following, and its dwell-time contribution was used to estimate the four within-trial pattern proportions, which summed to 100%. Thus, one trial could contain multiple patterns. Raters were blind to spatial features and study aims. For reliability assessment, the pattern with the largest proportion in each trial was treated as the dominant categorical label, yielding Cohen's $\kappa = 0.82$. Disagreements in segmentation or proportional allocation were resolved through discussion.

Step 2. Broader exploration strategies

To summarize higher-level tendencies across trials, pattern proportions were treated as compositional data. After adding a small pseudo-count ($\varepsilon = 1 \times 10^{-6}$) to handle zeros and renormalizing to sum to one, the data were transformed using centered log-ratios (CLR). K-means clustering was then applied to the CLR-transformed features, and the number of clusters was selected by maximizing the average silhouette score across candidate solutions (k = 2–6). These clusters were interpreted as broader exploration strategies and are reported in the Results.

2.5. *Deriving viewpoint-specific spatial features from point-cloud data*

Point-cloud analysis was used to characterize the spatial features available from each selected viewpoint. Compared with conventional 2D visibility methods, point-cloud data allow matched estimation of three-dimensional visibility structure and visible scene composition in historic landscapes (Lonergan & Hedley, 2016; Wróżyński et al., 2024). In the present study, this step was used to derive viewpoint-specific spatial features for subsequent association analyses.

Step 1. Data collection and preprocessing

Point-cloud data were collected using two complementary scanning approaches: handheld scanning and stationary scanning. Handheld scanning was used in narrow or irregular spaces such as secondary pathways, whereas stationary scanning provided higher accuracy in key areas such as waterfront zones. The resulting datasets were stored in RGB-colored PLY format for visual analysis and non-RGB LAS format for spatial measurement and structural analysis (Zhang & Yuyang, 2024).

During preprocessing, the point clouds were semantically segmented into bridges, vegetation, ground, buildings, and rockeries. Data density was reduced through grid sampling and voxelization based on grid-center points (Alsadik et al., 2014). A voxel size of 0.1 m was used to balance computational efficiency and spatial precision. Semantic labels were harmonized across the datasets before aggregation, and each voxel inherited the semantic class of its center point to ensure a single, reproducible class assignment. Water surfaces, which were not reliably captured because of specular reflection, were represented as flat planes and treated as valid geometric surfaces in later visibility computation. When intersected by LoS rays, these surfaces were assigned the class "water" rather than "sky."

Step 2. 3D viewshed computation

The 23 viewpoints selected for the VR study were also used for the viewshed analysis (Fig. 1c). For each viewpoint, LoS (Line-of-Sight) rays were generated at horizontal intervals of 5°, covering azimuth angles from 0° to 355°, and at vertical angles from −80° to 85°, approximating the human visual field (Fig. 1b). Maximum LoS length was set to 100 m to provide a consistent upper bound for comparability across viewpoints and to avoid over-emphasizing distant visibility beyond the scale of the garden.

For each ray j, we recorded the nearest intersection distance d_j (truncated at $d_{max} = 100$ m), an unobstructed flag u_j, and the semantic class c_j of the intersected object (voxel class or water plane). Rays with no intersection within d_{max} were coded as unobstructed ($u_j = 1$) and assigned the class "sky"; otherwise, $u_j = 0$. These ray-level records formed the common input for all subsequent spatial features.

To visualize visibility structure, the endpoints of LoS rays across vertical angles were connected to construct closed 3D viewsheds (Peng et al., 2024). These surfaces were used for descriptive visualization and for deriving vertical-band visibility measures. All scalar indicators reported in the Results were calculated from the underlying ray-level records.

Step 3. Spatial and scene-compositional features

Two types of features were derived from the ray-level records: scene-compositional features and spatial geometry.

For scene composition, we calculated the proportion of rays intersecting each visible class, including vegetation, water, buildings, rockeries, ground, bridges, and sky. Element proportions were computed as ray-count proportions, $p_c = n_c/N$, where n_c is the number of rays intersecting class c and N is the total number of rays per viewpoint; unobstructed rays contributed to the “sky” class.

For spatial geometry, we summarized visibility across vertical viewing bands reflecting typical human viewing ranges (Liu, 2020; Nijhuis, 2015): eye-level viewing (ELV, 5° to −5°), slightly downward (SDV, −6° to −30°), downward (look-down view, LDV, −31° to −70°), slightly upward (SUV, 6° to 15°), upward (look upward view, LUV, 16° to 50°), and overhead space (“Top”, 51° and above). These bands were used to describe viewpoint geometry in a perception-relevant way and to allow comparison across viewpoints.

Step 4. Core visual-spatial indicators

Four core indicators were used in the main analyses to summarize viewpoint-specific spatial geometry: openness, depth, orientation, and complexity. Their final formulas and interpretations are listed in Table 2, with full derivations provided in Appendix A1.

Depth describes visible layering and distance variation within a scene (Kaplan, 1987; Liu & Nijhuis, 2020). It was calculated through a layered computational procedure that produced a Layered Depth Index (LDI), an Absolute Depth Index (ADI), and a composite Depth Index (DI). Higher DI values indicate deeper and more layered scenes.

Orientation describes directional concentration in the horizontal visual field (Liu & Nijhuis, 2020). It was quantified using directional entropy so as to reflect both the number of visible directions and the extent of visibility along those directions. Higher values indicate stronger directional concentration and clearer spatial guidance.

Openness describes the proportion of visually unobstructed space within the viewing field (Stamps, 2005). It was calculated using a spherical area approximation of unobstructed rays (Inglis et al., 2022; Zhao et al., 2020). Higher values indicate more expansive views.

Complexity describes visual richness arising from diversity and spatial arrangement of scene features (Kuper, 2017). It was calculated using a weighted Shannon entropy approach, yielding a composition entropy (SEC), a scale-variation entropy (SES), and a composite Complexity Index (CI). Higher CI values indicate richer and more complex scenes (Stamps, 2003, 2004).

2.6. Association analyses

Associations between spatial features and visual behavior were examined using a staged analysis strategy with clearly differentiated roles for the statistical methods. Cross-classified linear mixed-effects models (LMMs) served as the primary inferential analyses. Spearman rank correlations were used only as descriptive diagnostics for monotonic trends and redundancy among spatial features. Random Forest (RF) models were used as secondary, non-inferential diagnostics to summarize nonlinear tendencies, predictor importance, and moderate predictive performance (Bzdok et al., 2018; Fife & D’Onofrio, 2023). For analysis, each participant-by-viewpoint eye-tracking trial was paired, through the shared viewpoint identifier, with the spatial-feature vector derived for the corresponding viewpoint.

Table 2
Visual-spatial indicators, formulas, and interpretations.

Indicator	Final formula	Principle and interpretation (details in Appendix A1)
Openness	Open = (Σ j u j w j)/(Σ j w j)	**Principle:** Fraction of the sampled viewing dome that is unobstructed (sky-visible) under the LoS sampling grid. **Interpretation:** Open ∈ [0,1]; higher Open means a larger unobstructed solid-angle share and a more expansive view.
Layered Depth Index (LDI)	LDI = H · CV, H = −Σ k p k ln(p k), CV = σ(d)/μ(d)	**Principle:** Combines how evenly visibility is distributed across distance layers (entropy of depth bins) with variability in viewing distances (CV). **Interpretation:** LDI ≥0; higher LDI indicates stronger near–far layering and greater depth variation.
Absolute Depth Index (ADI)	ADI = d vis max/d max	**Principle:** Normalized visible extent using the maximum observed viewing distance relative to the LoS maximum range. **Interpretation:** ADI ∈ [0,1]; higher ADI indicates longer visible extent.
Depth Index (DI)	DI = 0.5 · LDI∼ + 0.5 · ADI∼	**Principle:** Composite depth measure balancing layering (LDI) and extent (ADI) across viewpoints. **Interpretation:** Higher DI indicates deeper and more layered scenes.
Orientation (O)	O = 1 − H dir/ln(N sec), H dir = −Σ i p i ln(p i), p i = L i/Σ i L i	**Principle:** Directional concentration of visibility, measured as one minus normalized directional entropy of LoS lengths across azimuth sectors. **Interpretation:** O ∈ [0,1]; higher O indicates a clearer dominant direction and stronger spatial guidance; lower O indicates more isotropic visibility.
Composition entropy (SEC)	SEC = −Σ c p c ln(p c), p c = n c/N	**Principle:** Diversity of visible semantic elements quantified by Shannon entropy of class proportions. **Interpretation:** SEC ≥0; higher SEC indicates a more diverse and balanced element composition.
Scale-variation entropy (SES)	SES = −Σ c w c ln(w c), w c = CV c/Σ c CV c, CV c = σ(d|c)/μ(d|c)	**Principle:** Multi-scale richness quantified by entropy over class weights derived from within-class distance variability. **Interpretation:** SES ≥0; higher SES indicates stronger multi-scale variation across classes.
Complexity Index (CI)	CI = 0.5 · SEC∼ + 0.5 · SES∼	**Principle:** Composite complexity measure combining element diversity (SEC) and multi-scale variation (SES) after min–max normalization across viewpoints. **Interpretation:** If sub-indices are normalized to [0,1], then CI ∈ [0,1]; higher CI indicates richer and more complex views.

Step 1. Cross-classified structure and variance partitioning

Because observations were cross-classified repeated measures (participants by viewpoints), we first fitted random-intercepts models with random intercepts for participant and viewpoint. Intraclass correlations (ICCs) were used to assess the relative contribution of viewpoint-level versus participant-level variation. This step provided a rationale for emphasizing viewpoint-level spatial features in the subsequent models while retaining participant-level random effects.

Step 2. Primary inferential analyses using LMMs

Cross-classified LMMs were used as the primary inferential analyses to estimate associations between spatial features and visual behavior outcomes. Outcomes expressed as compositions were analyzed on the CLR scale, and continuous predictors were z-standardized. To avoid overloading a single model while retaining the theoretical distinction among spatial dimensions, the inferential analyses were organized into three separate feature-set model families.

The core-indicator models included depth, openness, orientation, and complexity as fixed effects and were used to evaluate **Prediction 1**. The viewpoint-geometry models included look-down visibility, eye-level visibility, and look-upward visibility as fixed effects, with particular emphasis on the differentiated profiles of eye-level and look-upward visibility specified in Prediction 2. The scene-composition models examined the visible proportions of dominant landscape elements as an additional feature set. All models included crossed random intercepts for participant and viewpoint to account for between-observer and between-scene variability.

