Understanding Building Layouts from Minimal 360° Views: Effects of Viewpoint Overlap, Sequencing, and Landmarks

Ahmed Aly
Institute for Geoinformatics, Universität Münster, Germany

Jakub Krukar
Institute for Geoinformatics, Universität Münster, Germany

Abstract

While 360° panoramic virtual tours are widely used for remote spatial exploration, there is no standardized method for determining the placement or sequencing of such viewpoints. Computational solutions like the Art Gallery Problem can determine minimal viewpoints that ensure the building's coverage, however cognitive research has not addressed whether these are sufficient to support the viewer's understanding of the space. In other words, would the hypothetical guards understand their gallery's layout? Drawing on the Network of Reference Frames theory and prior research on spatial integration mechanisms and associated integration costs, this study investigates how three design factors – visual overlap, route-based sequencing, and landmark availability – influence layout understanding from a minimal set of views. A reproducible web-based experimental framework was developed to enable systematic manipulation of viewpoint configurations and detailed behavioral logging. Fifty participants explored five systematically varied configurations across five building layouts and were tasked with recognizing the correct layout among visually similar alternatives, reporting confidence after each decision. Monetary incentives were used to encourage attentive exploration and constrain strategy use. Results revealed that while individual spatial ability is a strong predictor of layout understanding ($p < .001$), viewpoint design significantly modulates that understanding. Specifically, visual overlap between consecutive views significantly enhanced layout understanding only when paired with additional cues – namely landmarks or route-based sequencing. Among these, landmarks significantly improved layout recognition ($p < .05$), unlike route-based sequencing. Moreover, combining both cues yielded no additional benefit. These findings demonstrate how viewpoint configurations can be designed to better support layout understanding from a minimal set of views. In addition, the proposed replicable, web-based experimental pipeline enables systematic evaluation and optimization of viewpoint designs used in spatial learning and navigation tools.

2012 ACM Subject Classification Human-centered computing → Empirical studies in HCI; Human-centered computing → Virtual reality; Computing methodologies → Spatial and physical reasoning

Keywords and phrases Virtual tours, Art Gallery Problem, museum guard problem, isovists, layout understanding, 360° panoramas

Digital Object Identifier 10.4230/LIPIcs.COSIT.2026.9

Supplementary Material *Dataset (Raw data)*: `https://osf.io/8e7pf/`
Software (Experiment code): `https://github.com/ahmedsaly/LayoutRecStudy`
archived at `swh:1:dir:855b206332b133e5573f2a51680b197ee7850902`

Funding This project was supported through the JICE Seed Funding by the Joint Institute of Individualisation in a Changing Environment.

17th International Conference on Spatial Information Theory (COSIT 2026).
Editors: Sabine Timpf, Gabriele Filomena, Armand Kapaj, Rui Zhu, Nicholas Giudice, and Ed Manley; Article No. 9; pp. 9:1–9:23
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Virtual representations of architectural spaces have been increasingly used across multiple domains, including real estate, architecture, and interior design [43]. A particularly widespread approach is the use of *virtual tours* – interactive systems composed of sequential images, often in 360° panoramic format, that enable users to explore built environments remotely. The popularity of these systems, supported by commercial software such as Matterport or 3DVista, lies in their utility: users can form an impression of spatial structure and layout without physically entering the space.

In research contexts, similar image-based paradigms are used to study spatial cognition, navigation, and environmental understanding. For example, researchers simulate building exploration by presenting participants with a set of images [46, 24]. However, despite the use of such paradigms, there remains no standardized method for determining *where* to place viewpoint images, *how many* are required, or *what sequence* best supports spatial understanding. This highlights a central research gap: *How can viewpoint systems be systematically selected and organized to support spatial layout understanding?*

Addressing this question has both theoretical and practical relevance. From a theoretical perspective, it contributes to modeling how humans acquire survey knowledge from fragmented visual input – concepts central to theories describing how we integrate information across reference frames (see Section 2.1). Practically, it can inform the design of spatial learning tools, legible building layouts, indoor navigation systems, and architectural visualizations based on panoramic imagery. The primary contribution of this work is an empirical investigation of which viewpoint-system design factors support layout understanding from a minimal set of views.

An additional methodological contribution is a reproducible, web-based experimental framework for panoramic layout-understanding studies. It enables systematic manipulation of viewpoint configurations while capturing detailed behavioral responses of participants, and incorporates a monetary incentive structure to encourage attentive exploration. Using this framework, viewpoint positions were optimized using the Art Gallery Problem (AGP) algorithm to calculate the minimum number of views required for full geometric coverage (see Section 2.3). Across these optimized positions, the degree of overlap, the sequencing strategy, and the availability of landmarks were systematically added to test their impact.

Our findings reveal that layout understanding was strongly predicted by individual spatial ability and was significantly improved when landmarks were available within overlapping views. Furthermore, visual overlap alone was found to be insufficient without additional cues, and the monetary incentive structure also influenced exploration strategies.

2 Background

2.1 Reference frames and spatial integration

To understand how individuals integrate static panoramic images into a coherent mental model, we draw on the *Network of Reference Frames* theory [33]. It proposes that spatial knowledge is not stored in a single global coordinate system, but rather as a network of locally encoded *vista spaces* – areas visible from a single viewpoint [33, 37]. Each vista space is stored in its own reference frame, and transitions between vista spaces are represented by *perspective shifts* involving translation and rotation required to move from one reference frame to another. Comprehension of an environment, therefore, depends on the successful formation of these shifts [33].

Forming survey knowledge from panoramic imagery thus fundamentally relies on *spatial integration* – the mental unification of separately experienced vista spaces into a coherent environmental representation [33, 16, 22]. This integration relies on working-memory demands during mental transformation and alignment, introducing what is known as *integration cost* [34]. Integration costs are reflected in reduced spatial accuracy, increased task completion time, or higher error rates during recall [34]. These costs depend on how easily viewpoints can be aligned and linked into a consistent representation, and can be reduced by cues that support perspective-shift formation and reference-frame alignment.

For adjacent vista spaces, *visual overlap* provides a primary mechanism forming perspective shifts: when two vista spaces share a direct visual connection, the shared geometry provides the necessary translational and rotational components to link the frames [33]. As integration costs diminish when alignment occurs *during encoding* rather than being postponed to retrieval [34], integration during encoding is operationalized through an *overlap-based sequence*: presenting overlapping views consecutively so that each new panorama shares visual information with the previous one. In our exploratory design, this overlap-based sequence serves as the baseline for testing whether additional cues further facilitate integration and improve layout understanding.

2.2 Cues beyond overlap: landmarks and route-based sequencing

Beyond overlap, additional cues can support integration across views and help build a coherent layout representation. One such cue is *global landmarks*: distinctive objects visible from multiple locations can act as common anchors. Self-localization relative to a landmark allows the observer to reorient across different vista spaces, effectively providing the perspective shift between reference frames [33]. Landmarks within overlapping regions provide stable reference points that help align egocentric views to a shared allocentric structure, improving memory precision [1, 15, 25] and thus reducing integration costs [34]. In the context of panoramic viewpoint systems, this motivates placing landmarks so they remain visible across consecutive, overlapping views, providing persistent anchors across view transitions.

Another cue is continuous updating during movement. When moving between non-adjacent vista spaces, individuals can track their position and orientation relative to the last encoded vista via path integration, accumulating translational and rotational information to achieve the necessary perspective shift [32, 33]. In panoramic viewpoint systems, transitions between views are inherently discontinuous and therefore resemble teleportation-like navigation. Consistent with VR locomotion research, comparisons of teleportation with continuous movement indicate that discontinuous transitions can impair spatial updating and orientation and can bias spatial judgments such as distance perception [6, 29]. Accordingly, many studies focus on improving how teleportation is executed – for example, adding orientation-preserving mechanisms, rotational self-motion cues, or visual transition/flow cues to reduce disorientation and better support spatial updating [31, 42, 3].

However, these advances mainly refine transition mechanics rather than the placement of teleport targets. Recent work on scene-aware teleport graphs proposes structured sets of teleport nodes derived from environmental geometry [28], but how to select a minimal set of viewpoints that optimally supports users' understanding of the global layout remains largely unexplored. Our focus here is not on rendering smooth transitions between viewpoints, but on how a minimal set of discrete viewpoints is selected and ordered to support layout understanding. At the same time, continuous visual flow between viewpoints, even in the absence of active bodily movement, appears to be a key driver of spatial learning: smooth transitions between views enhance small-scale spatial learning, with passive viewing of these transitions performing as well as active movement [20]. A straightforward way to approximate some of this continuity in panoramic viewpoint systems is to structure the sequence of views so that it follows a plausible route through the environment. In this study, a *route-based sequence* was therefore tested: the same set of overlapping views was ordered to mirror how a person might naturally explore the building.

