Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Curating for interrupted visits: an exact audit of thematic coverage and narrative order in virtual galleries

Hsin-Hung Lin

hhlin@asia.edu.tw

Asia University

Research Article

Keywords: virtual curation, virtual galleries, narrative sequencing, robust optimization, interrupted visits, reproducibility

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11603365/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Curating for interrupted visits: an exact audit of thematic coverage and narrative order in virtual galleries

Hsin-Hung Lin
Department of Creative Product Design, Asia University, Taichung, Taiwan
Correspondence: hhlin@asia.edu.tw | ORCID: 0000-0001-5518-4669

Abstract

Virtual galleries can deliver coherent complete exhibitions while leaving early-departing visitors with unbalanced thematic exposure. We develop an auditable sequencing model combining diminishing thematic returns, narrative dependencies, profile-specific activity times and uncertain termination. Exact enumeration of twelve eight-exhibit layouts yields 483,840 layout–route pairs across 324 conditions. Mean-profile and shortest-route policies lose an average of 0.0126 and 0.1483 worst-profile proxy units relative to robust planning. Analytical results distinguish thematic submodularity from narrative complementarity and bound survival, timing and search uncertainty. Beam search extends computation to 24 exhibits with explicit unresolved optimality gaps. A collection-grounded application uses eight author-supplied paintings, a designed obstructed gallery and a working desktop authoring prototype. Across 54 application conditions, Euclidean-planned routes violate graph-replayed time caps in 35 cases. Width-64 beam search recovers every application optimum; an eight-start relocation baseline is also competitive, with mean feasible regret of 0.00005 at 330 s and 0.00036 under tight caps. The prototype connects normalized artwork views, parameter controls and route exports, with eight programmatic functional checks. Author-supplied images are distinguished from stipulated curatorial groups, geometry and uncalibrated visitor profiles. Out-of-set preferences, specification sensitivity and feasible-set changes constrain the recommendation. These results establish reproducible computational and software behavior, not measured improvements in visitor learning, presence or aesthetic experience.

Keywords: virtual curation; virtual galleries; narrative sequencing; robust optimization; interrupted visits; reproducibility

1 Introduction

Virtual curation involves more than placing digital objects in an immersive room. A curator chooses which relationships become visible, which concepts appear early, and which interpretive connections depend on encountering one object before another. Spatial efficiency can conflict with these choices. A short path may introduce an explanatory example before its historical premise; a coherent complete sequence may postpone an entire theme until the end. When visits can terminate early, the quality of a full itinerary is therefore insufficient to characterize the content actually made available along its prefixes.

Recent work demonstrates the breadth of this design space. Personalized XR research organizes ways to adapt environments to users [1]. Context-aware movement modeling connects trajectories to interpretable interactions [2]. Visual and haptic systems make virtual objects manipulable [3], while empirical studies examine depth-dependent interaction and aesthetic experience [4,5]. These developments motivate a narrower computational question: given a fixed exhibition and uncertain visitor profiles, which planned ordering protects meaningful exposure before a visit ends?

The present study treats this question as a decision audit. It does not estimate visitor preferences, predict abandonment from headset signals, or introduce a new general routing algorithm. Instead, every feasible ordering of a small gallery is evaluated under the same explicitly stated scoring assumptions. This exact benchmark exposes the cost of optimizing average utility, distance, or completed narratives while neglecting

another design requirement. The distinction is consequential: optimality under a stipulated score is a mathematical property, whereas curatorial effectiveness is an empirical and interpretive claim.

The research questions are: RQ1, what is sacrificed by replacing worst-profile planning with average-profile or distance planning? RQ2, how do interruption intensity, thematic–narrative weighting and a restrictive budget change that comparison? RQ3, when do profile, timing, survival-shape and narrative assumptions invalidate an advantage? RQ4, can bounded beam search recover exact small-instance solutions, and what uncertainty remains at larger sizes? RQ5, can this model support an executable collection-grounded authoring case, and how do obstacles alter route feasibility? The contribution is a formulation and evidence chain connecting prefix-sensitive rewards, sensitivity/search bounds, a matched-objective algorithm comparison and an inspectable application. It is not a new general-purpose routing solver or a validated visitor-response model.

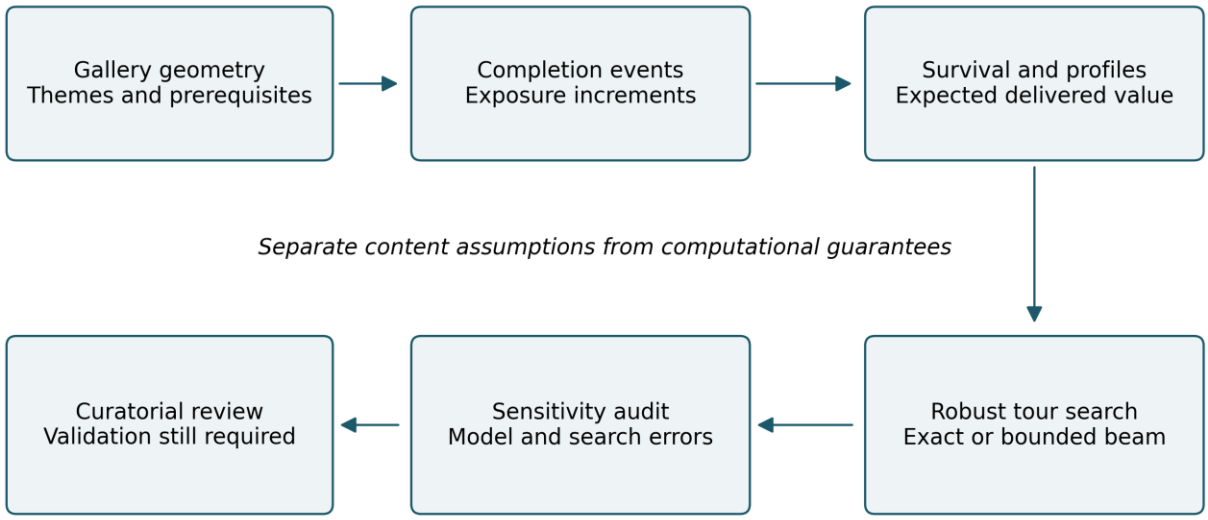

Fig. 1 Computational framework for interruption-sensitive virtual curation. Geometry and declared content relations determine completion events; survival and profile assumptions convert them to expected delivered value. Exact or bounded search returns a candidate, followed by specification and algorithmic audits. The final curatorial/empirical validation block is a required future step, not a completed user experiment. This is an original conceptual diagram corresponding to RQ1–RQ4.