Fixed-effect estimates with 95% confidence intervals, random-effect variance components, and marginal and conditional R^2 following Nakagawa were reported. Multiple comparisons within each outcome were controlled using the Benjamini–Hochberg false discovery rate procedure, with complete model estimates provided in Appendix A5.

Step 3. Descriptive diagnostics

Spearman rank correlation matrices were computed as descriptive diagnostics to summarize bivariate monotonic trends and identify redundancy among spatial features. When features were highly correlated, representative variables were retained to reduce redundancy and improve interpretability in the secondary diagnostic analyses.

RF models were then fitted as secondary descriptive tools to summarize nonlinear response patterns and moderate predictive performance. RF regression was used for continuous outcomes, and RF classification was used for categorical strategy outcomes where appropriate. Model performance was evaluated using cross-validation and summarized with standard metrics (e.g., R^2, MAE, MSE, and classification accuracy). Predictor importance was assessed using permutation importance, and partial dependence or accumulated local effects were inspected to summarize broad response shapes. These RF results were interpreted descriptively and were used mainly to identify scene ranges or patterns that were less clearly described by the mixed-effects associations.

2.7. Post-session debriefs for selected scenes

In addition to the quantitative analyses, brief post-session debriefs were conducted as a complementary interpretive source. Immediately after each VR scene block, participants were shown a preview of their gaze output, including heatmaps and scanpath visualizations, and were asked which scene features or overall scene qualities had drawn their attention and why (Hyrskykari et al., 2008).

Responses were recorded on a structured tally sheet and summarized by feature-mention frequency, salience rank (primary or secondary), and stated reason. Free-form responses were assigned to predefined feature categories, including architecture, inscriptions or plaques, rockeries or distinctive stones, water and reflections, vegetation or symbolic plantings, animals, and other. Reasons were categorized as aesthetic qualities, historical-cultural meaning, novelty or distinctiveness, restorative or relaxing feelings, or spatial guidance. Ambiguous responses were clarified during the debrief.

All 53 participants who completed the VR procedure also completed the debrief. For the descriptive evaluation of **Prediction 3**, viewpoints with RF cluster-classification accuracy below 70% were treated as lower-accuracy cases. VP5 and VP19 were selected from this group as illustrative cases for detailed interpretation. Scene selection was based on model performance and did not incorporate the debrief content. Model accuracy was assessed using the 51 valid eye-tracking datasets, whereas feature-mention frequencies were summarized from all 53 debrief responses.

The debriefs were used only to provide descriptive context for the selected scenes and were not treated as a parallel inferential source.

3. Results

This section first characterizes recurrent visual behavior patterns and broader exploration strategies observed in the VR eye-tracking task. It then summarizes the main spatial differences across the matched viewpoints, examines associations between spatial features and visual behavior, and finally uses post-session debriefs to interpret selected scenes in which observed patterns were less clearly described by spatial features alone.

3.1. Recurrent visual behavior patterns and broader exploration strategies

Eye-tracking experiments in immersive VR generated dynamic gaze data for 22 viewpoints, excluding VP3 because low-quality recordings did not provide reliable gaze data. The processed outputs included fixation heatmaps, trajectory diagrams, and fixation-duration summaries, which together supported the characterization of recurrent visual behavior patterns and broader exploration strategies (Fig. 3a–d, Appendix A4).

Across participants and viewpoints, four recurrent visual behavior patterns were identified through the trajectory-level coding procedure described in Section 2.4 and Appendix A2. The panels in Fig. 3e–h are included as illustrative examples of the visual signatures used during coding. Overall inter-rater agreement for the dominant trial-level pattern was high (Cohen's $\kappa = 0.82$). The exploratory scanning pattern (ESP) was characterized by broad, irregular trajectories and dispersed fixations, indicating wide-ranging visual exploration without sustained concentration on specific scene features. The hotspot attention pattern (HAP) was characterized by short trajectories and localized fixation clusters, indicating sustained attention to distinct scene features. The near–far alternation pattern (NFAP) was marked by repeated shifts between foreground and background features, suggesting sensitivity to depth contrasts. The route-oriented attention pattern (ROAP) followed the direction of paths or view axes and was concentrated toward vanishing points or central perspectives.

The occurrence of these patterns varied across viewpoints (Appendix A4). ROAP was more prevalent at VP5, VP7, and VP13, where linear pathways structured visual engagement. NFAP was relatively frequent at VP5 and VP17, corresponding to stronger depth contrasts. HAP dominated at VP1 and VP18–20, where prominent scene features attracted concentrated fixations. ESP was more evident at VP11 and VP13–15, suggesting scenes that encouraged extensive scanning with limited focus (Fig. 4a–d).

Based on k-means clustering of pattern compositions, three broader exploration strategies were identified. Environmental exploration (Cluster 0, n = 364) was dominated by ESP and reflected broad scene scanning without strong fixation. Specific element attention (Cluster 1, n = 316) was characterized by high HAP values and reflected targeted viewing of distinctive features with limited exploratory scanning. Holistic appreciation (Cluster 2, n = 442) combined HAP with NFAP,

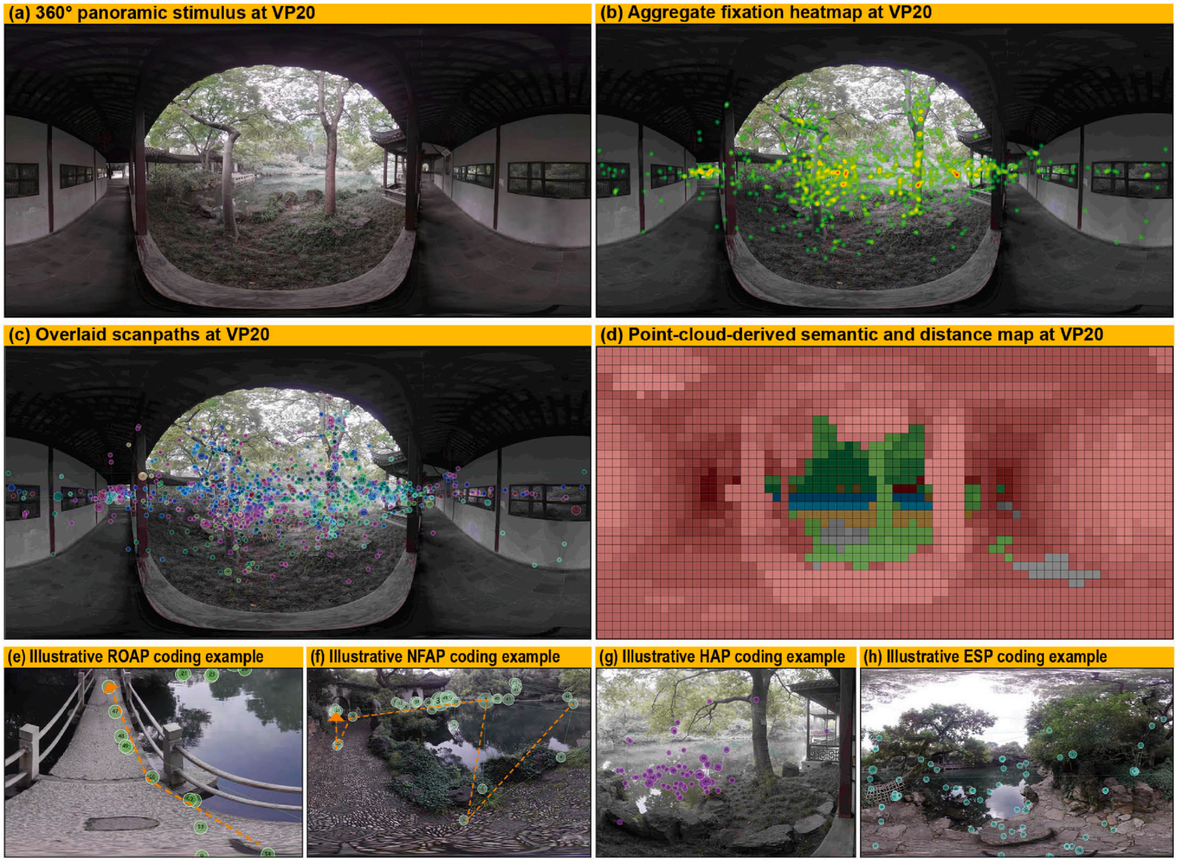


Fig. 3. Example of the eye-tracking experiment and illustrative scanpaths used in gaze-pattern coding: (a) panoramic scene of VP20; (b) aggregate fixation heatmap across valid participants for VP20; (c) aggregate eye-movement trajectories across valid participants for VP20; (d) visualization of VP20's spatial features and landscape elements; (e) an illustrative example of ROAP; (f) an illustrative example of NFAP; (g) an illustrative example of HAP; and (h) an illustrative example of ESP. Panels (e)–(h) are provided only to clarify the visual characteristics used during coding and are not intended as group-average trajectories. The full coding protocol and annotated segmentation example are provided in Appendix A2.

indicating both concentrated attention on landmarks and alternation between near and far elements, while route-following tendencies remained weak. These strategies also varied across viewpoints: some scenes, such as VP5 and VP19, showed a relatively balanced mix of strategies, whereas others, such as VP7, VP9, and VP10, were dominated by a single mode (Fig. 4e and f).

Taken together, these results indicate that visual exploration in the garden was not uniform. Instead, viewers showed recurrent and interpretable behavior patterns that provided the basis for the subsequent analyses of spatial associations.

3.2. Spatial characteristics of the matched viewpoints

The point-cloud analysis showed clear variation in spatial features across the 23 selected viewpoints (Fig. 5). These differences provided the environmental contrast used in the subsequent association analyses.

Visibility structure varied substantially across viewpoints. For example, VP7 and VP11 displayed exceptionally high upward and overhead visibility, suggesting relatively open vertical fields, whereas VP1 and VP2 showed much more limited vertical visibility, reflecting more introverted, architecturally bounded scenes (Table 3). Horizontal openness also varied, with viewpoints such as VP10 and VP16 offering more extended sightlines, whereas VP17 and VP2 presented more restricted visual volumes.

Scene composition likewise varied across viewpoints (Table 4). Some viewpoints, such as VP17–21, were dominated by buildings, while others, such as VP7–12, emphasized vegetation and waterscapes. Rockeries were especially prominent at VP7–9, and waterscapes at VP12–13, illustrating the diversity of scene composition within the garden.

The core spatial indicators also showed substantial spread across viewpoints (Table 5). *Openness* ranged from relatively expansive values at VP11 (0.4122) to very enclosed values at VP17 (0.0131). *Depth* indices were relatively high at VP1, VP4, and VP12, suggesting stronger visual layering, whereas VP20–22 showed simpler depth structure. *Complexity* was especially pronounced at VP12–13 and relatively low at VP17 and VP21.