Finally, integration efficiency is subject to individual differences. The ability to form and use perspective shifts varies by individual spatial ability [18]. Participants with stronger spatial abilities, such as higher mental rotation skills or a better sense of direction, demonstrate quicker and more accurate integration [18]. Therefore, spatial ability is measured and accounted for when analyzing differences in layout understanding between participants.

2.3 Methodological foundations: minimal viewpoints and incentives

One methodological concern is limiting the number of viewpoints while still supporting reliable layout understanding. The cognitive demands of spatial integration increase with the number of distinct vista spaces that must be linked into a coherent representation [33]. For example, pointing errors increase with the number of vista spaces traversed [49]. This motivates a key design goal for panoramic viewpoint systems: fewer, well-selected viewpoints should reduce cognitive load and support more efficient layout understanding. The challenge, however, is how to identify the smallest set of viewpoints that is sufficient for understanding the space. In the absence of a systematic method for determining what people need to see to understand a space, our starting assumption is to require *full visual coverage* of the space. This does not assume that full geometric coverage equals cognitive understanding; rather, it serves as a principled baseline for testing which further design factors support understanding. Under this condition, the question becomes: how can we identify the smallest set of viewpoints that collectively provides full coverage? This is a computational optimization problem, and it naturally maps to the Art Gallery Problem (AGP) [8, 39].

The AGP addresses the question of how many “guards” with 360° visibility are needed to observe every point in a polygonal environment of an art gallery [8, 39]. While classical theory gives general bounds on how many guards are needed, computing an *optimal* solution is NP-hard, especially when guards are allowed to be placed anywhere inside the polygon (point guards) [39]. A practical and widely used approach is the guard–witness formulation, which reduces coverage to a finite set of candidate guards and witness points and solves the resulting set-cover instance via Integer Linear Programming (ILP), iteratively refining the witness set until full coverage is achieved. Algorithm 966 [50] provides a robust implementation of this strategy and has been shown to solve thousands of benchmark instances efficiently on standard hardware, achieving optimal solutions in over 95% of more than 2,700 tested cases (including polygons with up to 2,500 vertices) [50]. In this study, this method is used to compute the minimum set of viewpoints required for full spatial coverage.

A second methodological concern is maintaining participant attention under potentially high cognitive demands. Because spatial integration is effortful and time-consuming, participants may adopt different viewing strategies depending on task framing. Monetary incentives can increase motivation and attentional effort, especially in tasks that require the deliberate allocation of cognitive resources [4, 21]. Therefore, in this study, a monetary incentive structure is used to encourage attentive exploration and integration across the available views rather than superficial task completion.

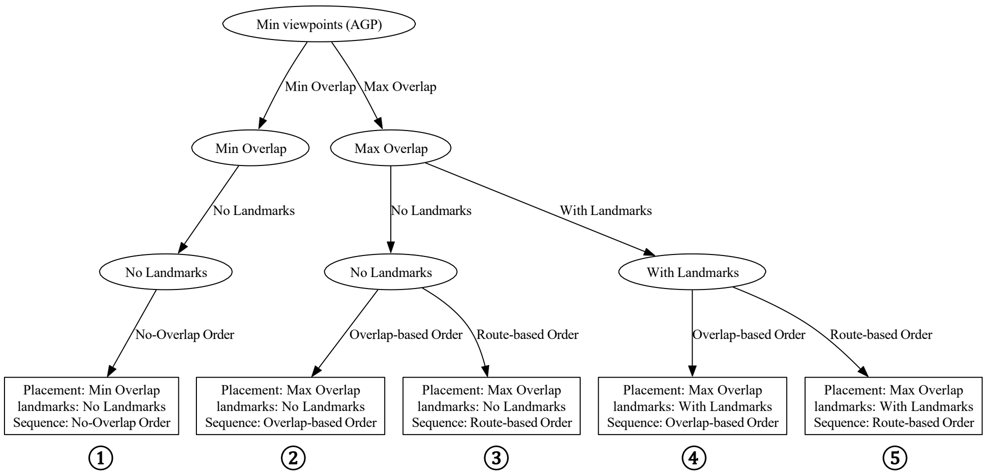


Figure 1 Overview of the five experimental conditions, varying overlap, sequencing strategy, and landmark presence.

3 Methods

3.1 Research question and experimental conditions

Building on the integration mechanisms reviewed in Section 2, to examine how viewpoint-system design influences *layout understanding* when indoor environments are explored from a minimal set of 360° panoramas, the following research question is posed:

(RQ). *How do viewpoint overlap, sequencing strategy, and landmark cues affect layout understanding from a minimal set of panoramic views?*

Here, *layout understanding* refers to the ability to form a mental representation of a building's overall structure (the number and arrangement of rooms or spaces, and their relative positions), sufficient to recognize a layout among visually similar alternatives.

To address this question, the following hypotheses are proposed:

- **H1 (Overlap).** Visual overlap between consecutive views is essential for layout understanding; lower layout understanding is expected when consecutive views provide no overlap compared to when overlap is high.
- **H2 (Cues beyond overlap).** When overlap is available, additional cues further improve layout understanding:
  - **H2a (Route-based order).** A route-based sequence is expected to yield higher layout understanding than an overlap-based sequence.
  - **H2b (Landmarks).** Landmark availability within overlap regions is expected to yield higher layout understanding than an overlap-based sequence without landmarks.
  - **H2c (Combined cues).** Combining route-based order and landmarks is expected to yield the strongest layout understanding.
- **H3 (Individual differences).** Higher individual spatial ability predicts better layout understanding.

To test these hypotheses, five experimental conditions were created and implemented across five building layouts. The conditions differed along three factors: viewpoint placement, viewing sequence, and the presence of landmarks. Figure 1 summarizes the construction

Table 1 Overlap statistics for the min- and max-overlap placements. *Total overlap* sums all pairwise isovist overlaps as a percentage of building area; regions seen by several viewpoints are counted once per pair, so the value can exceed 100%. *Overlapping viewpoints* is the number of viewpoints sharing overlap with at least one other, out of the total.

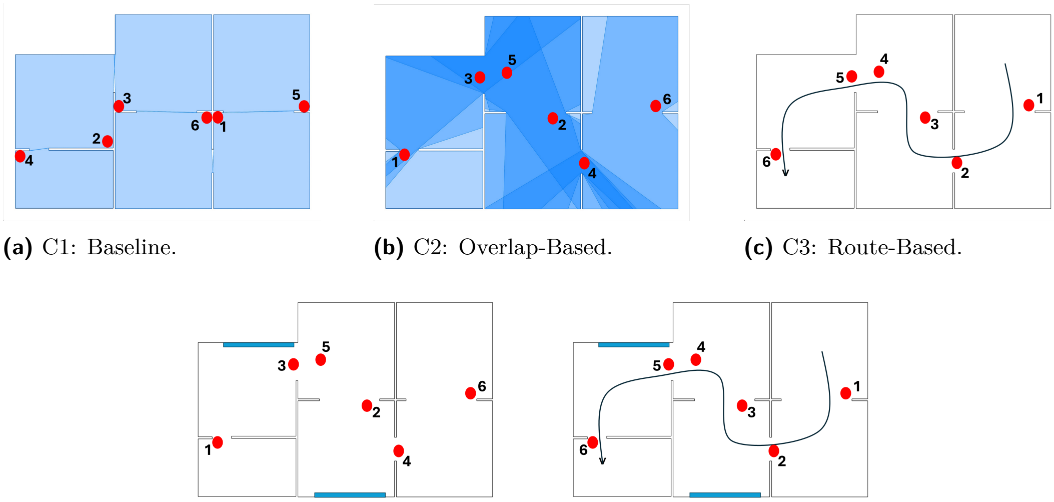
	Total overlap		**Overlapping viewpoints**	
Building	**Min**	**Max**	**Min**	**Max**
B1	3.18%	182.78%	2/4	4/4
B2	1.43%	285.93%	4/6	6/6
B3	3.45%	209.90%	2/5	5/5
B4	29.11%	92.85%	3/4	4/4
B5	0.70%	60.23%	2/6	6/6

pipeline and the resulting five conditions. The *minimal viewpoint number* required for full visual coverage was computed via an AGP solution (Section 2.3) and was held constant across conditions (within each building layout). These factors were then manipulated under the same full-coverage constraint:

- **Overlap level (viewpoint placement).** A *min-overlap* placement was constructed in which visual overlap of area seen between any pair of viewpoints was minimized as far as the layout permitted. A *max-overlap* placement was constructed in which overlap was maximized, aiming to enlarge the regions that can be jointly seen from multiple viewpoints and to increase the number of viewpoints that share overlap. Because strict "no overlap" between all viewpoint pairs was not always achievable, *min overlap* denotes the lowest feasible overall overlap, whereas *max overlap* denotes a placement in which overlap was made as extensive as feasible. The contrast between placements is summarized in Table 1 and visualized via isovists in Figure 2(a–b).
- **Sequencing strategy (transitions).** In an *overlap-based sequence*, views were ordered so that consecutive views overlapped. In a *route-based sequence*, consecutive overlap was preserved while the order was adjusted to approximate a plausible walking path through the environment. The walking path was designed manually, covering all rooms in the layout, visiting each only once, and staying in the center of the spaces (following findings that natural movement is biased towards *lowest isovist drift* [9].)
- **Landmark availability.** Distinctive landmarks were added and were placed to remain visible across consecutive overlapping views (i.e., within overlap regions).