Figure 1 separates three objects that are often conflated: the exhibition representation, the authoring objective, and the visitor response. The present analysis computes the first two. A visually convincing gallery or a numerically optimal route does not, by itself, validate the third.

2 Related work and conceptual positioning

2.1 Recent visual-computing directions

The 2025–2026 target-journal sample covers personalization [1], source-aware trajectory representation [2], real-time visuo-haptic interaction [3], and depth-sensitive selection [4]. These are selected relevant papers, not a census of The Visual Computer or a bibliometric trend estimate. Their common implication for this study is that a virtual environment has multiple levels of description: interaction mechanics, movement, adaptation, and interpretation. A gain at one level cannot automatically be called a gain at the others.

Huang et al. [1] synthesize heterogeneous personalization techniques. Here, profile uncertainty is represented explicitly rather than replaced by a single average user. Lemonari et al. [2] make movement around sources interpretable; our model instead assigns semantic consequences to a prescribed visit order. Zhou et al.

[3] illustrate the evidence required for a deployed VR system. We supply computational authoring evidence and make no headset frame-rate or tactile-performance claim. Bashar et al. [4], available here only as an abstract, motivate caution about activity duration; their experimental coefficients are not transferred to the present model.

Perceptual studies also caution against equating delivered content with received meaning. Jiménez-Navarro et al. [10] used two groups of 20 participants to test auditory suppression of visual performance. Their controlled manipulation concerns perception, whereas our completion event only denotes available exposure. Makled et al. [11] crossed rendering style and animation source with display modality in a 53-participant study. Their distinction between tracking fidelity and perceived behavioral realism motivates keeping representational detail separate from measured response. Neither paper supplies coefficients for the present model.

The methodological lesson from these studies and the system measurements in [3] is to align each claim with a distinct evaluation: computational optimality with an oracle or bound, system performance with measured implementation costs, and experiential quality with participants. Our revision strengthens the first category and reports limited timings for the second. It does not infer the third from either. The source-figure comparison in the accompanying journal analysis explains these distinctions without reproducing copyrighted figure panels.

2.2 Spatial generation, routing and aesthetic evidence

Full-text comparisons clarify the closest overlap. Rubio et al. [6] combine procedural room generation with evolutionary artwork arrangement and operator testing. Zhu et al. [7] combine interest-aware navigation, display detail and gaze interaction, with path and participant evaluations. Qian et al. [8] optimize reconstruction/display and exhibition layout, then assess visual experience and eye movements. Our decision variable is the ordering of fixed exhibits under possible early termination; the incremental evidence concerns delivered prefix value, explicit computational bounds and geometry-transfer failures. The accompanying matrix records the inspected sections and limits of these comparisons. These systems are not reimplemented as numerical baselines because their objectives, datasets and interaction pipelines differ.

The aesthetic comparison of Tomasetig et al. [5] used controlled observation stations and a fixed viewing order. That design helps isolate context effects but does not identify an optimal route. It also underscores why thematic exposure cannot be equated with liking, emotional involvement, memory, or presence. Our quality coefficient is an input to an interpretable surrogate objective, not a measured aesthetic quality score. The study therefore proposes a computational layer for curatorial deliberation, not an automated replacement for artistic judgment.

2.3 Relation to stochastic routing

Ferone et al. [9] formulate clustered reward collection under uncertain travel times and use simulation within adaptive large-neighborhood search. Their formulation and algorithm were inspected in full-text sections. The present problem shares time-sensitive reward collection, but the source of uncertainty and reward structure differ: visitors may terminate a fixed full-tour plan, themes saturate with repeated exposure, and prerequisite relationships reward an ordered pair rather than any completed set. Exhaustive enumeration is used to certify a small benchmark, not to compete with a scalable solver for a different problem.

These distinctions delimit the interpretation of the comparisons. A shortest full tour is a deliberately geometry-focused baseline. A no-exit policy isolates the omission of interruption; a coverage-only policy isolates narrative removal. Neither reproduces a published end-to-end VR system. The released literature matrix records question, method, data, finding, limitation, incremental scope, DOI, and access level for every cited paper.

3 Formal model

3.1 Geometry, duration and termination

Let π be a permutation of n exhibits, with entrance x_0 and exhibit coordinates x_i in metres. Each exhibit is visited once; no return to the entrance is required. The base synthetic benchmark uses obstacle-free straight-line navigation costs, without congestion, gaze-dependent detours or dynamic content. Section 4.8 replaces those costs with traversable graph distances in the application case. Neither cost model describes teleportation without a separate selection/orientation-time model. Profile s has positive speed v_s in metres per second and dwell multiplier m_s. Base dwell d_i is measured in seconds. Completion time includes travel and the full dwell at an exhibit:

$$T_{js}(\pi) = \sum_{1 \le \ell \le j} \left[\frac{\| x_{\pi\ell} - x_{\pi(\ell-1)} \|}{v_s} + m_s d_{\pi\ell} \right] \quad (1)$$

Equation (1) assumes deterministic activity times conditional on the profile. At the first position, the previous coordinate is the entrance. All terms have seconds as their unit. The full-tour feasibility rule is $\max_s T_{ns} \le B$. This is a restriction on an itinerary if completed, not a guarantee that a visitor remains in the gallery until B.

$$S_s(t) = Pr(L_s \ge t) = exp(-h_s t), \; h_s \ge 0 \quad (2)$$

L_s is a random termination time in seconds and h_s is a hazard in inverse seconds. The exponential survival law is a tractable stipulated model, not a fitted visitor distribution. It excludes route-induced boredom, fatigue feedback, voluntary skipping and re-entry. The zero-hazard limit means a visitor always completes any feasible route. General survival functions can replace the exponential in the expectation below; the simple exchange threshold later requires the exponential form.