Rather than being interpreted on their own, these spatial differences serve here as viewpoint-specific scene descriptors for the analyses below.

3.3. Associations between spatial features and visual behavior

To examine how visual behavior varied with viewpoint-specific spatial features, we related the core spatial indicators to the four visual behavior patterns and the three broader exploration strategies. Variance partitioning showed that viewpoint-level differences accounted for more variation than participant-level differences across the outcomes, supporting an emphasis on viewpoint-level features while retaining participant as a random intercept.

3.3.1. Mixed-effects associations

Across the four behavior patterns, the primary LMMs identified complexity as the most consistent core spatial indicator (Table 6). Higher complexity was associated with a marked increase in exploratory scanning (ESP) and with a decrease in focused hotspot attention (HAP).

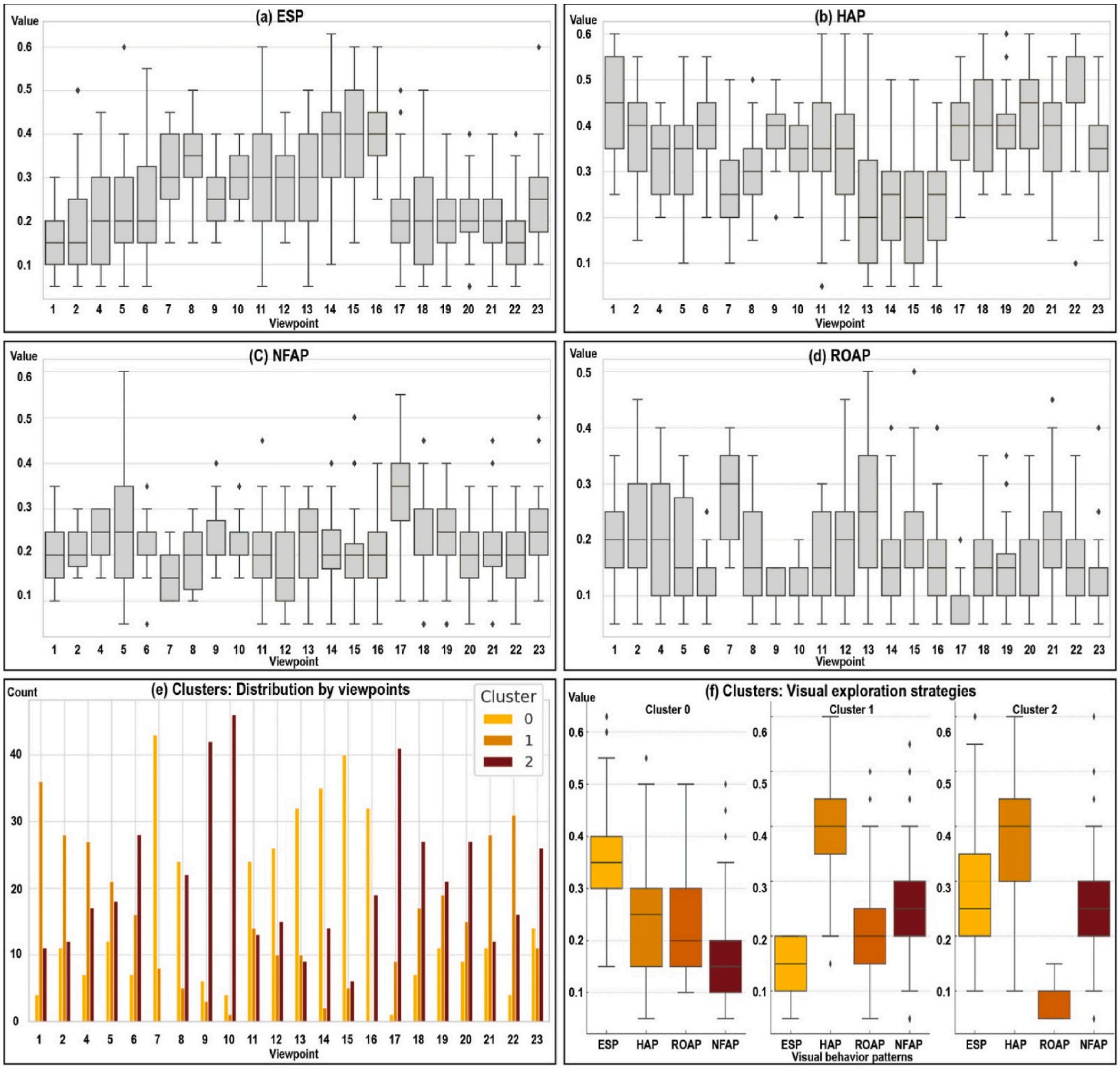


Fig. 4. Analysis results of visual behavior patterns and broader exploration strategies: (a)–(d) values and viewpoint distributions of ESP, HAP, NFAP, and ROAP; (e) clustering of exploration strategies and viewpoint distribution; (f) value distribution of each pattern within the three clusters.

Openness was negatively associated with ESP, indicating less exploratory scanning in more open scenes. Orientation was negatively associated with NFAP, suggesting lower near–far alternation where directional structure was stronger. Depth showed only a weak, trend-level positive association with HAP. The complete models also identified additional associations involving NFAP and ROAP. These results are reported in Appendix A5.2.

In parallel with the core-indicator models, the viewpoint-geometry models evaluated the differentiated vertical-visibility profiles specified in Prediction 2. Look-upward visibility (LUV) and eye-level visibility (ELV) showed non-equivalent relationships across the four recurrent visual behavior patterns (Table 7). Greater LUV was associated with more ESP and ROAP but less HAP and NFAP. Greater ELV was associated with more ESP and NFAP, less ROAP, and no clear association with HAP. These findings were consistent with Prediction 2 and indicate differentiated behavioral profiles rather than a simple one-to-one correspondence between a vertical visibility field and a single behavior pattern. The scene-composition models provided an additional feature set. Their principal associations are summarized in Table 8, with complete estimates reported in Appendix A5.

3.3.2. *Random Forest diagnostics*

After redundancy screening based on Spearman correlations (Appendix A5.1), RF analyses were used as secondary descriptive diagnostics to summarize nonlinear tendencies and predictive performance (Fig. 6). Models for ESP and HAP showed moderate predictive strength ($R^2 \approx 0.27$–0.28), while NFAP and ROAP remained weakly predicted ($R^2 < 0.20$). Key features for ESP included vegetation, complexity, depth, and limited building dominance, whereas HAP was more strongly associated with building presence, low vegetation, and openness extremes.

Local analyses further suggested that ESP was more common in scenes with moderate *openness* (0.2–0.3), *vegetation* above 40%, higher *complexity* (>0.6), and layered *depth* (>0.8). HAP tended to occur in scenes with lower *vegetation* (<0.3–0.4), lower *complexity* (<0.6), and either very low (<0.1) or relatively high (>0.3) *openness* (Figs. 7 and 8). These RF results are treated here as descriptive summaries of broad response shapes rather than as primary inferential evidence.

At the level of broader exploration strategies, RF classification identified *orientation*, *complexity*, vegetation, *depth*, and *openness* as the strongest discriminators between the three clusters (Fig. 9). Lower values of these features were more often associated with environmental exploration (Cluster 0), intermediate values with specific element

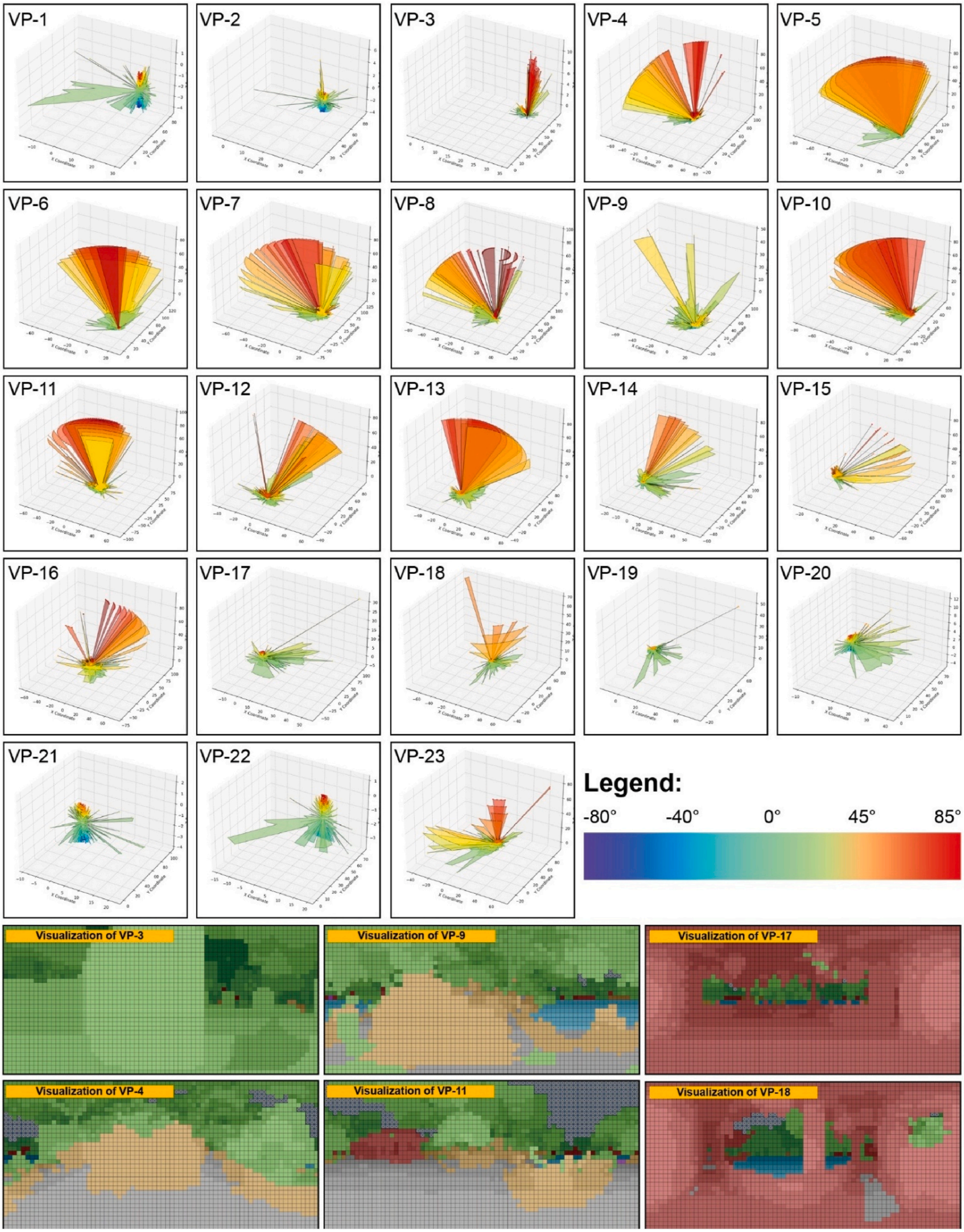


Fig. 5. Computed spatial features for the 23 viewpoints.
Upper panels: visualization of 3D viewsheds. **Lower panels:** visualization of selected panoramic environments, where hue represents element type and brightness indicates distance (red = buildings and structures, blue = waterscape, green = vegetation, purple = bridge, light gray = ground, yellow = rockery, dark gray = sky). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Results of visual area measurements.