Experimental conditions. Figure 2 illustrates the manipulations for one example layout. Each condition was implemented on each building layout as follows:

- **C1: Min overlap + no-overlap order + no landmarks (baseline).** A min-overlap placement is used, and views are ordered so that successive views do not overlap and that the sequence does not approximate a route.
- **C2: Max overlap + overlap-based order + no landmarks.** A max-overlap placement is used, and views are ordered so that each view overlaps with the next.
- **C3: Max overlap + route-based order + no landmarks.** The same placement as C2 is used, but views are ordered to approximate a route-like path while preserving consecutive overlap.
- **C4: Max overlap + overlap-based order + landmarks.** The same placement and order as C2 are used, but landmarks are added to be visible across overlap regions.
- **C5: Max overlap + route-based order + landmarks.** The same placement as C2 is used, with route-like ordering and landmarks, while maintaining consecutive overlap.



(a) C1: Baseline. **(b)** C2: Overlap-Based. **(c)** C3: Route-Based.

(d) C4: Overlap + Landmarks. **(e)** C5: All Cues Combined.

Figure 2 Five experimental conditions for Building B2. Red dots show viewpoints (numbers = order); blue bars = landmarks. In (a)–(b), semi-transparent isovists (not shown to participants) quantify overlap: (a) min-overlap placement, (b) max-overlap. C2–C5 use max-overlap placements; C1 uses min-overlap.

These conditions allow systematic testing of the proposed hypotheses: H1 is tested by comparing C1 against the overlap-rich conditions (C2 to C5); lower layout understanding in C1 would confirm the essential role of overlap in layout understanding. H2a is tested by comparing C2 vs. C3, where an improvement in C3 would indicate that route-based sequencing provides additional benefits beyond overlap-based sequence alone. H2b is tested by comparing C2 vs. C4, where an improvement in C4 would suggest that landmarks enhance layout understanding beyond overlap alone. H2c is tested by comparing C5 against C3 and C4, where an improvement in C5 would indicate additive or synergistic benefits from combining route-based sequencing and landmark cues. Spatial ability (H3) is included as an individual-differences predictor in the analysis.

3.2 Participants

Participants were recruited via Prolific and were required to be located in Germany or the Netherlands, fluent in English, and to have a Prolific approval rate between 95% and 100%; the study was restricted to desktop devices. Fifty-four participants were recruited, and $n = 50$ were retained after predefined attention and comprehension checks (34 male, 16 female; age 18–55, $M = 30.5$, $SD = 9.0$). The full set of exclusion criteria is provided in Appendix A.

3.3 Experimental platform

A browser-based framework was developed to support controlled presentation of panoramic views, enforcement of task constraints, and fine-grained behavioral logging in an online study. The framework was developed using `jsPsych` for experimental logic and flow [10] and

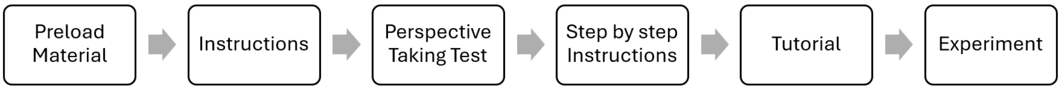


Figure 3 Overview of the six-phase experimental flow implemented in the study.

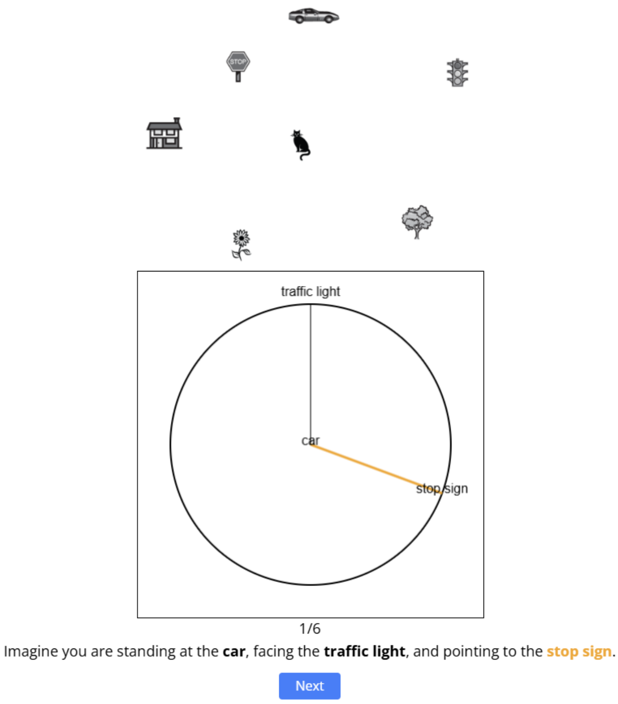


Figure 4 Perspective-Taking Test (PTT) trial. Participants imagined standing at one object (e.g., car), facing a second (e.g., traffic light), and indicated the direction to a third (e.g., stop sign).

`Panolens.js`[1] for rendering panoramic images. Interaction data were logged in real time using `DataPipe` [11], including view transitions, timestamps, layout selections, confidence ratings, and viewport behavior (active exploration vs. static viewing, explored view segments, and segment-level viewing durations). Data were stored on the Open Science Framework (OSF). The front end was deployed via GitHub Pages, and participants were recruited through Prolific.

3.4 Procedure and experimental flow

The study followed a six-phase procedure (Figure 3). First, all stimuli and interface assets were preloaded on participants' browser to avoid latency-related differences in loading and interaction timing. Second, general instructions were presented for both the spatial-ability assessment and the layout-recognition task. Third, spatial ability was assessed using a computerized Perspective-Taking Test (PTT; Figure 4). The PTT consisted of six trials in which participants were asked to imagine standing at one object, facing a second, and indicating the direction to a third object by clicking on a circular response display. Angular

[1] `https://github.com/pchen66/panolens.js`

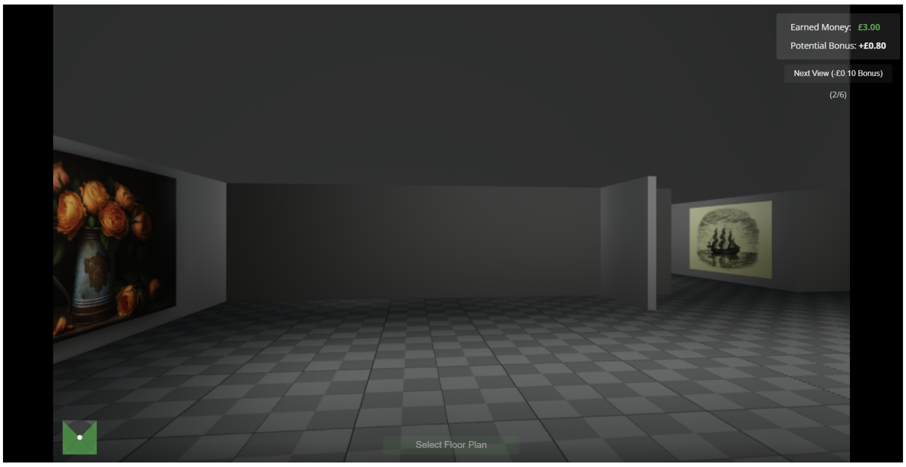


Figure 5 Main experimental interface displaying the 360° panorama, progress indicator (bottom-left), bonus status (top-right), and layout selection button (bottom-center).

error was computed per trial and aggregated into a participant-level score, which was normalized as a z-score for between-participant comparisons (lower values indicating better spatial ability). The PTT was selected as it measures rotational perspective-taking, which is critical for integrating spatial information across multiple viewpoints. The original Spatial Orientation Test (SOT) was introduced in prior work [26, 17], and the computerized version used here followed a later adaptation [12].

Fourth, step-by-step task instructions were provided for the panoramic interface, exploration mechanics, reward logic, and layout selection process. Fifth, a practice tutorial trial was completed to familiarize participants with the full sequence of exploration, decision-making, and confidence reporting. The tutorial used a simplified two-view layout with two alternative floor plans to choose between, where the two alternatives differed in a single room. Sixth, the main task was completed across five trials, each involving one building under one experimental condition. On each trial, 360° panoramic views were explored (Figure 5) before a layout was selected from a set of alternatives, after which confidence was rated on a five-point scale (Figure 6) and trial feedback (correct/incorrect, correct answer, and bonus balance) was presented. Layout understanding was operationalized via correct layout recognition among visually similar alternatives.