3.2 Thematic saturation and narrative dependencies

Each exhibit i contributes a dimensionless coefficient a_{ik} in [0,1] to theme k. In the benchmark each exhibit has one theme and a coefficient between .55 and .95; other coefficients are zero. For a completed set A, thematic coverage is a saturating proxy:

$$F_k(A) = 1 - \prod_{i \in A} (1 - a_{ik}) \quad (3)$$

The product is a modeling choice for diminishing returns. It need not represent statistically independent learning events. Two exhibits with coefficients .7 and .8 produce .94, rather than 1.5; the second contributes .24 after the first. Order does not affect completed-set coverage, but it changes which marginal gain is delivered early. For prefix A_{j-1}, the increment is:

$$g_{jk} = a_{\pi j,k} \prod_{\ell < j} (1 - a_{\pi\ell,k}) \quad (4)$$

Profile weights w_{sk} are nonnegative and sum to one. Completing exhibit j requires survival to T_{js}. Linearity of expectation gives the exact expected thematic proxy, without sampling termination times:

$$C_s(\pi) = \sum_{j=1 \ldots n} S_s(T_{js}) \sum_{k=1 \ldots K} w_{sk} g_{jk} \quad (5)$$

A narrative edge (a,b) means that exhibit a introduces a premise used by b. The edge earns a unit only if a precedes b and the visitor finishes b. This is a deliberately strict order convention: retrospective interpretation receives no credit. With edge set E and uniform edge weights, the narrative proxy is:

$$N_s(\pi) = \frac{1}{|E|} \sum_{(a,b) \in E} \mathbb{1}\{r_a < r_b\} \, S_s(T_{r_b s}) \quad (6)$$

r_i denotes the position of exhibit i. Equation (6) makes narrative complementarity explicit. An isolated prerequisite has no edge reward until its dependent appears. The model does not require adjacent presentation, assess factual accuracy, or penalize misleading narratives. These omitted dimensions require a curator-defined extension rather than an interpretation of N as validated understanding.

$$U_s(\pi;\theta) = \theta C_s(\pi) + (1-\theta)N_s(\pi),\ 0 \le \theta \le 1 \quad (7)$$

Both terms lie in [0,1], so their convex combination is dimensionless. Equal numerical ranges do not make the constructs empirically commensurate: θ is a normative design weight. We vary it rather than estimate it from nonexistent visitor responses.

3.3 Robust planning and competing policies

$$\pi^* \in arg\,max_{\pi\in\Pi(B)}\, min_{s\in\{1,2,3\}}\, U_s(\pi;\theta) \quad (8)$$

Π(B) is the set of full permutations satisfying the time cap for all three modeled profiles. Equation (8) protects the least-scoring modeled profile. It is not a confidence guarantee for a visitor population. The mean-profile policy maximizes the equal-weight average of U_s; the shortest policy minimizes total travel distance within the identical feasible set. The no-exit policy sets every survival probability to one, maximizes the number of correctly ordered narrative edges, and breaks ties by distance. Completed-set thematic coverage is route-invariant, which makes this no-exit simplification exact for the benchmark. The coverage-only policy maximizes $\min_s C_s$ and is then evaluated using the original θ.

A sixth baseline is a feasible-prefix greedy policy. At each position it chooses the next exhibit with the largest minimum-profile prefix score divided by mean elapsed time, among choices that admit a feasible complete permutation. The implementation uses the exact candidate set for that feasibility test. It is therefore an oracle-assisted greedy diagnostic, not a scalable heuristic with an independent runtime advantage. This distinction prevents an unfair claim that exhaustive optimization is practically superior to a cheap online algorithm.

$$R(\pi) = min_s\, U_s(\pi^*) - min_s\, U_s(\pi) \ge 0 \quad (9)$$

Nonnegative regret in Equation (9) follows from construction on the same feasible set; observing it is not a discovery. Scientifically informative outputs are its magnitude, the sacrificed average score and distance, and whether the ordering remains useful when the modeled profiles are wrong. Ties within 10^{-12} are resolved by shorter travel, then permutation order.