VP	LDV	SDV	ELV	SUV	LUV	Top
1	35.3	602.9	3346.8	225.0	140.6	29.0
2	60.7	412.9	804.1	268.0	349.8	17.7
4	15.7	395.6	2840.3	3627.7	50591.4	34290.1
5	32.2	237.5	2797.9	4535.9	109932.1	33939.1
6	36.5	254.7	2980.4	4104.4	59447.0	17167.0
7	57.3	402.5	6309.3	9984.2	124560.7	33631.7
8	47.1	263.8	3608.0	9048.0	63684.4	40158.9
9	35.45	294.7	6332.8	11876.0	21015.5	579.5
10	46.0	439.4	8352.4	4474.1	81440.9	56180.4
11	73.4	731.6	5309.0	7842.1	150555.1	82058.8
12	36.4	383.3	6702.7	14947	53936.8	31698.8
13	41.0	533.6	7254.9	8393.2	93722.6	50432.5
14	54.6	447.9	5455.6	8429.7	28639.6	1472.0
15	55.7	417.9	1549.4	1837.3	64166.9	9901.4
16	65.5	690.6	8853.0	6692.4	93511.1	40239.9
17	52.6	256.5	9161.7	12594.0	3106.7	115.8
18	39.5	332.8	6609.6	3083.8	39499.8	98.3
19	65.1	528.4	4387.4	593.5	1631.3	53.5
20	61.7	487.7	4028.1	3439.9	1154.4	47.3
21	67.9	551.7	3420.6	275.5	180.4	34.6
22	45.9	472.5	1471.0	115.0	132.0	25.7
23	60.4	463.9	8132.2	13571.0	35206.2	14645.3

attention (Cluster 1), and higher values with holistic appreciation (Cluster 2). Overall classification accuracy remained moderate, indicating that spatial features were informative for broad tendencies but did not fully describe every scene-specific strategy.

3.3.3. Summary of spatial associations

The analyses suggest that spatial features did not contribute in a uniform way. The primary LMMs identified relatively stable associations for the core indicators of *complexity*, *openness*, *orientation*, and, at a trend level, *depth*, whereas the RF diagnostics suggested more nonlinear response patterns for *openness*, vegetation, *complexity*, and *depth*. The viewpoint-geometry LMMs revealed differentiated vertical profiles across the behavior patterns. LUV was positively associated with ESP and ROAP but negatively associated with HAP and NFAP, whereas ELV was positively associated with ESP and NFAP, negatively associated with ROAP, and showed no clear association with HAP. These results were consistent with **Prediction 2** and indicate that vertical visibility fields should not be treated as interchangeable measures of overall visibility.

3.4. Post-session debriefs for selected scenes

Post-session debriefs provided descriptive context for the three exploration strategies. Across these strategies, participants broadly distinguished dispersed attention to vegetation-dominated or spatially layered scenes, attention to distinctive individual elements, and appreciation of overall scene coherence. Reported reasons included aesthetic or restorative qualities, novelty or distinctiveness, historical-cultural meaning, and spatial layering.

Among the lower-accuracy viewpoints identified using the criterion described in Section 2.7, VP5 and VP19 are presented as two illustrative cases (Fig. 9b, including 10 VPs). At VP5, 31 of the 53 debrief participants (58.5%) identified a specific tree as a salient or focal element, while 36 participants (67.9%) mentioned koi. At VP19, 24 participants (45.3%) identified inscriptions or plaques as focal elements. Because participants could mention more than one element, these percentages are not mutually exclusive.

The responses at VP5 converged on the tree and koi, whereas those at VP19 converged on inscriptions or plaques (Fig. 10). These focal elements were not represented as separate predictors in the spatial models and therefore provided additional context for the discrepancies between the predicted and observed exploration-strategy profiles. These observations were descriptively consistent with Prediction 3. However, focal-target selection was not modeled as a separate quantitative outcome; the debrief findings should therefore be interpreted as complementary descriptive evidence rather than as a causal or independent inferential test.

3.5. Descriptive synthesis across analyses

Taken together, the results were consistent with **Predictions 1** and **2** and provided qualified descriptive support for Prediction 3. Prediction 1 was reflected in the contrasting adjusted associations of complexity and openness with exploratory viewing, although the RF diagnostics indicated that the openness relationship was range-dependent rather than uniformly linear. **Prediction 2** was reflected in the differentiated behavioral profiles of upward and eye-level visibility. **Prediction 3** received more limited support because it was evaluated through selected model discrepancies and post-session debriefs rather than through a separate quantitative model of focal-target selection. Table 9 summarizes the key evidence underlying these evaluations.

Table 4
Results of proportions of visible landscape elements in different scenarios.

VP	Sky	Ground	Water	Vegetation	Buildings	Rockery	Bridge
1	0.0000	0.0874	0.0093	0.0459	0.8364	0.0210	0.0000
2	0.0000	0.0125	0.0004	0.0878	0.8815	0.0177	0.0000
4	0.0383	0.2655	0.0040	0.4500	0.0020	0.2401	0.0000
5	0.0826	0.2103	0.1120	0.4093	0.0117	0.1741	0.0000
6	0.0500	0.2744	0.0919	0.3936	0.0105	0.1797	0.0000
7	0.0463	0.1160	0.0375	0.4525	0.0048	0.3413	0.0016
8	0.0584	0.0322	0.0185	0.5302	0.0012	0.3594	0.0000
9	0.0040	0.1023	0.0548	0.4996	0.0048	0.3340	0.0004
10	0.0774	0.1567	0.0766	0.3791	0.0109	0.2977	0.0016
11	0.1261	0.3634	0.0044	0.3292	0.0471	0.1289	0.0008
12	0.0210	0.0375	0.2119	0.5077	0.0125	0.0334	0.1761
13	0.0830	0.0016	0.3143	0.3973	0.0153	0.0113	0.1773
14	0.0093	0.2023	0.0250	0.5165	0.0052	0.2361	0.0056
15	0.0278	0.2671	0.0016	0.4065	0.0669	0.2297	0.0004
16	0.0572	0.2917	0.0250	0.4311	0.0516	0.1434	0.0000
17	0.0012	0.0004	0.0073	0.0608	0.9271	0.0032	0.0000
18	0.0137	0.0262	0.0290	0.0802	0.8493	0.0016	0.0000
19	0.0004	0.0403	0.0137	0.0496	0.8618	0.0342	0.0000
20	0.0000	0.0161	0.0121	0.0705	0.8892	0.0121	0.0000
21	0.0000	0.0356	0.0006	0.0379	0.9243	0.0017	0.0000
22	0.0000	0.0266	0.0060	0.1092	0.8533	0.0048	0.0000
23	0.0193	0.2393	0.0431	0.3638	0.3243	0.0101	0.0000

Table 5
Results of visual indicator computation.

VP	Openness	Depth			Orientation	Complexity		
		LDI	ADI	DI		SEC	SES	CI
1	0.1752	3.7412	1.0000	0.9829	0.0056	0.6285	1.8468	0.5952
2	0.0168	1.6892	0.3564	0.3228	0.0375	0.4542	1.3384	0.4310
4	0.2281	3.8434	1.0000	1.0177	0.2365	1.2136	1.4128	0.6315
5	0.3861	3.1010	0.595	0.6866	0.1333	1.5010	1.2498	0.6614
6	0.3626	3.2526	1.0000	0.9154	0.1478	1.4471	1.4077	0.6864
7	0.2306	2.9771	1.0000	0.8676	0.0500	1.2772	1.5549	0.6810
8	0.1040	3.5776	1.0000	0.9767	0.1045	1.0629	1.4687	0.6087
9	0.1387	2.9972	1.0000	0.8711	0.0982	1.1565	1.4370	0.6236
10	0.2745	3.4443	0.6457	0.7714	0.1088	1.4731	1.5859	0.7355
11	0.4122	2.6921	1.0000	0.8183	0.0436	1.4326	1.6305	0.7365
12	0.2247	3.6270	1.0000	0.9802	0.0696	1.3512	1.7599	0.7481
13	0.3534	3.1369	1.0000	0.8953	0.0337	1.3687	1.7322	0.7456
14	0.1701	3.0526	1.0000	0.8807	0.0973	1.1976	1.5733	0.6663
15	0.2266	3.2881	0.3302	0.5866	0.0826	1.3504	1.5290	0.6924
16	0.3062	2.9593	1.0000	0.8646	0.0553	1.4094	1.5757	0.7178
17	0.0131	1.9232	1.0000	0.6851	0.1177	0.3060	1.0612	0.3287
18	0.0819	3.4322	0.5937	0.7433	0.1423	0.6083	1.5415	0.5169
19	0.0680	1.2995	0.7129	0.4335	0.0776	0.5839	1.5197	0.5058
20	0.0366	1.1235	0.5569	0.3250	0.1262	0.4647	1.5259	0.4786
21	0.0419	0.8833	0.4791	0.2445	0.0480	0.4357	1.3331	0.4253
22	0.0206	0.8547	0.666	0.3330	0.0472	0.5303	1.2891	0.4375
23	0.2446	2.6352	1.0000	0.8084	0.0590	1.3334	1.5158	0.6851

Table 6
Selected fixed-effect estimates from primary LMMs for core spatial indicators.

Pattern	Core spatial feature	β	95% CI	q (BH)
ESP	Complexity	0.44	[0.34, 0.55]	<0.001
ESP	Openness	−0.28	[-0.39, −0.17]	<0.001
HAP	Complexity	−0.17	[-0.26, −0.07]	0.002
HAP	Depth (DI)*	0.04	[-0.00, 0.08]	0.089
NFAP	Orientation	−0.05	[-0.08, −0.02]	0.001
ROAP	—	—	—	—

Notes: Only effects meeting the reporting threshold ($q < 0.10$) are shown, together with one theory-relevant trend-level depth effect ($q = 0.089$).