During the main task, a standardized panoramic interface was used. Each panorama was subdivided into four 90° segments, and all segments (i.e., the full scene) were required to be viewed at least briefly before participants could advance to the next panorama or select a layout. Viewing progress was displayed via an indicator. Panoramas were explored via mouse dragging or keyboard input. The interface was designed to maintain visual consistency across screen sizes and browser types.

3.5 Stimuli, layouts, and layout alternatives

Five building layouts were selected from the HouseExpo dataset [30]. As the raw layouts contained geometric noise from SLAM-based scanning artifacts, vertex simplification and wall regularization were applied using a custom script to improve architectural realism and to ensure compatibility with AGP-based viewpoint computation. Candidate layouts were screened to support all five experimental conditions described in Section 3.1. Selection was

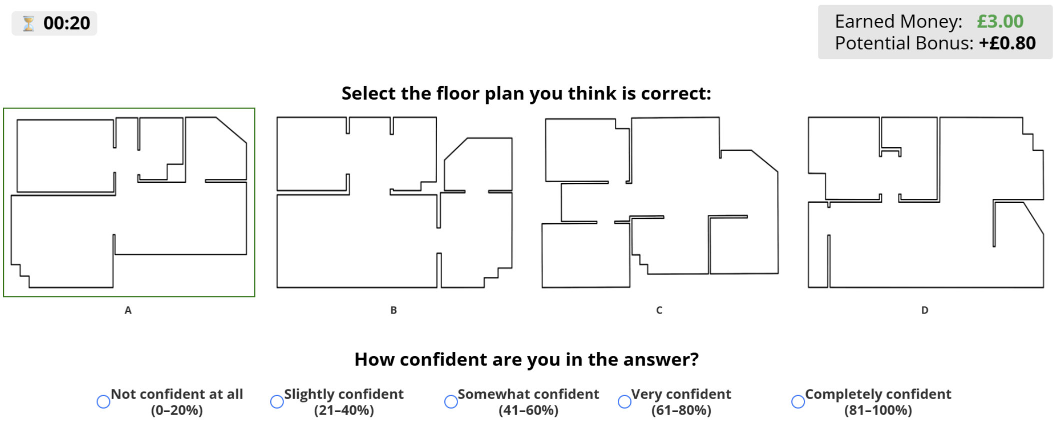


Figure 6 Layout-choice screen (example Building B4). Participants selected the correct layout among three visually similar alternatives within a 40 s decision window, then reported confidence on a five-point scale.

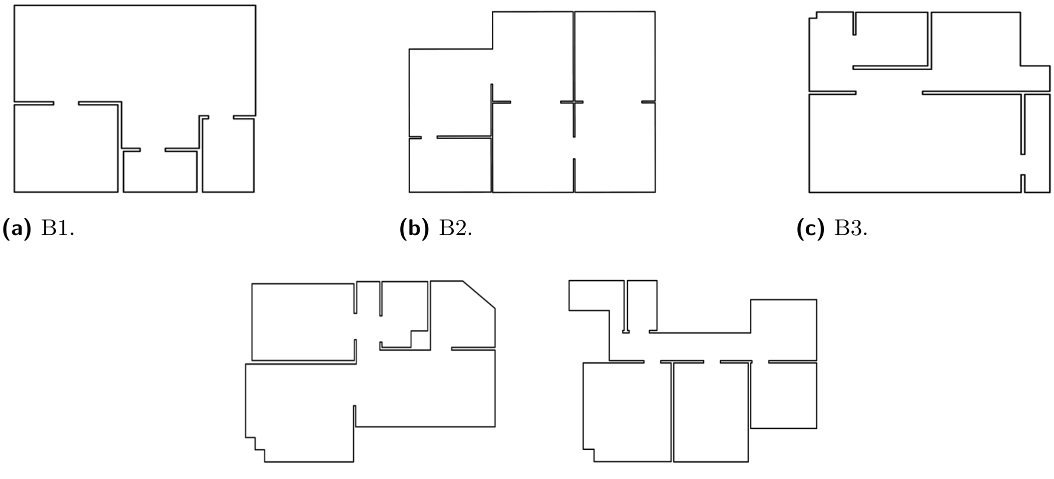
Table 2 Condition assignments across layouts for the five experimental combinations. The combinations forming a balanced Latin square in which each condition appears exactly once per building and once per combination.

Combination	Building				
	B1	**B2**	**B3**	**B4**	**B5**
1	C2	C4	C1	C3	C5
2	C5	C1	C3	C2	C4
3	C3	C5	C4	C1	C2
4	C1	C2	C5	C4	C3
5	C4	C3	C2	C5	C1

further constrained such that the AGP 966 algorithm yielded between 4 and 6 viewpoints per layout. The final five layouts were selected to be comparable in overall size yet structurally distinct. Selected rooms were reshaped manually to introduce distinctive geometric features (e.g., diagonal corners, L-shapes, or elongated rectangles), resulting in layouts that vary in the number of distinctive spaces (with some where most rooms were unique, and some with only one or two highly distinctive rooms). Layouts were also inspected to ensure that no two viewpoints within a layout were redundant or visually ambiguous. The final set of five layouts is shown in Figure 7.

To counterbalance the five buildings and five experimental conditions (C1–C5), participants were assigned to one of five combinations (Table 2). Each combination mapped one condition to each building, so that each participant experienced all five conditions, one per building. Using five distinct buildings – rather than repeating a single layout across conditions – ensured that each condition was experienced on a previously unseen building, preventing layout familiarity from carrying over between trials. Building order was randomized within participants to mitigate sequence effects.

Each layout was 3D-modeled in Rhino/Grasshopper and imported into Unity. Lighting, materials, and camera parameters were configured to approximate realistic viewing conditions. Viewpoint locations were initialized from the AGP-derived number of viewpoints and then validated and adjusted in Isovists.org app to match the five condition placements. The



(a) B1. **(b)** B2. **(c)** B3.

(d) B4. **(e)** B5.

Figure 7 Final set of five building layouts used in the experiment. Layouts were selected for structural variety and full compatibility with all five experimental conditions. Manual refinements enhanced room distinctiveness and eliminated redundant viewpoints.

resulting viewpoint configurations were imported into Unity, where an automated script applied the five experimental conditions (assigning viewpoints, inserting landmarks, ordering views, and exporting 360° panoramic images). These panoramas served as the core visual stimuli.

For each trial, one correct layout was presented to participants alongside three visually similar alternatives. The alternatives were created by editing the target layout so that they would match its overall complexity and distinctive geometric cues while differing in spatial configuration (in most cases this was achieved by rearranging room positions). Orientation was held constant across alternatives to avoid rotational confounds. An example set of alternatives is shown in Figure 6. Each participant received a base amount of £3 and could earn up to £4 through performance-based bonuses (£0.80 per layout × 5 layouts).

3.6 Pilot-informed protocol refinement

An iterative refinement process was conducted to align participant behaviour with the intended construct of *layout understanding* (integration across multiple vistas rather than single-view feature matching). Two pilot phases were run to validate the platform and identify unintended strategies. A first pilot phase with $n = 18$ volunteers validated the technical setup and gathered usability feedback across the five layout–condition combinations. A second pilot phase on Prolific (54 recruited; 42 retained after the checks in Appendix A) evaluated how interface access and incentive framing shaped exploration.

In both pilots, the layout-choice screen could be accessed after any number of viewed panoramas. A performance-based bonus scheme was used, in which a base payment of £3 was provided, and a bonus of up to £4 was offered, corresponding to five layout trials, with up to £0.80 available per layout if the correct layout was selected. The first two views were free and additional views were associated with small deductions (e.g., £0.10 per additional view). A single optional revisit was permitted after all views had been seen, with a one-time deduction of £0.10, and an incorrect choice resulted in forfeiture of the bonus for that layout. Under this structure, frequent early decisions were observed: layouts were often selected after only one or two panoramas, and decision times on the layout-choice screen were relatively large (median decision time = 31.5 s, IQR: 20.0–46.8). Interaction logs and participants' feedback indicated that decisions were often made via *feature matching*–memorizing local details and then excluding alternatives–rather than by integrating the full set of vistas into a coherent layout representation. Because the use of this strategy by some participants confounded the results, the pilot data were not used for hypothesis testing.

The final protocol was then fixed for hypothesis testing: the tutorial trial was redesigned so that the view distinguishing the two alternatives was shown first rather than second, discouraging participants from relying on the last-viewed panorama for feature matching; access to the layout-choice screen was locked until all panoramas in a trial had been viewed at least once; per-view penalties were removed in favour of a flat £0.80 bonus for a correct answer; and a 40 s decision window was imposed to limit prolonged feature-matching on the alternatives. Evidence that these changes shifted behaviour toward the intended integrative strategy is reported in Section 4.1.