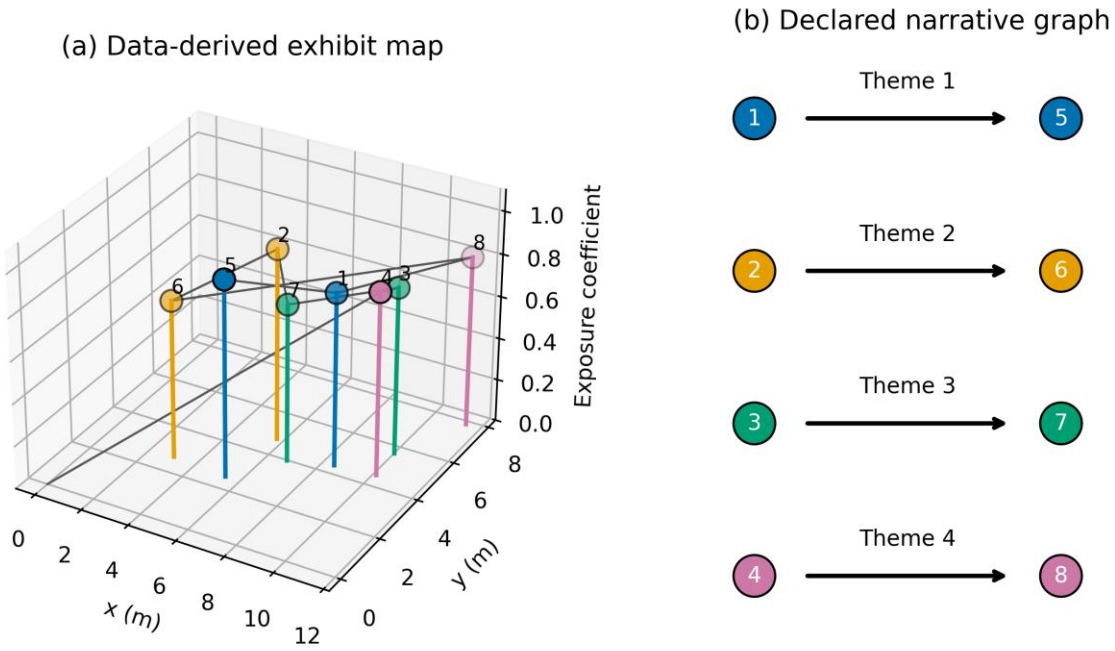

Fig. 2 Data-derived representation of the worked synthetic gallery. Panel (a) uses the exact input coordinates and exposure coefficients; height is a dimensionless coefficient, not physical pedestal height. Lines connect the chosen robust sequence. Panel (b)

shows the four declared prerequisite edges and theme identities. The original diagram is generated from inputs.json and the selected route in results.csv; it is neither a photograph nor an implemented headset scene.

The two views in Figure 2 have different analytical functions. Euclidean coordinates enter Equation (1), the coefficient axis enters Equations (3)–(5), and the directed edges enter Equation (6). Encoding a coefficient as height makes the assumptions inspectable but does not introduce a third spatial dimension into the route-cost calculation. Exhibit labels remain visible so the same object can be followed across the spatial and narrative views.

4 Analytical checks and interpretation

4.1 Exact prefix accounting

Let J be the number of exhibits completed before termination and set $S_s(T_{0s})=1$. Because completion times are ordered, the probability of a particular completed prefix is:

$$Pr(J = j) \;=\; S_s(T_{js}) \;-\; S_s(T_{j+1,s}),\;\; j < n;\;\; Pr(J = n) = S_s(T_{ns}) \quad (10)$$

For j=0 the first term is one. Multiplying each prefix probability by its completed thematic and ordered-edge score and summing over j gives Equations (5)–(7). The equality follows by telescoping: every marginal thematic increment and every satisfied edge is counted in exactly the prefixes that reach its completion time. This alternative computation is used as a numerical check. It validates the implementation of the stipulated expectation, not the psychological meaning of the scoring model.

4.2 A local ordering reversal

Consider two different-theme exhibits a and b with equal effective durations d after a common prefix. Assume swapping them changes neither subsequent completion times nor other narrative rewards. Let v_a and v_β be their already-weighted marginal thematic rewards and r the already-weighted reward for edge a→b. Under hazard h, common-prefix survival multiplies both orders and can be removed. The difference between a-before-b and b-before-a is:

$$\Delta \;=\; (v_a - v_b)[exp(-hd) - exp(-2hd)] \;+\; r\, exp(-2hd) \quad (11)$$

$$a\; before\; b \;\Leftrightarrow\; r \;\geq\; (v_b - v_a)[exp(hd) - 1] \quad (12)$$

When $v_\beta > v_a$, a narrative prerequisite is favored only if its edge reward compensates for delaying the larger thematic gain. Set v_a=.10, v_β=.30, r=.02 and d=30 s. The critical hazard is ln(1.1)/30 = 0.003177 s^{-1}. At h=.002, the required narrative reward is approximately .01237 and a-before-b is favored; at h=.008 the threshold is .05425 and b-before-a is favored. These are weighted proxy units, not observed comprehension probabilities. The result is a restricted two-item exchange identity. It does not imply a globally optimal sorting rule, because real swaps alter geometry, repeated-theme marginals and other dependency edges.

4.3 Budget and uncertainty boundaries

The minimum across profiles is monotone in each profile score, but it is nonsmooth where the worst profile changes. Consequently an average-profile improvement may lower the protected score. Expanding the profile set cannot improve the optimum on an unchanged feasible set. Reducing B also cannot improve it. These structural facts provide useful debugging checks; they do not establish which profile set is appropriate. All modeled profiles are treated as plausible scenarios with no estimated population frequencies.

For timing misspecification, a multiplier c scales travel and dwell durations together. Completion probabilities are recomputed at cT. We report both full-plan feasibility and a deadline-censored score that assigns zero marginal reward to completions later than B:

$$S_s(T_{js}; B, c) \;=\; exp(-h_s c T_{js})\, \mathbb{1}\{cT_{js} \le B\} \quad (13)$$

This cutoff evaluates the exposure of the fixed route if a hard session limit is enforced. It must not be confused with re-optimizing a shorter tour or an adaptive recourse policy. An infeasible full tour may still deliver a useful prefix, but that does not make the original full-tour planning guarantee valid.

4.4 Diminishing returns do not justify a generic greedy guarantee

For A contained in B and exhibit i outside B, the marginal coverage of theme k satisfies:

$$F_k(A \cup \{i\}) - F_k(A) \;=\; a_{ik}\prod_{u\in A}(1-a_{uk}) \;\ge\; a_{ik}\prod_{u\in B}(1-a_{uk}) \quad (14)$$

All coefficients lie in [0,1], so adding factors cannot increase the product. Hence completed-set thematic coverage is monotone and submodular. This places one component within the classical diminishing-returns setting [12]. The full route problem does not inherit the cardinality-constrained greedy guarantee: completion times depend on ordering and geometry, and pair rewards introduce complementarities. A prerequisite pair has zero completed-pair reward for either singleton but positive reward when both are available in order. Consequently, the original greedy-prefix diagnostic and the new beam method must be evaluated on this objective rather than assigned a generic approximation ratio.