Table 7
Strongest mixed-effects associations for viewpoint-geometry features across visual behavior patterns.

Pattern	ELV, β (95% CI)	q	LUV, β (95% CI)	q
ESP	0.072 [0.045, 0.099]	<0.001	0.136 [0.112, 0.159]	<0.001
HAP	−0.002 [−0.026, 0.022]	0.856	−0.163 [−0.170, −0.156]	<0.001
NFAP	0.059 [0.034, 0.084]	<0.001	−0.085 [−0.105, −0.066]	<0.001
ROAP	−0.129 [−0.159, −0.098]	<0.001	0.113 [0.086, 0.140]	<0.001

Notes: ELV = eye-level visibility; LUV = look-upward visibility. β values are fixed-effect estimates from cross-classified LMMs with CLR-transformed outcomes and z-standardized predictors. The models included LDV, ELV, and LUV simultaneously, with crossed random intercepts for participant and viewpoint. q values are Benjamini–Hochberg-adjusted within each outcome. Complete estimates are reported in Appendix A5.3.

4. Discussion

Three findings organize the discussion: the contrasting relationships of complexity and openness with recurrent visual behavior, the differentiated behavioral profiles of vertically defined visibility, and the complementary contribution of post-session debriefs to selected model discrepancies. The following sections interpret these findings in relation to prior work and consider their contribution, practical implications, and evidential limits.

Table 8
Strongest mixed-effects associations for scene-compositional features across visual behavior patterns.

Pattern	Strongest scene-compositional feature	β	95% CI	q
ESP	Vegetation	0.190	[0.138, 0.242]	<0.001
HAP	Building	0.104	[0.017, 0.190]	0.056
NFAP	—	—	—	—
ROAP	Water	0.061	[0.026, 0.096]	0.002

Notes: For each pattern, the table reports the scene-compositional feature with the smallest adjusted q value among those meeting the reporting threshold. Full estimates are reported in Appendix A5.4.

4.1. Spatial patterns in relation to prior work

At a broad level, the results are consistent with earlier environmental-psychology research showing that visual engagement varies with spatial organization. More specifically, the present findings distinguish visual complexity from spatial openness. In the joint models, higher complexity was associated with more exploratory scanning and less hotspot attention, whereas greater openness was associated with less exploratory scanning. This contrast indicates that informational richness and unobstructed visual extent should not be treated as interchangeable dimensions of scene structure (Herzog & Bryce, 2007; Kaplan & Kaplan, 1989; Stamps, 2004; Torralba et al., 2006). By estimating these relationships at matched viewpoints, the study extends broader accounts of scene organization and visual engagement to a sequentially experienced heritage setting (Dupont et al., 2014; Henderson, 2003).

The viewpoint-geometry results provide a further distinction. Greater look-upward visibility was associated with more ESP but less HAP and NFAP, whereas greater eye-level visibility was associated with less ROAP. Rather than reflecting a common effect of overall visibility, these profiles indicate that different portions of the vertical visual field corresponded to different forms of visual engagement. This form of differentiation is difficult to identify using predominantly horizontal or two-dimensional scene descriptions. Scene-compositional associations were more selective: vegetation was most clearly associated with ESP,

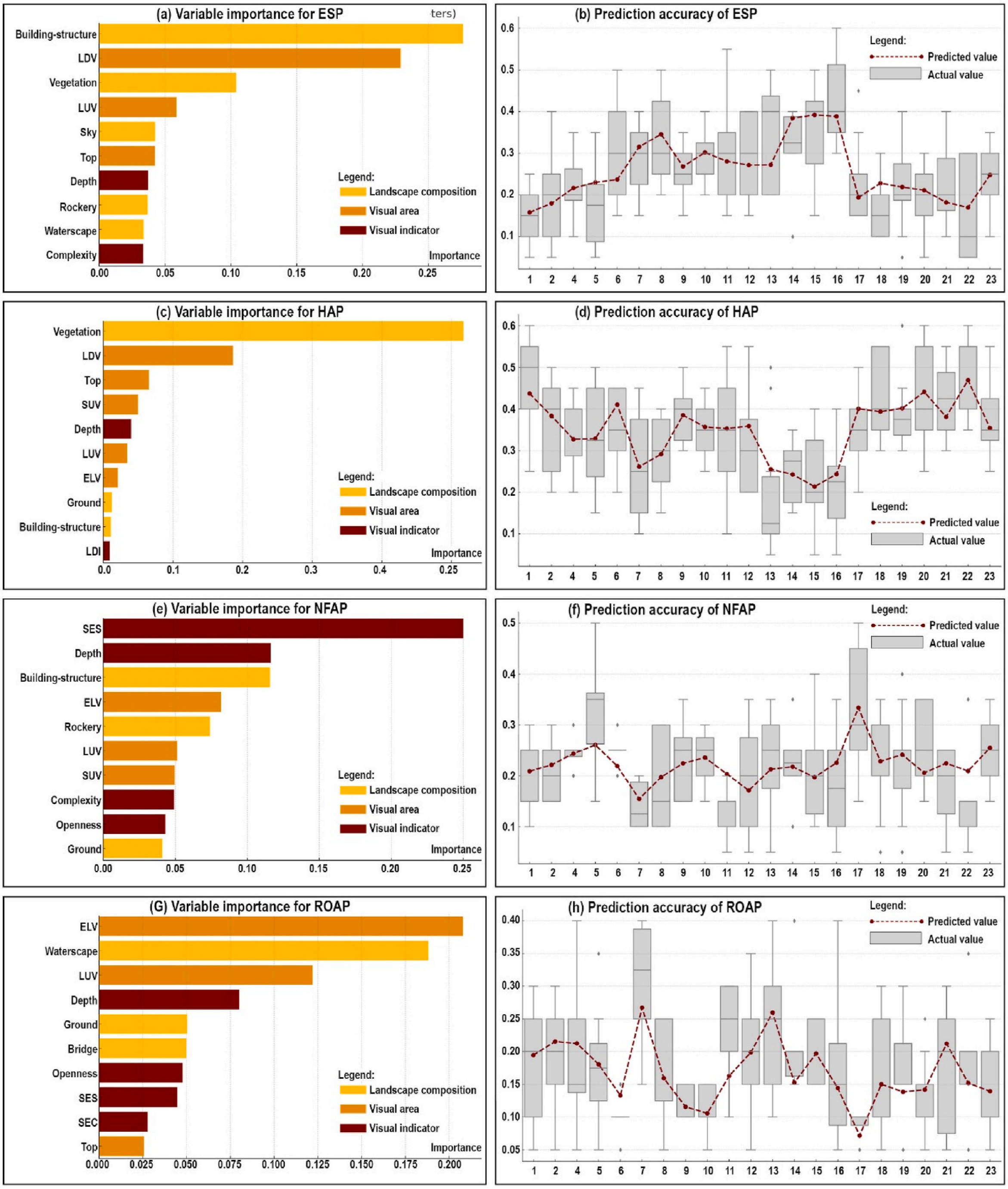


Fig. 6. RF analysis results: (a), (c), (e), and (g) factor importance results for the four visual behavior patterns; (b), (d), (f), and (h) comparisons between predicted and observed values for the four patterns.

water with ROAP, and building visibility more weakly with HAP. The results therefore identify a multidimensional association structure rather than a single spatial feature that uniformly explains visual behavior.

The RF diagnostics qualified the average mixed-effects associations by showing that some relationships varied across the observed feature ranges, particularly those involving openness and complexity, and to a lesser extent depth and vegetation. This is consistent with previous suggestions that enclosure, complexity, and scene organization may not operate uniformly across environmental conditions (Kaplan & Kaplan, 1989; Liu & Nijhuis, 2020, 2022; Nasar, 1994). In particular, the relatively elevated ESP values at moderate openness do not contradict the negative adjusted LMM coefficient. Instead, they indicate that an average directional association can coexist with range-dependent variation.

Finally, the post-session debriefs provided additional context for the selected discrepancies at VP5 and VP19. Participants repeatedly identified a tree, koi, and inscriptions or plaques as focal elements that were not represented as separate predictors in the spatial models (Henderson & Hayes, 2018; Hwang et al., 2011; Tilley, 2006). These observations are consistent with the possibility that spatial geometry provides a more complete account of broad engagement tendencies than of specific focal-target selection. However, because focal targets were not modeled as an independent quantitative outcome, the debrief evidence should be

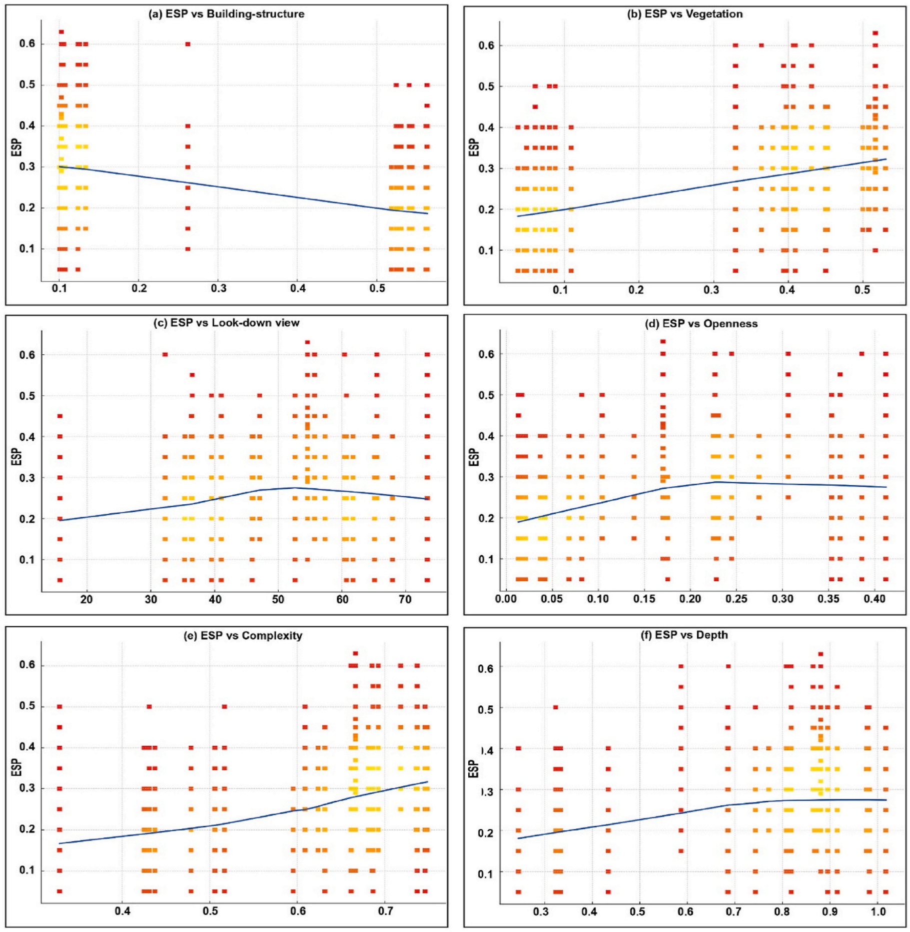


Fig. 7. RF local analysis of important variables associated with ESP: (a) building structure; (b) vegetation; (c) look-down view; (d) openness; (e) complexity; (f) depth.