4 Results

4.1 Effect of protocol refinements on exploration behavior

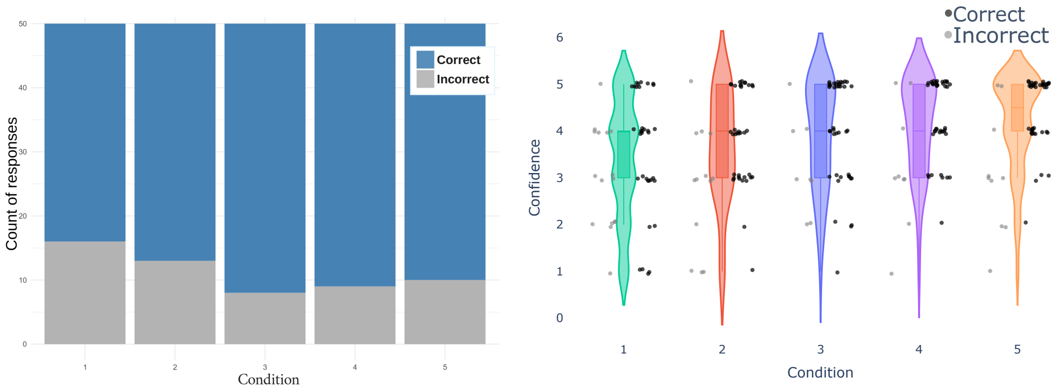
Before analysing task performance, we verified that the final protocol shifted exploration behaviour toward the intended integrative strategy and away from the single-view feature matching observed in the pilots. Median per-panorama viewing time fell from 23.9 s (IQR: 14.5–44.5) in the pilot to 13.7 s (IQR: 8.7–22.9) in the final protocol, and time spent on the final panorama fell from 18.9 s (IQR: 11.6–36.7) to 10.4 s (IQR: 7.2–17.3), consistent with reduced reliance on memorising isolated views, and on the last view in particular. Median decision time on the layout-choice screen also decreased, from 31.5 s (IQR: 20.0–46.8) to 23.5 s (IQR: 15.5–36.1) despite the 40 s cap, suggesting that participants selected the layout matching an already-formed representation rather than comparing features across alternatives. The optional revisit, available only after all views had been seen, was used 61 times across the 250 retained trials, compared with only 2 times in the pilots. Together, these patterns indicate improved compliance with the intended integrative strategy, and only data from the final protocol were retained for the reported analyses.

4.2 Descriptive overview of performance measures

We analysed three measures related to task performance: (i) correctness (objective accuracy), (ii) confidence ratings (subjective certainty), and (iii) an integrated confidence–correctness score.

Correctness. Correctness measures whether participants selected the correct layout from four options available and provides a direct indicator of task success. Figure 8a shows correct versus incorrect responses per condition (50 responses per condition; $N = 250$). Overall, 194 responses were correct. Accuracy was lowest in Condition 1 and higher in Conditions 3–5, suggesting that additional cues beyond baseline viewing may benefit recognition.

However, correctness alone does not distinguish between informed recognition and lucky guesses. Two participants may both answer correctly while differing substantially in certainty and depth of understanding, motivating the confidence ratings examined next.



(a) Correct vs. incorrect responses by condition. **(b)** Confidence ratings by condition.

Figure 8 Descriptive performance summaries across conditions (50 responses per condition).

Confidence. Confidence ratings provide a metacognitive estimate of participants' certainty in their layout judgments, potentially distinguishing informed decisions from guesses, thereby reflecting varying levels of layout understanding. Figure 8b shows the distribution of confidence ratings by condition. Consistent with the correctness pattern, Conditions 3–5 exhibit higher confidence distributions, whereas Condition 1 shows broader and lower confidence, indicating greater uncertainty.

To evaluate whether confidence meaningfully tracks performance in this task, calibration was assessed using a Confidence–Accuracy Characteristic (CAC) analysis. CAC plots, widely used to assess the diagnostic value of confidence, relate confidence levels to the proportion of correct responses at each level [23, 35, 48, 45]. As shown in Figure 9, accuracy increased with confidence overall. The small dip at confidence level 2 coincides with the smaller number of responses at that level. This confidence–accuracy alignment suggests that participants' confidence ratings tracked response accuracy overall and were therefore informative for distinguishing stronger from weaker layout understanding.

A limitation of raw confidence is that it does not reflect whether certainty is justified: high confidence after an incorrect choice corresponds to misplaced certainty and should not be interpreted as strong comprehension. To better capture meaningful understanding across conditions, a combined metric was therefore used.

Combined confidence–correctness score. To integrate subjective certainty with objective correctness, a combined score was computed:

$$\text{Combined Score} = \begin{cases} \text{Confidence rating (1–5)} & \text{if the answer is correct} \\ 0 & \text{if the answer is incorrect.} \end{cases}$$

This yields a six-level ordinal measure (0–5), where higher values reflect correct and confident recognition. The combined score aligns closely with the study goal of capturing *successful and justified* layout understanding, and was therefore used as the primary dependent variable in the main model. A correctness-only model is also reported as a robustness check.

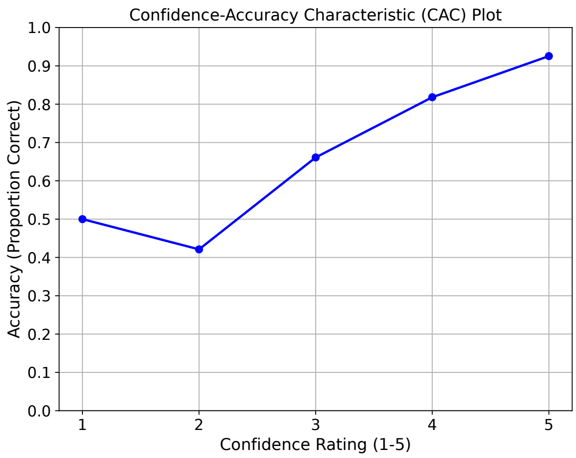


Figure 9 Confidence–accuracy characteristic (CAC) plot. Each point shows the proportion of correct responses at a given confidence level (1–5).

4.3 Modeling approach

Two complementary models were fitted in R [41] to quantify effects of viewing condition and spatial ability: (i) a mixed-effects logistic regression on correctness (robustness check), and (ii) a cumulative link mixed model on the combined confidence–correctness score suitable for ordinal data (primary analysis). Spatial ability was operationalized by the standardized PTT score (`PTT_z`).

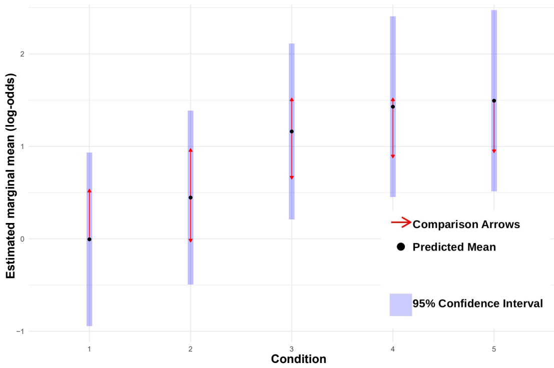
4.3.1 Robustness check: correctness model (binary DV)

Correctness was modeled as a binary outcome (0/1) using a mixed-effects logistic regression (`lme4` [2]) with fixed effects of `Condition` and `PTT_z`, and a random intercept for `Layout`:

$$\texttt{Correctness} \sim \texttt{Condition} + \texttt{PTT\_z} + (1 \mid \texttt{Layout}). \tag{1}$$

A random intercept for Participant was initially included but was removed from the final model: once spatial ability (`PTT_z`) was added as a fixed effect, the participant-level variance dropped to zero (a singular fit), indicating that individual differences in correctness were fully captured by `PTT_z`. In this model, Spatial ability significantly predicted correctness (better spatial ability corresponding to higher accuracy; $p < .001$). In contrast, condition differences were not statistically significant after adjustment (all contrasts $p > .17$). Thus, although descriptive differences were visible, correctness alone did not yield strong inferential differences between conditions. Moreover, correctness does not capture participants' internal certainty: correct answers obtained through guessing are indistinguishable from correct answers based on confident understanding. This motivates the combined-score model as the primary analysis.

A post-hoc simulation-based power analysis (500 iterations) using the `simr` package [14] indicated that achieved power was 99.8% for spatial ability but only 42.8% for the condition effect, showing that the binary outcome was underpowered to detect condition differences.



(a) Estimated marginal means by condition.