4.5 Survival shape and an interpretable perturbation bound

The prefix expectation remains valid for any nonincreasing survival function with S(0)=1. To test the exponential restriction, we use a Weibull family calibrated to the same median m_s, not the same mean:

$$S_s(t;\kappa) = exp[-ln(2)(t/m_s)^{\kappa}],\; m_s = ln(2)/h_s,\; \kappa > 0 \quad (15)$$

Time t and median m_s are in seconds and shape κ is dimensionless. At κ=1 this is the base exponential. Shapes .6 and 1.8 describe decreasing and increasing hazards. Their common median isolates a particular shape perturbation; it does not match mean duration, early-exit frequency or every quantile. These distributions remain stipulated stress scenarios, not fitted visitor laws.

For a fixed route, write b_{js} for its nonnegative, already-weighted thematic and narrative reward delivered at event j. The total undiscounted reward is at most one. Under exponential termination:

$$\frac{\partial U_s}{\partial h_s} = -\sum_j b_{js} T_{js} exp(-h_s T_{js}) \le 0;\; \frac{\partial^2 U_s}{\partial h_s{}^2} = \sum_j b_{js} T_{js}{}^2 exp(-h_s T_{js}) \ge 0 \quad (16)$$

The derivatives have units of seconds and seconds squared because U is dimensionless and h has inverse-second units. They describe a fixed sequence, not a derivative through discrete reoptimization. Late, valuable events create sensitivity. The computed worked-route derivative is checked against central differences with step 10^{-7} s^{-1}; the largest absolute discrepancy is 8.31e-09 s. At a switch of the binding profile or route, the optimized minimum can be nondifferentiable.

If two survival laws differ by at most ε at every relevant completion event and for every planning profile, nonnegative rewards summing to at most one give the following bound on an unchanged feasible set:

$$|U_s - \tilde{U}_s| \le \varepsilon;\; |min_s U_s - min_s \tilde{U}_s| \le \varepsilon;\; R_{transfer} \le 2\varepsilon \quad (17)$$

The first inequality is the triangle inequality applied to the weighted event sum. The minimum operator is 1-Lipschitz in the largest coordinate error. For the final inequality, insert and subtract the two objective values of the route optimized under the perturbed law; its optimality cancels the middle term. This bounds regret from selecting an exact optimizer under the wrong survival law. It does not cover changed activity times or feasibility. A large ε yields a weak bound, which is informative about the information needed for calibration.

4.6 Narrative interpretation as a contested model parameter

Equation (6) gives no credit when a dependent exhibit appears before its premise. Retrospective interpretation may nevertheless be meaningful. We therefore retain forward credit one, grant reverse-order credit $\eta \in [0,1]$, and deliver either reward only after both exhibits have been completed:

$$N_s(\pi;\eta) = \frac{1}{|E|}\sum_{(a,b)\in E}[\mathbb{1}\{r_a < r_\beta\} + \eta\mathbb{1}\{r_\beta < r_a\}]\, S_s(T_{max(r_a,r_\beta),s}) \quad (18)$$

The parameter is dimensionless and normative. At $\eta=0$ the original model is recovered. At $\eta=1$ the pair no longer requires a direction, but still requires two completed exhibits. We neither estimate η nor interpret it as recall probability. Reoptimizing over η tests whether the design recommendation depends on the strict interpretive convention, rather than merely on numerical uncertainty.

4.7 An admissible beam search with an explicit certificate

A prefix p stores the visited set A, last position, completion times t_s, residual thematic products $\rho_k = \prod_{i\in A}(1-a_{ik})$, and accumulated unconditional expected value $u_s(p)$. Extending by exhibit i updates travel and dwell time as in Equation (1), updates the residual product, and adds only the newly delivered rewards:

$$u_s(p \oplus i) = u_s(p) + S_s(t'_s)[\theta\textstyle\sum_k w_{sk}\rho_k a_{ik} + (1-\theta)\frac{e_i(p)}{|E|}] \quad (19)$$

Here $e_i(p)$ counts edges completed in the permitted direction by appending i, and t'_s is the updated completion time. This search implementation uses the strict narrative model and base exponential survival. It does not call the enumerated full-tour feasibility oracle used by the original greedy-prefix diagnostic.

Let ρ_k^* be the residual product after all exhibits, and let Erem count edges whose dependent endpoint has not yet been visited. Ignoring future travel and permitting every remaining edge gives an optimistic upper bound:

$$UB_s(p) = u_s(p) + S_s(t_s)[\theta\textstyle\sum_k w_{sk}(\rho_k - \rho_k{}^*) + (1-\theta)\frac{Erem}{|E|}];\;\; UB(p) = min_s UB_s(p) \quad (20)$$

Every future survival is at most current survival. Remaining thematic increments telescope to the residual-product difference, and each counted edge can contribute at most one. Edges already reversed are excluded. Thus each profile-specific continuation value is bounded above, and taking their minimum still bounds the robust continuation objective. The bound is admissible but may be loose because it assigns all future rewards the current survival value.

With remaining exhibits R and current location ℓ, every future exhibit must have one incoming segment from ℓ or another remaining exhibit. Ignoring connectivity yields the necessary time bound:

$$LB_s(p) = t_s + m_s\textstyle\sum_{i\in R} d_i + \frac{1}{v_s}\sum_{i\in R} min_{v\in(R\cup\{\ell\})\setminus\{i\}} dist(v,i) \quad (21)$$

All terms are seconds. The incoming minima may form disconnected cycles, so passing this test does not establish feasibility; failing it for any profile safely prunes the prefix. The search expands surviving prefixes, ranks them by UB and keeps at most W per depth. It records the largest upper bound D over every discarded prefix. Any feasible full route belongs either to a retained terminal branch or a discarded branch; therefore, for a returned best score L:

$$L \le OPT \le max(L,D);\;\; G_{cert} = max(L,D) - L \quad (22)$$

No approximation ratio is assumed. If no complete feasible route survives, the run is reported as a failure rather than a zero-valued route. Bounds are checked against all exact completions of 120 sampled prefixes in the first layout, including value and time bounds. For a fixed W, at most Wn candidates are expanded at a depth. The straightforward incoming-bound implementation examines $O(n^2)$ pair distances per candidate, giving

$O(Wn^4)$ arithmetic work before sorting and $O(Wn^2)$ retained candidate storage. These are implementation bounds, not a claim of asymptotic novelty.