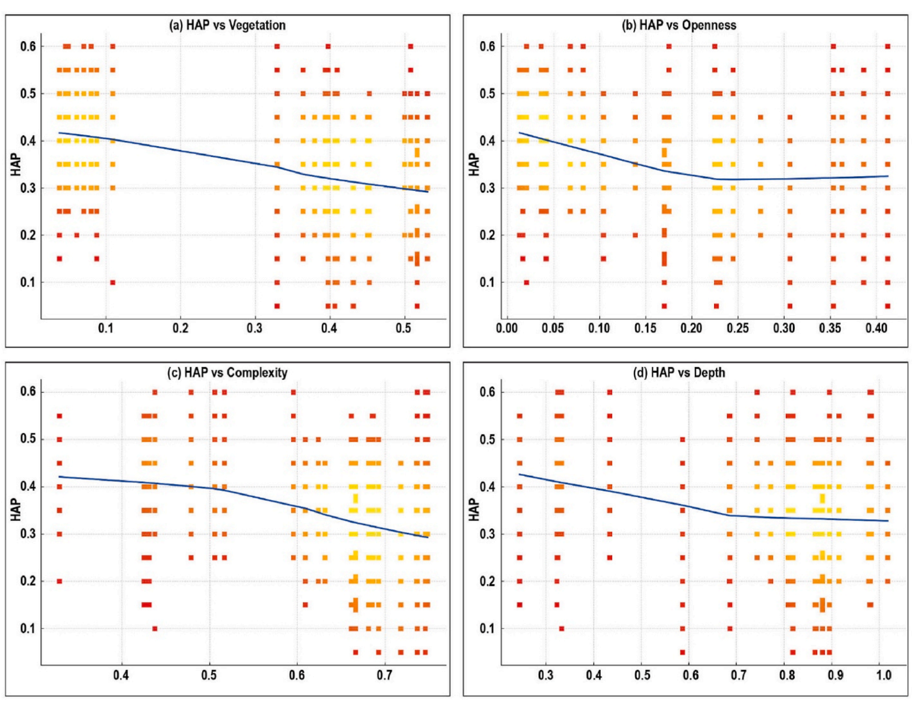


Fig. 8. RF local analysis of important variables associated with HAP: (a) vegetation; (b) openness; (c) complexity; (d) depth.

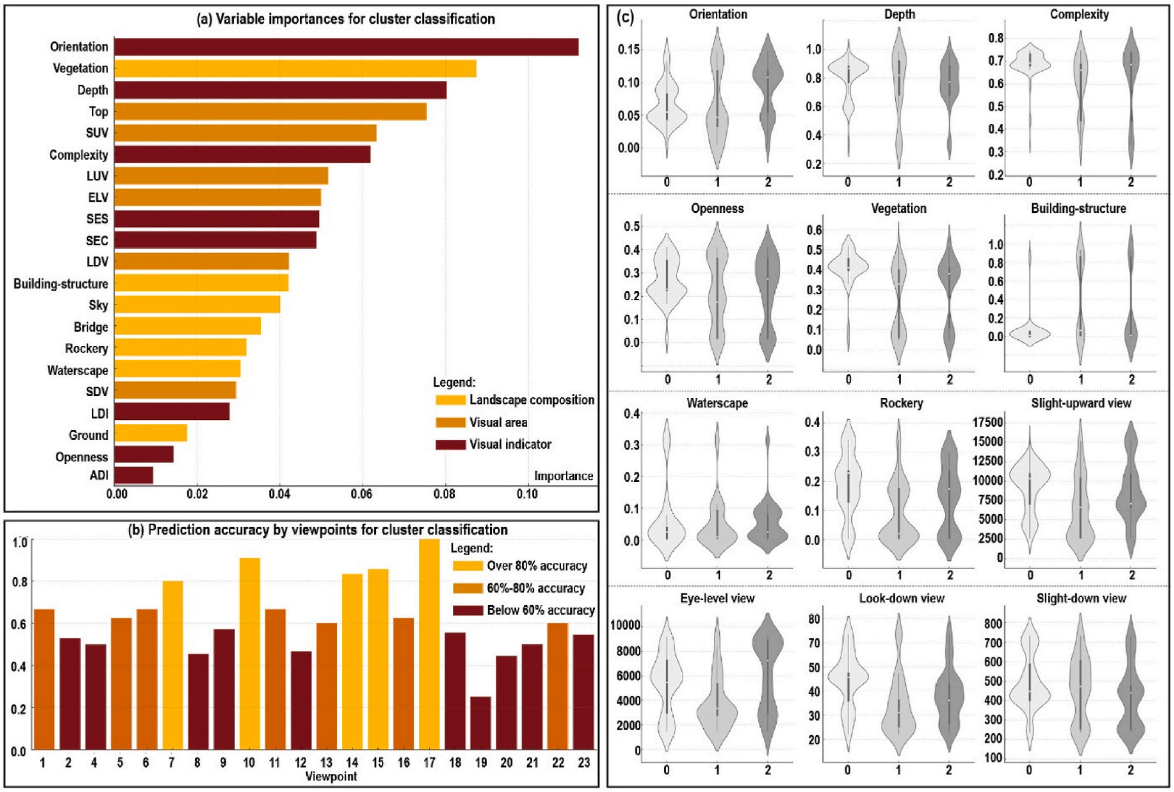


Fig. 9. RF analysis results for broader exploration strategies: (a) factor importance; (b) prediction accuracy and corresponding viewpoints; (c) local analysis of effective variables.

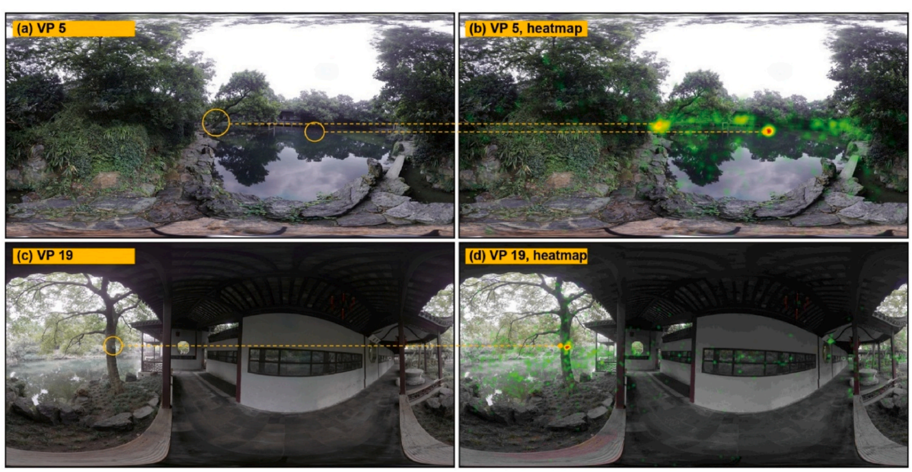


Fig. 10. High-discrepancy scenes selected for the descriptive evaluation of Prediction 3: (a) panoramic scene and (b) aggregate fixation heatmap at VP5; (c) panoramic scene and (d) aggregate fixation heatmap at VP19.

interpreted as complementary and descriptive rather than as a direct comparative or causal test.

4.2. Focused contribution and practical implications

The main value of the study lies in providing a finer-grained account of scene–behavior correspondence in a heritage setting. By matching recurrent visual behavior to viewpoint-specific spatial features, the analysis shows how scene-level differences in visual engagement can be described in relation to spatial geometry and scene composition at the level of individual scenes rather than only at the level of broader environmental categories. It also shows that spatial features were not equally informative: some features, especially complexity and openness, were more consistently associated with visual behavior, whereas others were more selective and scene-dependent. The contribution further lies in distinguishing contrasting relationships among global spatial indicators, particularly complexity and openness, and differentiated profiles among vertical visibility fields, particularly upward and eye-level visibility. Rather than identifying a single dominant spatial predictor, the findings indicate that different dimensions of viewpoint structure correspond to different configurations of recurrent visual behavior. In this respect, the study helps specify where spatial description is more informative and where additional interpretive context may still be needed (Peng et al., 2026).

Table 9
Key evidence relevant to the three directional predictions.

Directional prediction	Key evidence
Prediction 1: Complexity and openness will show contrasting relationships with exploratory and focused viewing.	In the joint core-indicator LMMs, complexity was positively associated with ESP (β = 0.44, 95% CI [0.34, 0.55], q < 0.001) and negatively associated with HAP (β = −0.17, 95% CI [−0.26, −0.07], q = 0.002). Openness was negatively associated with ESP (β = −0.28, 95% CI [−0.39, −0.17], q < 0.001). RF diagnostics indicated that ESP was relatively frequent at moderate openness (0.20–0.30), suggesting a range-dependent rather than uniformly linear openness relationship.
Prediction 2: Upward and eye-level visibility will exhibit differentiated behavioral profiles.	Greater look-upward visibility was associated with more ESP (β = 0.136, 95% CI [0.112, 0.159], q < 0.001), less HAP (β = −0.163, 95% CI [−0.170, −0.156], q < 0.001), and less NFAP (β = −0.085, 95% CI [−0.105, −0.066], q < 0.001). Greater eye-level visibility showed its clearest association with less ROAP (β = −0.129, 95% CI [−0.159, −0.098], q < 0.001).
Prediction 3: Selected scenes that are less fully described by spatial models will show recurrent attention to distinctive or meaning-laden focal elements.	RF models showed moderate predictive performance for ESP and HAP ($R^2 \approx 0.27$–0.28) and weaker performance for NFAP and ROAP ($R^2 < 0.20$). At VP5, the focal tree was mentioned by (31/53, 58.5%) participants and koi by (36/53, 67.9%); at VP19, inscriptions were mentioned by (24/53, 45.3%). These elements were not represented as separate predictors in the spatial models.