Contrast	Estimate (log-odds)	SE	*z*-ratio	*p*-value (Dunnett)
H1: C1 as reference (testing if overlap is essential)				
C1 vs C2	0.452	0.384	1.177	0.5703
C1 vs C3	1.166	0.388	3.007	**0.0099**
C1 vs C4	1.434	0.403	3.562	**0.0014**
C1 vs C5	1.499	0.406	3.696	**0.0009**
H2a,b: C2 as reference (testing route sequencing and landmarks)				
C2 vs C3	0.715	0.384	1.860	0.1965
C2 vs C4	0.983	0.397	2.473	**0.0474**
C2 vs C5	1.047	0.400	2.615	**0.0322**
H2c: C5 as reference (testing additive benefit of both cues)				
C5 vs C3	-0.333	0.395	-0.841	0.7783
C5 vs C4	-0.065	0.403	-0.160	0.9954

(b) Planned condition contrasts (log-odds).

Figure 10 Model-based effects for the combined confidence–correctness score. (a) Estimated marginal means (points) with 95% CIs (purple bands); red arrows aid pairwise comparison (less overlap between arrows = stronger differences), and higher values indicate stronger layout understanding. (b) Condition contrasts (log-odds): 0 = no difference; positive values favour the second condition in each label, negative the first; bold p-values are significant.

4.3.2 Primary model: combined score (ordinal DV)

The primary analysis modeled the combined confidence–correctness score (0–5) using a cumulative link mixed model (logit link), fitted with the `ordinal` package [7], with fixed effects of `Condition` and `PTT_z` and random intercepts for `Participant` and `Layout`:

$$\texttt{CombinedScore} \sim \texttt{Condition} + \texttt{PTT\_z} + (1 \mid \texttt{Participant}) + (1 \mid \texttt{Layout}). \qquad (2)$$

Spatial ability was a significant predictor in this model ($p < .001$), indicating that participants with better spatial ability tended to achieve higher combined scores.

For the condition effect, estimated marginal means (Figure 10a) show an increasing trend from Condition 1 toward Conditions 3–5. Post hoc pairwise comparisons, computed with `emmeans` [27] using Dunnett adjustment,[2] (Figure 10b) showed that Conditions 3, 4, and 5 significantly outperformed Condition 1, while Condition 2 did not ($p = .57$). This pattern suggests that maximal viewpoint overlap alone (Condition 2) does not substantially enhance layout understanding. Both landmarks (C4) and their combination with route-based sequencing (C5) led to significantly higher scores compared to Condition 2 ($p = .047$ and $p = .032$, respectively), whereas route-based sequencing alone (C3 vs. C2) did not yield a significant improvement ($p = .20$). Condition 5 did not differ significantly from either Condition 3 or Condition 4, indicating that combining both cues produced no additive benefit beyond the presence of a single cue.

Revisit behavior. Revisits were most frequent in Condition 1 (18 revisits, with approximately 90% involving revisiting all available views), whereas other conditions ranged between 8–12 revisits with an average revisit extent of approximately 74%. This pattern is consistent with higher uncertainty or integration demand in the baseline condition.

[2] Dunnett adjustment accounts for multiple comparisons by testing each condition against a designated reference. In this experiment, C1, C2, and C5 were each used as reference levels to test H1 and H2a–c.

A post-hoc power analysis for the combined-score model yielded 98.4% power for the condition effect. Data were simulated (500 iterations) from the fitted model parameters – fixed effects, thresholds, and random-effect variances – and each iteration was evaluated with a likelihood-ratio test, confirming that this ordinal measure was well-powered to detect condition differences.

5 Discussion

This study examined how viewpoint-system design influences *layout understanding* from a minimal set of 360° panoramas. Motivated by the Network of Reference Frames theory [33], we varied three factors that can support spatial integration across discrete vista spaces: visual overlap (to enable correspondence between consecutive views), route-based ordering (to approximate continuity), and landmarks (to provide stable anchors). Using a combined confidence–correctness score as the primary outcome (with a correctness-only robustness check), we interpret the results in relation to the hypotheses and discuss theoretical, applied, and methodological implications.

Across analyses, spatial ability was a strong predictor of performance: participants with lower `PTT_z` (better perspective-taking ability) achieved higher correctness and higher combined scores, supporting H3. This is consistent with prior work showing that spatial integration efficiency varies across individuals and depends on the ability to form and use perspective shifts [18, 33, 34]. Participants with better perspective-taking ability may therefore have been better able to align viewpoints and integrate multiple vistas into a coherent layout representation.

Condition differences were not statistically significant in the correctness-only model, but were detectable once confidence was incorporated via the combined score. This is consistent with the post-hoc power analyses reported above: on the same sample, the binary correctness outcome afforded low power for condition effects (42.8%), whereas the combined score afforded high power (98.4%), indicating that the coarser binary measure limited sensitivity to condition differences. Taken together, the combined-score results suggest the following patterns:

- **Overlap alone is insufficient.** Condition 2, which maximized visual overlap without route-based sequencing or landmarks, did not significantly outperform the baseline Condition 1. Performance improved only when overlap was combined with additional cues, as in Conditions 3, 4, and 5. This suggests that visual overlap alone does not guarantee integrative spatial understanding: it appears to provide a foundation that the additional cues build on, rather than being effective on its own. H1 is therefore only partially supported.
- **Landmarks, not route-based sequencing, aid integration.** Conditions 4 and 5, which included landmarks, significantly outperformed Condition 2, whereas Condition 3, which added route-based sequencing without landmarks, did not. This suggests that, once overlap is available, landmarks rather than route-based sequencing are the cue that improves layout understanding, supporting H2b while not supporting H2a.
- **No additive benefit from combining cues.** Condition 5, which combined both landmarks and route-based sequencing, did not significantly outperform either Condition 3 or Condition 4. This indicates that, once overlap is available, the presence of either cue may be sufficient, and combining them does not yield additional gains. Thus, H2c is not supported.

Behavioral evidence was consistent with this interpretation. Revisit behavior was highest in the baseline condition (Condition 1), indicating higher uncertainty and/or higher integration demand in the least supportive configuration. Revisiting can be interpreted as an attempt to refine perspective shifts through repeated exposure, which has been shown to reduce integration costs [36, 33, 34].

These results offer several theoretical contributions. The Network of Reference Frames theory was used to inspire the experimental design and to explain the mechanism underlying participants' performance, rather than being tested as a theory. Within this scope, the results are consistent with the spatial-integration mechanism implied by the Network of Reference Frames theory: spatial knowledge is constructed by linking multiple vista spaces via perspective shifts, and both task design (viewpoint cues) and individual ability (spatial perspective-taking) influence how effectively this integration occurs [33, 34]. More specifically, this study extends this theory-informed line of work by operationalizing key integration-related ideas in a controlled panoramic-viewpoint design.

Visual overlap can enable correspondences between consecutive views, but it was not sufficient on its own to reliably improve layout understanding. This supports the idea that integration costs are not only a matter of whether overlap exists, but also whether reference-frame relations can be maintained consistently across transitions [34]. Landmarks likely reduced integration costs by providing stable, re-identifiable anchors within overlap regions, thereby strengthening perspective shifts and helping participants maintain a consistent orientation across views. This interpretation is consistent with prior findings that landmarks and allocentric cues support cross-view alignment and spatial memory precision [1, 15, 25].

The weak incremental effect of route-based ordering in the current layouts suggests boundary conditions. A route-like sequence may not yield strong benefits when views remain discrete and the environment is small (4–6 viewpoints), because ordering alone may not provide enough continuity to support sustained updating across transitions. This does not imply that movement-related continuity is unimportant in general; rather, it suggests that the route-based ordering used here may have been too weak to reproduce the benefits observed in VR locomotion research, where continuous movement or enriched transitions can improve spatial updating [31, 42, 3]. At the same time, our findings complement that literature by shifting the focus from *how transitions are rendered* to *how a minimal set of viewpoints is placed and sequenced*.

From an applied perspective, for panoramic viewpoint systems intended to support building understanding, the results suggest that maximizing overlap alone is unlikely to be sufficient. Adding structured cues – particularly landmarks that persist across overlapping vistas – is a practical way to improve users' comprehension. More broadly, this work provides a reproducible experimental pipeline for systematically testing different viewpoint-system configurations under controlled constraints and with fine-grained behavioral logging. This offers an initial step toward evidence-based design rules for panoramic tours and related spatial-learning interfaces.

Beyond the design factors, the refinement process showed how strongly task framing shapes exploration: the per-view penalty used in the pilots encouraged early decisions, whereas the flat reward with enforced full exploration promoted more integrative viewing (Section 4.1). This underscores the importance of carefully calibrated framing in online spatial-cognition studies, where participants may otherwise adopt confounding shortcut strategies.