4.8 Navigation geometry and a fixed-route timing bound

For a connected navigation graph $G=(V,E^G)$, with nonnegative edge lengths ℓ_e in metres and stations embedded at graph vertices, substitute graph distance for the Euclidean segment length in Equation (1):

$$d_G(u,v) = min_{p:u\to v}\sum_{e\in p}\ell_e;\ \ \Delta t_{uv,s} = \frac{d_G(u,v)}{v_s} \quad (23)$$

The paths are shortest only on the declared graph. Four-neighbor movement adds direction restrictions as well as obstacle detours; it is not continuous-space geodesic navigation. A physically deployable system must account for body clearance and turning. Here the point-agent graph is a transparent authoring abstraction. Artwork wall positions and viewing stations are distinct: a stipulated 1.5 m virtual height and source pixel aspect ratio determine display size, while station coordinates determine route costs.

For the same full ordering, let T^E_{js} and T^G_{js} denote Euclidean and graph completion times, $\delta T_{js}=T^G_{js}-T^E_{js}\geq 0$, and b_{js} the nonnegative mixed thematic/narrative increment before survival discounting. These coefficients are dimensionless and sum to at most one. With the same exponential hazard $h_s\geq 0$ in both evaluations:

$$U_s^{\,E} - U_s^{\,G} = \sum_j b_{js} e^{-h_s T_{js}^{\,E}}[1-e^{-h_s\delta T_{js}}] \ \leq\ min(1, h_s max_j \delta T_{js}) \quad (24)$$

The loss is nonnegative. The bound follows from $1-\exp(-x)\leq x$ for $x\geq 0$ and $\sum_j b_{js}\leq 1$. Hazard has units s^{-1} and delay has units s, so the bound is dimensionless. It applies to a fixed route with unchanged rewards and deterministic times; it does not certify the route's time feasibility, preserve its optimality, or account for a changed exit process. A small score perturbation can coexist with a hard budget violation. Appendix C provides a complete eight-step evaluation.

5 Computational design and reproducibility

All inputs to the base benchmark were generated for this study and are explicitly synthetic. Twelve independent layouts use eight exhibit coordinates sampled uniformly over $x\in[1,12]$ m and $y\in[1,8]$ m, with entrance (0,0). Dwell durations are integer draws from 18 through 30 s, and exposure coefficients are uniform on [.55,.95]. Exhibits 1–4 introduce themes 1–4; exhibits 5–8 revisit them, with edges 1→5, 2→6, 3→7 and 4→8. This repeated structure isolates order-dependent narrative completion from theme repetition. The separate collection case in Section 6.10 uses author-supplied images and provisional curatorial groups, while retaining explicitly designed geometry and visitor assumptions.

Table 1 Stipulated visitor profiles; weights are dimensionless, speed is m/s, and base hazard is s^{-1}

Profile	Theme weights (1,2,3,4)	Speed	Dwell factor	Hazard
1	[0.55, 0.15, 0.15, 0.15]	0.7	1.1	0.002
2	[0.1, 0.55, 0.2, 0.15]	1.0	1.0	0.003
3	[0.1, 0.15, 0.25, 0.5]	1.3	0.9	0.004

The factorial design crosses 12 layouts, hazard multipliers {.5,1,2}, $\theta\in$ {.25,.50,.75}, and three budget regimes. The tight budget is the ceiling of the shortest feasible maximum-profile completion time plus 5 s; the other caps are 330 and 420 s. There are 324 design conditions and 1,944 policy rows. Each layout has 8!=40,320 candidate orders, giving 483,840 distinct layout–route pairs. Reusing a route under another weight or hazard setting does not create a new independent dataset.

We calculate expectations exactly for every route rather than simulate individual visitors. Independent random sampling occurs only in generating the 12 layouts and in resampling those layouts for a design-

dispersion interval. The pseudo-random seed is 20261102; bootstrap seed is 20261103. Across 1,080 route/profile checks, a separate prefix-outcome summation agrees with marginal-reward expectations to 2.22e-16. The comparison samples ten lexicographically spaced routes for every layout, hazard multiplier and profile. It is not a second independent optimization solver. An additional scalar implementation independently recomputed geometry, set-product coverage and edge completion for all 1,944 selected policy rows (5,832 profile evaluations); the maximum absolute discrepancy was 5.69×10^{-14}.

For uncertainty reporting, each layout contributes one average regret per policy across its 27 factorial settings. Five thousand paired bootstrap resamples of the 12 layouts produce central 95% percentile ranges. These ranges describe sensitivity to the stipulated design ensemble and its resampling; they are not confidence intervals for real museum visitors. Deterministic route enumeration has no Monte Carlo error. We do not report p-values or treat the 1,944 dependent policy rows as independent observations.

Out-of-set preference tests use one dominant theme with weight .85 and three weights .05, paired with hazards .001 or .008 s^{-1}, speed 1 m/s and dwell multiplier 1. These eight conditions are outside the planning profiles. For each layout, policies selected at θ=.5, base hazards and B=330 s are held fixed. Timing tests independently apply $c\in\{.8,1,1.2,1.5\}$. These perturbations stress model misspecification; they are not a probability sample of plausible audiences.

The computational implementation uses Python 3.12.14, NumPy 2.3.5, SciPy 1.17.0 and pandas 2.2.3. Figures are generated with Matplotlib 3.10.8. The main calculation took approximately 13.7 s in the execution environment; this is a reproducibility observation, not a controlled hardware benchmark. All inputs, selected routes, condition summaries, worked-route arrays and source code are provided in Online Resource 1.

The additional design audit contains 108 survival-transfer rows (12 layouts × 3 shape values × 3 policies), 60 narrative-credit rows, 120 beam runs with exact eight-exhibit comparators, and 48 larger beam runs. The larger extension uses 12 independently generated layouts: four each with 12, 16 and 24 exhibits, seed 20261104, the same geometry/dwell/coefficient generators, and repeated-theme edges $i\to i+4$ wherever defined. Its time cap is 40n s. Widths are {1,4,16,64,256} for exact comparisons and {4,16,64,256} for larger cases. Eight-exhibit runs use both 330 s and the original tight cap. These repeated policy/model evaluations are not independent visitor observations.