Note: β values are fixed-effect estimates from cross-classified LMMs with CLR-transformed outcomes and z-standardized predictors; q values are Benjamini–Hochberg-adjusted. RF results are interpreted as descriptive diagnostics rather than as parallel inferential evidence. Model discrepancies were assessed using the 51 valid eye-tracking datasets, whereas debrief frequencies were based on all 53 participants who completed the procedure. Participants could mention more than one focal element; therefore, the reported percentages are not mutually exclusive.

The study also has methodological value in showing how viewpoint-matched three-dimensional spatial features and immersive eye-tracking can be brought into closer correspondence at the level of individual scenes. Compared with studies that rely mainly on photographs, aggregate ratings, or coarse scene proxies, this approach allows scene-specific environmental description and recurrent visual behavior to be examined within the same sequence of views (Adhanom et al., 2023; Wróżyński et al., 2024). The three-dimensional component is particularly relevant because it permits upward, eye-level, downward, and overhead visibility fields to be represented separately rather than collapsed into a predominantly horizontal or image-based scene description. The methodological value here lies in demonstrating a workable scene-level approach within one heritage setting rather than in proposing a universally applicable framework.

The practical implications should likewise remain modest. The results suggest that viewpoint-level features such as openness, complexity, layering, and the presence of memorable scene features are relevant when considering how visitors visually engage with heritage scenes. This interpretation is consistent with landscape perception research linking openness, depth, complexity, and informational richness to environmental perception and preference, and with the concept of imageability, through which distinctive spatial elements contribute to how environments are recognized and remembered (Lynch, 1976; Peng et al., 2025; Peng et al., 2026). This may be useful for low-impact stewardship strategies concerned with preserving sightlines, maintaining layered scene sequences, and recognizing features that repeatedly attract attention. The findings also suggest that the preservation of vertical relationships, including canopy openings, upper architectural visibility, and eye-level framing, may be relevant to the experiential character of particular viewpoints. Such an implication is also aligned with heritage-setting guidance that emphasizes significant sightlines, important views, visual relationships, and the wider setting as part of heritage significance and protection (Peng et al., 2026; Sukwai et al., 2022). At the same time, the findings do not justify prescriptive thresholds to be transferred directly to other sites. Landscape-perception studies are sensitive to scene and participant sampling, while heritage impact assessment is inherently dependent on site-specific attributes, values, and management contexts (Peng et al., 2026;Zhang et al., 2025). Any practical use of the results should therefore be understood as site-sensitive and exploratory rather than as a universal design recipe.

4.3. Limitations and future work

Several limitations should be noted. First, the study was conducted in immersive VR rather than in situ. Although immersive viewing offers advantages over static image-based tasks, it does not fully reproduce real locomotion, multisensory input, weather, soundscape, or social context, all of which may influence visual behavior during actual site visits (Lawrence et al., 2020). The present findings therefore apply most directly to matched immersive viewing rather than to fully naturalistic heritage experience.

Second, the study is based on a single heritage garden. The value of the case lies in its suitability for observing scene-by-scene variation within a sequentially experienced heritage environment, not in statistical representativeness. The findings should therefore be read as analytic and case-based. They identify a pattern of scene–behavior correspondence in this setting and suggest questions that can be examined in other heritage environments with different spatial grammars, planting structures, and visitor traditions.

Third, the participant pool was intentionally homogeneous. Fifty-three Chinese participants from non-design backgrounds completed the procedure and post-session debrief, and 51 were retained for the eye-tracking analyses after participant-level recording-quality review. All participants had previously visited classical gardens in Suzhou, but none had visited or reported prior awareness of Jichang Garden. This reduced variation associated with professional design training and site-specific memory, but general familiarity with Jiangnan classical gardens may still have influenced attention to inscriptions, rockeries, water, pavilions, or symbolic vegetation. Because familiarity was not measured using a graded scale, its contribution could not be modeled directly. Future studies should include participants with more diverse disciplinary, cultural, and experiential backgrounds.

Fourth, viewpoint order and starting orientation were fixed in order to preserve the garden's route-like sequence. While this choice supported the experiential logic of the case, it may also have introduced anchoring, learning, or fatigue effects. Future studies could combine route-preserving designs with partial randomization or counterbalancing to assess how robust the observed patterns are under different presentation structures.

Finally, **Prediction 3** was evaluated descriptively using viewpoints selected according to the model-discrepancy criterion described in Section 2.7 and the corresponding post-session debriefs. Because focal-target selection was not modeled as a separate quantitative outcome, the observations at VP5 and VP19 should be interpreted as case-specific complementary evidence rather than as a causal or independent inferential test. Future research should model focal-target selection directly across a larger number of scenes and sites.

5. Conclusions

Using immersive eye-tracking, this study identified four recurrent visual behavior patterns and three broader exploration strategies across viewpoint-matched scenes in Jichang Garden. Higher complexity was associated with more exploratory scanning and less hotspot attention, whereas greater openness was associated with less exploratory scanning. These contrasting directions indicate that complexity and openness represent distinct dimensions of scene structure. Vertically differentiated visibility also showed distinct behavioral profiles: greater upward visibility was associated with more exploratory scanning but less hotspot attention and near–far alternation, whereas greater eye-level visibility was associated with less route-oriented attention.

Post-session debriefs provided complementary context for selected scenes in which participants repeatedly attended to a tree, koi, and inscriptions that were not represented as separate spatial predictors. These observations suggest that viewpoint-level spatial features may characterize broad engagement tendencies more fully than specific focal-target selection, although the latter was not modeled independently. Overall, the study provides case-based, scene-level evidence that multidimensional spatial features, including vertical visibility, correspond to recurrent visual behavior in differentiated ways within one heritage garden.

CRediT authorship contribution statement

Yuyang Peng: Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Steffen Nijhuis:** Conceptualization, Supervision, Writing – review & editing. **Yingwen Yu:** Conceptualization, Investigation, Methodology, Writing – review & editing. **Giorgio Agugiaro:** Supervision. **Guanting Zhang:** Software.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvp.2026.103144.

References

Aalto, P., & Steinert, M. (2025). Emergence of eye-tracking in architectural research: A review of studies 1976–2021. *Architectural Science Review, 68*(3), 213–223. https://doi.org/10.1080/00038628.2024.2382796

Adhanom, I. B., MacNeilage, P., & Folmer, E. (2023). Eye tracking in virtual reality: A broad review of applications and challenges. *Virtual Reality, 27*(2), 1481–1505. https://doi.org/10.1007/s10055-022-00738-z

Alsadik, B., Gerke, M., & Vosselman, G. (2014). Visibility analysis of point cloud in close range photogrammetry. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., II-5*, 9–16. https://doi.org/10.5194/isprsannals-II-5-9-2014

Berto, R., Massaccesi, S., & Pasini, M. (2008). Do eye movements measured across high and low fascination photographs differ? Addressing Kaplan's fascination hypothesis. *Journal of Environmental Psychology, 28*(2), 185–191. https://doi.org/10.1016/j.jenvp.2007.11.004

Bollen, Z., Wyss, A. M., Guizar Rosales, E., Le Houcq Corbi, Z., & Knoch, D. (2025). The role of attitudes and attention in pro-environmental decision-making: An eye-tracking study. *Journal of Environmental Psychology, 107*, Article 102776, 10.1016/j.jenvp.2025.102776.

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods, 15*(4), 233–234. https://doi.org/10.1038/nmeth.4642

Chen, B., Gong, C., & Li, S. (2022). Looking at buildings or trees? Association of human nature relatedness with eye movements in outdoor space. *Journal of Environmental Psychology, 80*, Article 101756, 10.1016/j.jenvp.2022.101756.

Dewhurst, R., Foulsham, T., Jarodzka, H., Johansson, R., Holmqvist, K., & Nyström, M. (2018). How task demands influence scanpath similarity in a sequential number-search task. *Vision Research, 149*, 9–23. https://doi.org/10.1016/j.visres.2018.05.006

Di Pietro, L., Guglielmetti Mugion, R., & Renzi, M. F. (2018). Heritage and identity: Technology, values and visitor experiences. *Journal of Heritage Tourism, 13*(2), 97–103. https://doi.org/10.1080/1743873X.2017.1384478

Dupont, L., Antrop, M., & Van Eetvelde, V. (2014). Eye-tracking analysis in landscape perception research: Influence of photograph properties and landscape characteristics. *Landscape Research, 39*(4), 417–432. https://doi.org/10.1080/01426397.2013.773966

Dupont, L., Antrop, M., & Van Eetvelde, V. (2015). Does landscape related expertise influence the visual perception of landscape photographs? Implications for participatory landscape planning and management. *Landscape and Urban Planning, 141*, 68–77. https://doi.org/10.1016/j.landurbplan.2015.05.003

Evans, D., & Chamberlain, B. (2024). In pursuit of eye tracking for visual landscape assessments. *Land, 13*(8), 1184. https://www.mdpi.com/2073-445X/13/8/1184.

Fife, D. A., & D'Onofrio, J. (2023). Common, uncommon, and novel applications of random forest in psychological research. *Behavior Research Methods, 55*(5), 2447–2466. https://doi.org/10.3758/s13428-022-01901-9

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*(2), 6. https://doi.org/10.1167/8.2.6, 6.

Frost, S., Kannis-Dymand, L., Schaffer, V., Millear, P., Allen, A., Stallman, H., Mason, J., Wood, A., & Atkinson-Nolte, J. (2022). Virtual immersion in nature and psychological well-being: A systematic literature review. *Journal of Environmental Psychology, 80*, Article 101765. https://doi.org/10.1016/j.jenvp.2022.101765

Gentin, S. (2011). Outdoor recreation and ethnicity in Europe—A review. *Urban Forestry and Urban Greening, 10*(3), 153–161. https://doi.org/10.1016/j.ufug.2011.05.002

Ginzarly, M., Pereira Roders, A., & Teller, J. (2019). Mapping historic urban landscape values through social media. *Journal of Cultural Heritage, 36*, 1–11. https://doi.org/10.1016/j.culher.2018.10.002

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

Henderson, J. M., & 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. https://doi.org/10.1167/18.6.10, 10.

Herzog, T. R. (1992). A cognitive analysis of preference for urban spaces. *Journal of Environmental Psychology, 12*(3), 237–248. https://doi.org/10.1016/S0272-4944(05)80138-0

Herzog, T. R., & Bryce, A. G. (2007). Mystery and preference in within-forest settings. *Environment and Behavior, 39*(6), 779–796. https://doi.org/10.1177/0013916506298796

Hwang, A. D., Wang, H. C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research, 51*(10), 1192–1205. https://doi.org/10.1016/j.visres.2011.03.010

Hyrskykari, A., Ovaska, S., Majaranta, P., Räihä, K.-J., & Lehtinen, M. (2008). Gaze path stimulation in retrospective think-aloud. *Journal of Eye Movement Research, 2*(4), 1–18. https://www.mdpi.com/1995-8692/2/4/24.