5.1 Limitations

While the combined confidence–correctness score provided an informative primary outcome, it collapses all incorrect trials to zero regardless of the reported confidence. As a result, the score does not preserve potentially informative differences among errors (e.g., low-confidence mistakes versus high-confidence mistakes). Although the CAC analysis showed that confidence generally tracked correctness, high-confidence errors can still occur, reflecting overconfidence or "lucky" responding. Further refinement of scoring methods is therefore warranted. For example, calibration-aware measures such as the Brier Score [5] or Confidence-Based Marking (CBM) [13] explicitly reward alignment between confidence and correctness while penalizing miscalibrated responses. Incorporating such measures (either analytically or within the incentive structure) could encourage more cautious and better-calibrated judgments and provide a fuller picture of performance.

Layout understanding was measured through a multiple-choice layout recognition test. This format is objective and scalable, but it may not fully reflect participants' internal spatial representations. As observed in the pilot phases, some participants developed response strategies based on the characteristics of the choice alternatives themselves (e.g., feature matching and elimination) rather than the spatial information gathered from the viewpoints. Moreover, the same selected layout may reflect different reasoning strategies, depending on how similar the distractor layouts are to the correct layout. Although the final protocol was designed to reduce such shortcut strategies (e.g., unlocking the choice screen only after all views were explored and limiting decision time), performance in a multiple-choice format remains partly dependent on the distractor set. Complementary assessment formats – such as freehand sketching [44] or spatial reconstruction tasks – could provide deeper insight into participants' mental models, at the cost of additional scoring complexity of final results.

The tested environments were small single-floor layouts with relatively few viewpoints. This may have limited the observable benefit of route-based ordering: in larger or multi-floor environments with longer paths and more transitions, continuity cues may play a stronger role in supporting integration.

Relatedly, the five buildings were selected to be structurally distinct rather than matched on complexity, but no quantitative complexity measure was applied. Prior work shows that structural or geometric layout complexity can affect the spatial knowledge acquired [38], so differences in complexity across the five buildings may have contributed to the between-layout variability in performance. This variability was absorbed by the layout random intercept in all models, so it did not bias the condition or spatial-ability estimates; nonetheless, an explicit complexity measure – for example, based on isovist or space-syntax properties – would allow this variability to be characterised rather than only controlled for, and is a worthwhile direction for future stimulus design.

Finally, prior experience with virtual environments or video games was not collected, yet such experience may influence comfort with panoramic navigation and exploration efficiency, and could thus contribute to individual differences independently of spatial ability; future studies could include it as a covariate. In addition, the sample was not balanced by gender (34 male, 16 female), which may limit the generalisability of the findings.

5.2 Future work

Several directions can extend this work and address its limitations. One open question concerns *orientation alignment*. Because initial panorama headings were randomized, participants may have incurred additional alignment costs at each view. Misaligned orientations are known to increase integration costs; for example, Meilinger [34] reported substantially higher errors when integrating information across 90°-rotated viewpoints compared to aligned ones. Future studies can compare (i) consistent global orientation (e.g., "north-up" initialization), (ii) alignment toward the next overlap region, and (iii) adaptive initialization based on the preceding viewpoint, to test whether reducing alignment disparity improves integration and confidence.

Another promising direction is to leverage the hierarchical structure of spatial memory. Hierarchical theories of spatial cognition propose that not all vista spaces are equally represented in memory [33, 19, 47]. Smaller or peripheral spaces may be encoded at lower hierarchical levels, while more central or integrative spaces may serve as higher-level anchors. A key rationale behind this structuring lies in clustering effects – where spatial representations are organized not as isolated views but as grouped regions or spatial units. Rather than presenting all views as independent and equally important, future studies could present spatially clustered views as cohesive groups, emphasizing the transitions between these higher-order clusters instead of transitions between individual views. For example, a central hall and its adjacent rooms could be presented as one group, followed by other overlapping or connected groups. One possible basis for defining such groups is the concept of *m-spaces* [40], i.e., to the minimal set of convex spaces required to cover an environment. A hierarchical presentation could then begin with the most connected m-spaces, followed by successively linked clusters, reflecting how spatial knowledge may be naturally organized in memory.

6 Conclusion

This exploratory study examined how viewpoint-system design supports building-layout understanding from a minimal set of 360° panoramas. Building on the Network of Reference Frames theory and spatial-integration concepts, three design factors were systematically manipulated: visual overlap between consecutive views, route-based sequencing, and landmark availability. A reproducible, web-based experimental framework was developed to implement these manipulations under controlled conditions and to log exploration behavior and responses in detail.

The results indicate that individual spatial ability – specifically *perspective-taking* – was a strong predictor of layout understanding ($p < .001$). In addition, viewpoint design was found to modulate performance: high visual overlap alone did not reliably improve performance, meaning that a hypothetical guard from the Art Gallery Problem solution would be unlikely to understand the gallery layout even with the help of overlapping views. Yet, overlap supported layout understanding as soon as it was paired with additional cues. Among the tested cues, landmarks produced the clearest improvement in layout understanding ($p < .05$). Route-based sequencing showed weaker effects in the present setting, and combining landmarks with route-based sequencing did not yield additional benefits beyond either cue alone.

Overall, these findings suggest a simple design implication for panoramic tours: overlap is helpful, but not sufficient on its own; stable landmarks visible across overlapping views can make layout understanding easier. At the same time, a central contribution of this work is the reproducible experimental pipeline. Because the framework is modular, new design variations (e.g., stronger continuity cues, different orientation rules, or alternative evaluation tasks) can be implemented quickly, tested efficiently, and compared under the same controlled setup. This provides a practical foundation for future studies to test additional design factors and to develop more general, evidence-based rules for generating effective viewpoint systems.

References

1 Christina Adamou, Marios N. Avraamides, and Jonathan W. Kelly. Integration of visuospatial information encoded from different viewpoints. *Psychonomic Bulletin & Review*, 21(3):659–665, 2014. doi:10.3758/s13423-013-0538-5.

2 Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi:10.18637/jss.v067.i01.

3 Jiwan Bhandari, Paul MacNeilage, and Eelke Folmer. Teleportation without spatial disorientation using optical flow cues. In *Proceedings of the 44th Graphics Interface Conference*, GI '18, pages 162–167, Waterloo, CAN, 2018. Canadian Human-Computer Communications Society. doi:10.20380/GI2018.22.

4 Sarah E Bonner and Geoffrey B Sprinkle. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society*, 27(4):303–345, 2002. doi:10.1016/S0361-3682(01)00052-6.

5 Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950. doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

6 Lucia A. Cherep, Alex F. Lim, Jonathan W. Kelly, Devi Acharya, Alfredo Velasco, Emanuel Bustamante, Alec G. Ostrander, and Stephen B. Gilbert. Spatial cognitive implications of teleporting through virtual environments. *Journal of Experimental Psychology: Applied*, 26(3):480–492, 2020. doi:10.1037/xap0000263.

7 Rune H. B. Christensen. *ordinal—Regression Models for Ordinal Data*, 2023. R package version 2023.12-4.1. URL: https://CRAN.R-project.org/package=ordinal.

8 V. Chvátal. A combinatorial theorem in plane geometry. *Journal of Combinatorial Theory, Series B*, 18:39–41, 1975. doi:10.1016/0095-8956(75)90061-1.

9 Ruth Conroy. *Spatial navigation in immersive virtual environments.* Phd thesis, University College London, 2001. URL: https://discovery.ucl.ac.uk/id/eprint/1111.

10 Joshua R. de Leeuw. jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1):1–12, 2015. doi:10.3758/s13428-014-0458-y.

11 Joshua R. de Leeuw. Datapipe: Born-open data collection for online experiments. *Behavior Research Methods*, 56(3):2499–2506, 2024. doi:10.3758/s13428-023-02161-x.

12 Alinda Friedman, Bernd Kohler, Peri Gunalp, Alexander P. Boone, and Mary Hegarty. A computerized spatial orientation test. *Behavior Research Methods*, 52(2):799–812, 2020. doi:10.3758/s13428-019-01277-3.

13 A. R. Gardner-Medwin. Confidence assessment in the teaching of basic science. *Research in Learning Technology*, 3(1), 1995. doi:10.3402/rlt.v3i1.9597.

14 Peter Green and Catriona J. MacLeod. Simr: an r package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4):493–498, 2016. doi:10.1111/2041-210X.12504.

15 Nathan Greenauer, Catherine Mello, Jonathan W. Kelly, and Marios N. Avraamides. Integrating spatial information across experiences. *Psychological Research*, 77(5):540–554, 2013. doi:10.1007/s00426-012-0452-x.

16 Nathan Greenauer and David Waller. Micro- and macroreference frames: Specifying the relations between spatial categories in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36:938–957, 2010. doi:10.1037/a0019647.

17 Mary Hegarty and David Waller. A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32(2):175–191, 2004. doi:10.1016/j.intell.2003.12.001.

18 Mary Hegarty and David A. Waller. Individual differences in spatial abilities. In Priti Shah and Akira Miyake, editors, *The Cambridge Handbook of Visuospatial Thinking*, pages 121–169. Cambridge University Press, Cambridge, 2005. doi:10.1017/CBO9780511610448.005.