The beam routine carries no exhaustive route set. All successful returned routes are independently rescored in a separate scalar implementation: 146 routes and 438 profile evaluations, maximum discrepancy 2.22e-16. Timings cover one sequential execution per configuration in the shared Python environment and exclude figure generation. They are reproducibility diagnostics, not controlled latency benchmarks or headset frame rates. Online Resource 1 contains every input, route, failure and timing; plotting summaries retain failures in their denominator.

6 Results

6.1 Geometry and the price of protecting a profile

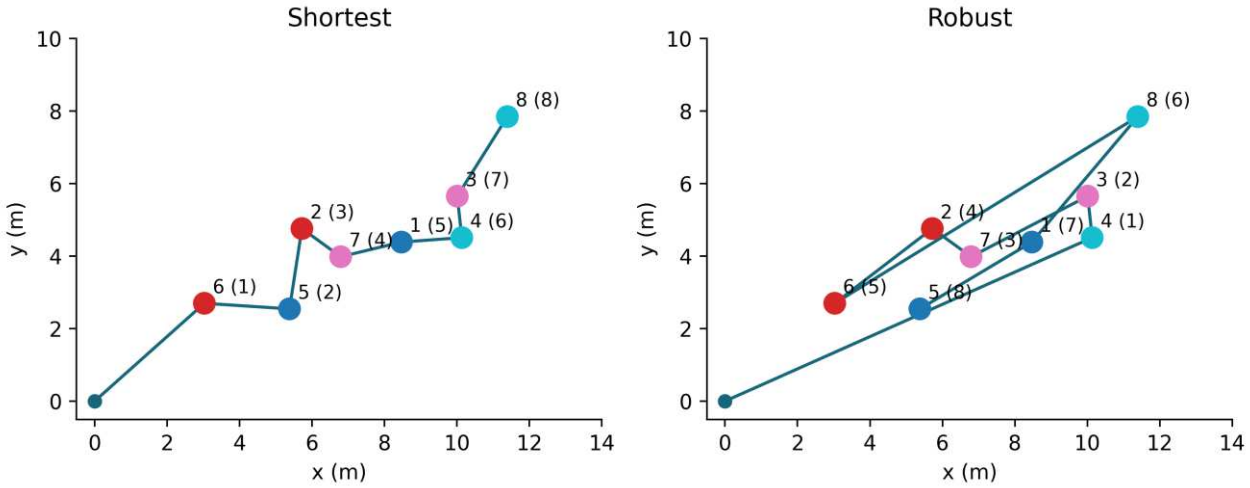

Fig. 3 Shortest and robust routes for the first synthetic layout at θ=.5, base hazards and B=330 s. Exhibit labels give object ID followed by visit position in parentheses. Coordinates are in metres; the entrance is (0,0). The geometry comes directly from inputs.json and the selected sequences from results.csv. Lines depict planned navigation costs, not recorded visitor movement

For the worked gallery, the shortest route travels 17.110 m, whereas the robust route travels 38.524 m. The robust worst-profile score is .666733, compared with .374379 for shortest routing. Most of this particular difference concerns narrative ordering: the minimum narrative proxy is .585941 versus .116282. This contrast should not be read as evidence that longer routes are better. Distance is only one contributor to the stipulated duration, and the narrative score explicitly penalizes reversed prerequisites.

Table 2 Worked-gallery comparison; all scores are dimensionless synthetic proxies

Policy	Worst score	Mean score	Travel m	Max time s
Robust	0.6667	0.6815	38.52	272.83
Mean-profile	0.6481	0.6924	34.14	266.56
Shortest	0.3744	0.4370	17.11	242.24
No-exit	0.6212	0.6839	26.00	254.95
Coverage-only	0.5131	0.5335	42.53	278.56
Greedy-prefix	0.5099	0.5515	35.94	269.14

The stronger comparison is against mean-profile optimization. Robust planning raises the worst score from .648107 to .666733, an absolute increase of .018627, while reducing the mean score from .692447 to .681453. Travel increases by 4.389 m. Thus protection has an explicit cost: the policy does not dominate on average utility or distance. A curator who accepts the specified average-profile distribution could reasonably select the mean policy. The model makes that value judgment visible rather than hiding it inside an aggregate ranking.