Inglis, N. C., Vukomanovic, J., Costanza, J., & Singh, K. K. (2022). From viewsheds to viewscapes: Trends in landscape visibility and visual quality research. *Landscape and Urban Planning, 224*, Article 104424, 10.1016/j.landurbplan.2022.104424.

Jeong, D., Jeong, M., Yang, U., & Han, K. (2022). Eyes on me: Investigating the role and influence of eye-tracking data on user modeling in virtual reality. *PLoS One, 17*(12), Article e0278970, 10.1371/journal.pone.0278970.

Kaplan, S. (1987). Aesthetics, affect, and cognition:environmental preference from an evolutionary perspective. *Environment and Behavior, 19*(1), 3–32. https://doi.org/10.1177/0013916587191001

Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology, 15*(3), 169–182. https://doi.org/10.1016/0272-4944(95)90001-2

Kaplan, R., & Kaplan, S. (1989). *The experience of nature: A psychological perspective*. Cambridge university press.

Ke, L., Zhang, Z., Ma, Y., Xiao, Y., Wu, S., Wang, X., Liu, X., & He, J. (2023). Evaluating flight performance and eye movement patterns using virtual reality flight simulator. *Journal of Visualized Experiments, 195*, Article e65170. https://doi.org/10.3791/65170

Keswick, M. (2003). *The Chinese garden: History, art, and architecture*. Harvard University Press.

Kuper, R. (2017). Evaluations of landscape preference, complexity, and coherence for designed digital landscape models. *Landscape and Urban Planning, 157*, 407–421. https://doi.org/10.1016/j.landurbplan.2016.09.002

Lawrence, R. K., Edwards, M., Chan, G. W. C., Cox, J. A., & Goodhew, S. C. (2020). Does cultural background predict the spatial distribution of attention? *Culture and Brain, 8*(2), 137–165. https://doi.org/10.1007/s40167-019-00086-x

Liu, Q., Zhu, Z., Zeng, X., Zhuo, Z., Ye, B., Fang, L., Huang, Q., & Lai, P. (2021). The impact of landscape complexity on preference ratings and eye fixation of various urban green space settings. *Urban Forestry and Urban Greening, 66*, Article 127411. https://doi.org/10.1016/j.ufug.2021.127411

Liu, M., & Nijhuis, S. (2020). Mapping landscape spaces: Methods for understanding spatial-visual characteristics in landscape design. *Environmental Impact Assessment Review, 82*, Article 106376, 10.1016/j.eiar.2020.106376.

Liu, M., & Nijhuis, S. (2022). Talking about landscape spaces. Towards a spatial-visual landscape design vocabulary. *The Design Journal, 25*(2), 263–281. https://doi.org/10.1080/14606925.2021.2021672

Lonergan, C., & Hedley, N. (2016). Unpacking isovists: A framework for 3D spatial visibility analysis. *Cartography and Geographic Information Science, 43*(2), 87–102. https://doi.org/10.1080/15230406.2015.1065761

Lynch, K. (1976). *Managing the sense of a Region*. Cambridge, MA: Mit Press.

Nasar, J. L. (1994). Urban design aesthetics:the evaluative qualities of building exteriors. *Environment and Behavior, 26*(3), 377–401. https://doi.org/10.1177/001391659402600305

Nijhuis, S. (2015). GIS-based landscape design research: Stourhead landscape garden as a case study. *A+ BE| Architecture and the Built Environment*, (13), 1–338. https://doi.org/10.7480/abe.2015.13

Peng, Y., Li, W., Nijhuis, S., Yu, Y., & Wu, Z. (2026). Seeing heritage through green and blue: Assessing the visual influence of blue-Green infrastructure (BGI) in historic urban areas (HUAs). *Environmental Impact Assessment Review, 118*, Article 108301. https://doi.org/10.1016/j.eiar.2025.108301

Peng, Y., Nijhuis, S., Geng, M., & Yu, Y. (2025). Enhancing visual attribute comprehension of urban heritage landscapes using combined GIS-based visual analysis methods: West Lake as a case study. *Environmental Impact Assessment Review, 115*, Article 108032, 10.1016/j.eiar.2025.108032.

Peng, Y., Nijhuis, S., Wang, Z., Yu, Y., Verbree, E., & van Oosterom, P. (2026). When to stop? A visual impact assessment framework for incremental urban and community expansion in rural heritage landscapes. *Sustainable Cities and Society, 143*, Article 107364, 10.1016/j.scs.2026.107364.

Peng, Y., Nijhuis, S., Wu, Z., & Yu, Y. (2026a). Beyond scenic views: An integrative review of visual research on heritage landscapes—themes, methods, and typologies across scales. *Landscape and Urban Planning, 276*, Article 105742, 10.1016/j.landurbplan.2026.105742.

Peng, Y., Nijhuis, S., Wu, Z., & Yu, Y. (2026b). From comparison to combination: Street view imagery (SVI) and 3D model-based analyses for urban visual environment assessment. *Landscape Architecture and Sustainability, 3*(1), Article 100034, 10.1016/j.las.2026.100034.

Peng, Y., Zhang, G., Nijhuis, S., Agugiaro, G., & Stoter, J. E. (2024). Towards a framework for point-cloud-based visual analysis of historic gardens: Jichang Garden as a case study. *Urban Forestry and Urban Greening, 91*, Article 128159, 10.1016/j.ufug.2023.128159.

Pfaller, M., Kroczek, L. O. H., Lange, B., Fülöp, R., Müller, M., & Mühlberger, A. (2021). Social Presence as a Moderator of the Effect of Agent Behavior on Emotional Experience in Social Interactions in Virtual Reality [Original Research]. *Frontiers in Virtual Reality, 2*. https://doi.org/10.3389/frvir.2021.741138, 2021.

Poria, Y., Biran, A., & Reichel, A. (2009). Visitors' preferences for interpretation at heritage sites. *Journal of Travel Research, 48*(1), 92–105. https://doi.org/10.1177/0047287508328657

Qiu, Y., Pan, H., Kalantari, Z., Giusti, M., & Che, S. (2023). The natural focus: Combining deep learning and eye-tracking to understand public perceptions of urban ecosystem aesthetics. *Ecological Indicators, 156*, Article 111181, 10.1016/j.ecolind.2023.111181.

Shu, X., Hao, P., Wang, Y., Chen, X., Shen, X., Kuai, H., & Dong, L. (2018). Influences of family culture and garden life on landscape development in classical gardens on yangtze Delta-a case study of jichang garden, wuxi, China. In *XXX International Horticultural Congress IHC2018: VII conference on landscape and urban horticulture, IV conference on 1279*.

Smith, L. (2006). *Uses of heritage*. Routledge.

Stamps, A. E. (2003). Advances in visual diversity and entropy. *Environment and Planning B: Planning and Design, 30*(3), 449–463. https://doi.org/10.1068/b12986

Stamps, A. E. (2004). Mystery, complexity, legibility and coherence: A meta-analysis. *Journal of Environmental Psychology, 24*(1), 1–16. https://doi.org/10.1016/S0272-4944(03)00023-9

Stamps, A. E. (2005). Enclosure and safety in urbanscapes. *Environment and Behavior, 37*(1), 102–133. https://doi.org/10.1177/0013916504266806

Sukwai, J., Mishima, N., & Srinurak, N. (2022). Identifying visual sensitive areas: An evaluation of view corridors to support nature-culture heritage conservation in Chiang Mai historic city. *Built Heritage, 6*(1), 23, 10.1186/s43238-022-00071-z.

Taylor, K., & Lennon, J. (2011). Cultural landscapes: A bridge between culture and nature? *International Journal of Heritage Studies, 17*(6), 537–554. https://doi.org/10.1080/13527258.2011.618246

Tilley, C. (2006). Introduction:Identity, place, landscape and Heritage. *Journal of Material Culture, 11*(1–2), 7–32. https://doi.org/10.1177/1359183506062990

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*(4), 766.

Uzzell, D. L. (1996). Creating place identity through heritage interpretation. *International Journal of Heritage Studies, 1*(4), 219–228. https://doi.org/10.1080/13527259608722151

Williams, C. C., & Castelhano, M. S. (2019). The changing landscape: High-level influences on eye movement guidance in scenes. *Vision, 3*(3). https://doi.org/10.3390/vision3030033

Wróżyński, R., Pyszny, K., & Wróżyńska, M. (2024). Reaching beyond GIS for comprehensive 3D visibility analysis. *Landscape and Urban Planning, 247*, Article 105074, 10.1016/j.landurbplan.2024.105074.

Yang, H. (2022). On the design of garden. In *A treatise on the garden of jiangnan: A study on the art of Chinese classical garden* (pp. 59–550). Springer. https://doi.org/10.1007/978-981-16-6924-8.

Yuan, S., Browning, M. H. E. M., Li, F., Duchowski, A., Hallo, J., & Nicolette, M. (2025). Can visual attention help uncover the effects of visitor density on urban park experience? An eye-tracking and virtual reality study. *Journal of Environmental Psychology, 106*, Article 102631, 10.1016/j.jenvp.2025.102631.

Zhang, C., Lin, G., Peng, Y., & Yu, Y. (2025). A workflow for urban Heritage digitization: From UAV photogrammetry to immersive VR interaction with multi-layer evaluation. *Drones, 9*(10), 716. https://www.mdpi.com/2504-446X/9/10/716.

Zhang, G., & Peng, Y. (2024). Method for quantitative analysis of visual space of chinese traditional gardens based on LiDAR point cloud: A case study of jichang garden. *Landscape Architecture, 31*(7), 108–114.

Zhao, Y., Wu, B., Wu, J., Shu, S., Liang, H., Liu, M., Badenko, V., Fedotov, A., Yao, S., & Yu, B. (2020). Mapping 3D visibility in an urban street environment from mobile LiDAR point clouds. *GIScience and Remote Sensing, 57*(6), 797–812, 10.1080/15481603.2020.1804248.

Zou, Y., Yang, Y., Li, Y., Liao, J., & Xiao, H. (2023). How do tourists' heritage spatial perceptions affect place identity? A case study of quanzhou, China. *Journal of Hospitality and Tourism Management, 55*, 460–470. https://doi.org/10.1016/j.jhtm.2023.05.018