19 Stephen C. Hirtle and John Jonides. Evidence of hierarchies in cognitive maps. *Memory and Cognition*, 13(3):208–217, 1985. doi:10.3758/BF03197683.

20 Corinne A. Holmes, Steven A. Marchette, and Nora S. Newcombe. Multiple views of space: Continuous visual flow enhances small-scale spatial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(6):851–861, 2017. doi:10.1037/xlm0000346.

21 Ronald Hübner and Jan Schlösser. Monetary reward increases attentional effort in the flanker task. *Psychonomic Bulletin & Review*, 17(6):821–826, 2010. doi:10.3758/PBR.17.6.821.

22 Toru Ishikawa and Daniel R. Montello. Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, 52(2):93–129, 2006. doi:10.1016/j.cogpsych.2005.08.003.

23 Peter Juslin, Nils Olsson, and Anders Winman. Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence–accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5):1304–1316, 1996. doi:10.1037/0278-7393.22.5.1304.

24 Amy A Kalia, Gordon E Legge, and Nicholas A Giudice. Learning building layouts with non-geometric visual information: The effects of visual impairment and age. *Perception*, 37(11):1677–1699, 2008. doi:10.1068/p5915.

25 Jonathan W. Kelly and Timothy P. McNamara. Reference frames during the acquisition and development of spatial memories. *Cognition*, 116(3):409–420, 2010. doi:10.1016/j.cognition.2010.06.002.

26 Maria Kozhevnikov and Mary Hegarty. A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & cognition*, 29(5):745–756, 2001. doi:10.3758/BF03200477.

27 Russell V. Lenth. *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2025. R package version 1.11.1. doi:10.32614/CRAN.package.emmeans.

28 Changyang Li, Haikun Huang, Jyh-Ming Lien, and Lap-Fai Yu. Synthesizing scene-aware virtual reality teleport graphs. *ACM Trans. Graph.*, 40(6), December 2021. doi:10.1145/3478513.3480478.

29 Hengshan Li, Panagiotis Mavros, Jakub Krukar, and Christoph Hölscher. The effect of navigation method and visual display on distance perception in a large-scale virtual building. *Cognitive Processing*, 22:239–259, 2021. doi:10.1007/s10339-020-01011-4.

30 Tingguang Li, Danny Ho, Chenming Li, Delong Zhu, Chaoqun Wang, and Max Q.-H. Meng. Houseexpo: A large-scale 2d indoor layout dataset for learning-based algorithms on mobile robots. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5839–5846, 2020. doi:10.1109/IROS45743.2020.9341284.

31 Alex F. Lim, Jonathan W. Kelly, Nathan C. Sepich, Lucia A. Cherep, Grace C. Freed, and Stephen B. Gilbert. Rotational self-motion cues improve spatial learning when teleporting in virtual environments. In *Proceedings of the 2020 ACM Symposium on Spatial User Interaction*, SUI '20, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3385959.3418443.

32 Jack M. Loomis, Roberta L. Klatzky, Reginald G. Golledge, and John W. Philbeck. Human navigation by path integration. In Reginald G. Golledge, editor, *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*, pages 125–151. Johns Hopkins University Press, Baltimore, MD, 1999.

33 Tobias Meilinger. The network of reference frames theory: A synthesis of graphs and cognitive maps. In Christian Freksa, Nora S. Newcombe, Peter Gärdenfors, and Stefan Wölfl, editors, *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, pages 344–360. Springer Berlin Heidelberg, 2008. doi:10.1007/978-3-540-87601-4_25.

34 Tobias Meilinger, Alain Berthoz, and Jan M. Wiener. The integration of spatial information across different viewpoints. *Memory & Cognition*, 39(6):1042–1054, 2011. doi:10.3758/s13421-011-0088-x.

35 Laura Mickes. Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, 4(2):93–102, 2015. doi:10.1016/j.jarmac.2015.01.003.

36 Daniel R. Montello. Spatial orientation and the angularity of urban routes: A field study. *Environment and Behavior*, 23(1):47–69, 1991. doi:10.1177/0013916591231003.

37 Daniel R. Montello. Scale and multiple psychologies of space. In Andrew U. Frank and Irene Campari, editors, *Spatial Information Theory A Theoretical Basis for GIS*, pages 312–321. Springer, 1993. doi:10.1007/3-540-57207-4_21.

38 Michael J. O'Neill. Evaluation of a conceptual model of architectural legibility. *Environment and Behavior*, 23(3):259–284, 1991. doi:10.1177/0013916591233001.

39 Joseph O'Rourke. *Art Gallery Theorems and Algorithms*. Oxford University Press, USA, 1987.

40 J Peponis, J Wineman, M Rashid, S Hong Kim, and S Bafna. On the description of shape and spatial configuration inside buildings: Convex partitions and their local properties. *Environment and Planning B: Planning and Design*, 24(5):761–781, 1997. doi:10.1068/B240761.

41 R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. URL: https://www.R-project.org/.

42 Kasra Rahimi, Colin Banigan, and Eric D. Ragan. Scene transitions and teleportation in virtual reality and the implications for spatial awareness and sickness. *IEEE Transactions on Visualization and Computer Graphics*, 26(6):2273–2287, 2020. doi:10.1109/TVCG.2018.2884468.

43 Yugandhar Shinde, Kyeongsuk Lee, Beyza Kiper, Makayla Simpson, and Sogand Hasanzadeh. A systematic literature review on 360° panoramic applications in architecture, engineering, and construction (aec) industry. *Journal of Information Technology in Construction*, 28:405–437, 2023. doi:10.36680/j.itcon.2023.021.

44 M. Simonet, C. Vater, C. Abati, S. Zhong, P. Mavros, A. Schwering, M. Raubal, C. Hölscher, and J. Krukar. Probing mental representations of space through sketch mapping: a scoping review. *Cognitive Research: Principles and Implications*, 10(1):59, 2025. doi:10.1186/s41235-025-00667-w.

45 Alexander M Smith and Amy-May Leach. Confidence can be used to discriminate between accurate and inaccurate lie decisions. *Perspectives on Psychological Science*, 14(6):1062–1071, 2019. doi:10.1177/1745691619863431.

46 Dajana Snopková, Hana Švedová, Petr Kubicek, and Zdeněk Stachoň. Navigation in indoor environments: Does the type of visual learning stimulus matter? *ISPRS International Journal of Geo-Information*, 8, 2019. doi:10.3390/ijgi8060251.

47 Albert Stevens and Patty Coupe. Distortions in judged spatial relations. *Cognitive Psychology*, 10(4):422–437, 1978. doi:10.1016/0010-0285(78)90006-3.

48 Eylul Tekin, K. Andrew DeSoto, John H. Wixted, and Henry L. Roediger III. Applying confidence accuracy characteristic plots to old/new recognition memory experiments. *Memory*, 29(4):427–443, 2021. doi:10.1080/09658211.2021.1901937.

49 Perry W. Thorndyke and Barbara Hayes-Roth. Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14(4):560–589, 1982. doi:10.1016/0010-0285(82)90019-6.

50 Davi C. Tozoni, Pedro J. de Rezende, and Cid C. de Souza. Algorithm 966: A practical iterative algorithm for the art gallery problem using integer linear programming. *ACM Trans. Math. Softw.*, 43(2), 2016. doi:10.1145/2890491.

A Participant exclusion criteria

Exclusion criteria were predefined and applied conservatively, so that a participant was excluded only when multiple independent indicators of inattentiveness or non-compliance co-occurred. General checks across all phases removed participants who used a mobile device, submitted incomplete data (fewer than five output files), or restarted the experiment and submitted duplicate files; none were excluded on these grounds.

In the second pilot phase, where the layout-choice screen could be opened after any number of viewed panoramas, exclusion was based on the following indicators. A *nonsensical-confidence* failure occurred when a participant decided after a single view in the tutorial while reporting very high confidence: because the tutorial was unsolvable from the first view alone, and participants who attempted to decide after the first view were explicitly warned of this, yet could still proceed. Reporting very high confidence in this situation flagged the participant as providing nonsensical responses, and led to exclusion on its own.

Three further indicators were defined: (i) deciding after a single tutorial panorama without high confidence, (ii) viewing fewer than three panoramas across all five trials – both reflecting consistent guessing or feature matching rather than layout integration – and (iii) skipping two or more of the three instruction screens (viewing time below a minimum threshold), reflecting inattention. A participant who met at least two of these three indicators was also excluded. In total, 12 of 54 participants were excluded.

In the final protocol, full exploration of all panoramas was enforced by the interface, so only two checks remained: below-threshold instruction-screen viewing times (as above) and a total task duration under five minutes. Participants were excluded only if they met both; 4 of 54 were excluded, yielding the final analytic sample of 50.