6.2 Baselines, ablations and budget restrictiveness

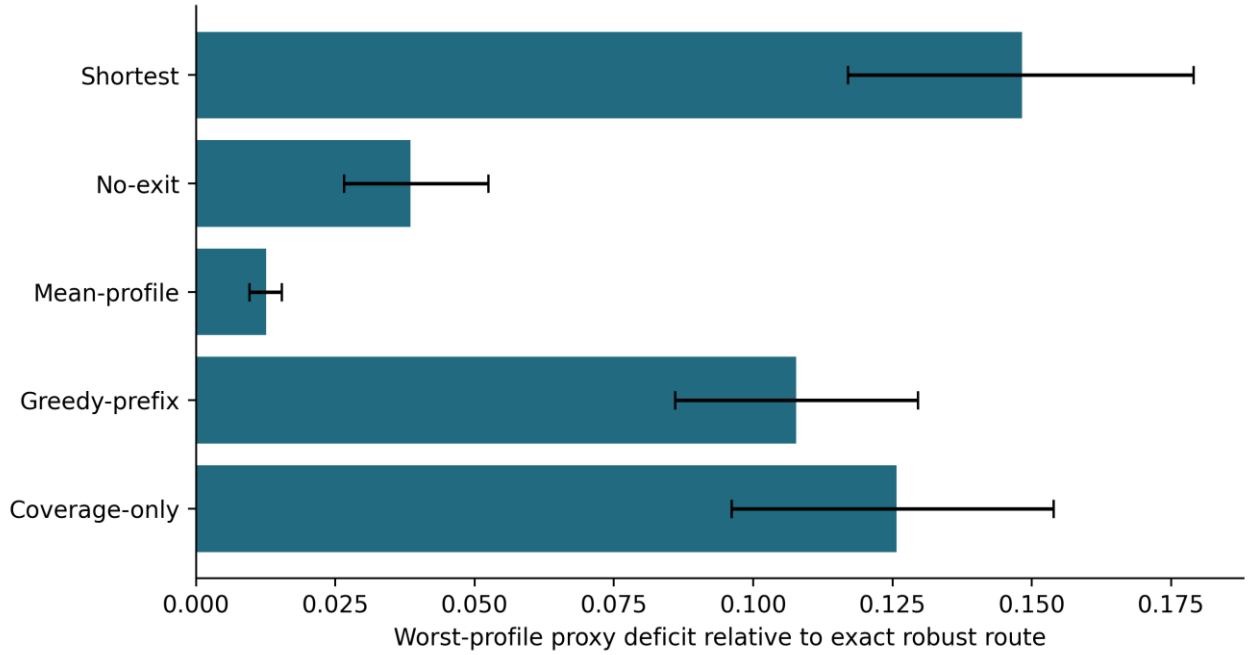

Fig. 4 Mean worst-profile deficit relative to exact robust planning across the 324 design conditions. Error bars are paired 95% percentile design-resampling ranges over 12 synthetic layouts, with 5,000 resamples. They are not visitor confidence intervals. Values are traceable to summary.csv; the robust policy has zero deficit by definition

Table 3 Conditional design-ensemble regret; lower values are closer to the exact robust objective

Policy	Mean	Resampling range	Maximum
Coverage-only	0.1257	[0.0961, 0.1539]	0.5748
Greedy-prefix	0.1077	[0.0860, 0.1295]	0.4655
Mean-profile	0.0126	[0.0096, 0.0154]	0.0703
No-exit	0.0385	[0.0266, 0.0524]	0.1629
Robust	0.0000	[0.0000, 0.0000]	0.0000
Shortest	0.1483	[0.1170, 0.1790]	0.4688

Average-profile optimization is the closest of the tested alternative criteria. Its mean deficit is .012622, substantially smaller than the .148283 deficit of distance-only planning. Coverage-only and greedy-prefix policies also lose narrative value, with mean deficits .125710 and .107749. Removing interruption gives .038477. These differences reflect objective mismatch and selected-order effects. They do not establish algorithmic superiority over genetic, reinforcement-learning or large-neighborhood-search methods that solve the same objective; such methods were not compared.

The tight budgets leave only 25–307 of 40,320 orders feasible per layout (mean 95.9), while 330 s leaves 36,918–40,320 and 420 s leaves all routes feasible. The wide budget is intentionally a saturation control: once all routes are feasible, increasing the cap cannot improve the optimum. Tight caps sharply constrain the room available for thematic and narrative rearrangement. Reporting the feasible-set size distinguishes genuine budget sensitivity from a parameter that is numerically inactive.

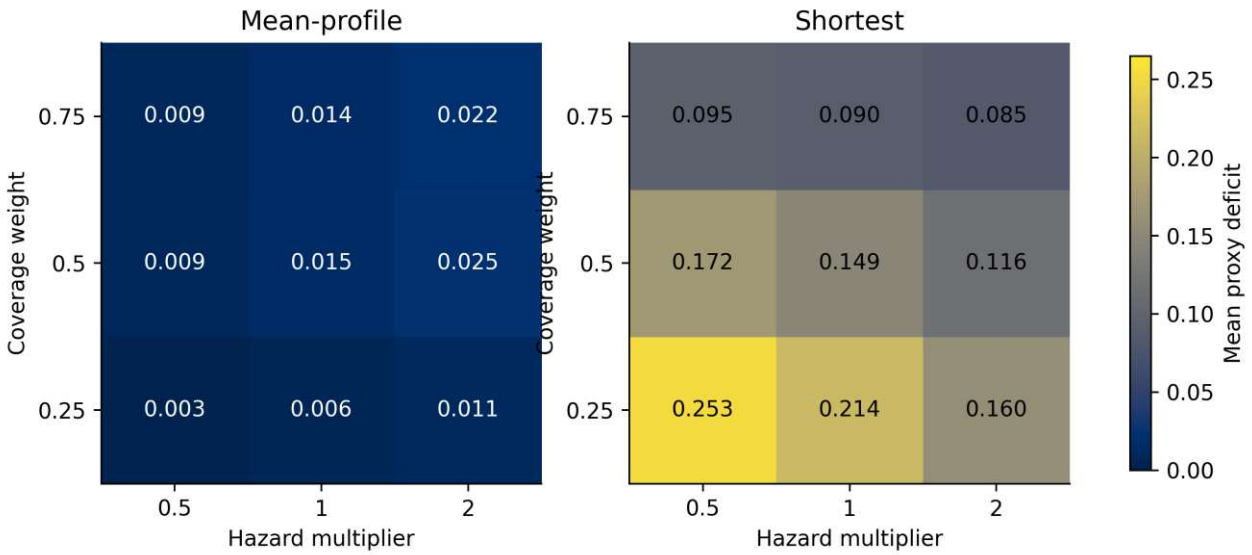

Fig. 5 Mean deficits for average-profile and shortest-route planning by hazard multiplier and thematic weight. Each cell averages over the 12 layouts and three budget regimes. The shared scale and printed values support comparison without interpreting colour alone. The data are grouped from results.csv; differences are conditional on the synthetic profile set

Figure 5 separates interruption severity from the curator's thematic weight. A larger hazard does not guarantee a larger regret: it can diminish both policies' rewards or change which profile binds. Likewise, raising θ reduces the reward for prerequisite completion, potentially narrowing a disadvantage caused by reversed edges while changing the thematic ordering. The interaction is a property of coupled objectives, not a universal monotonic rule.

6.3 A multidimensional reading of the feasible set

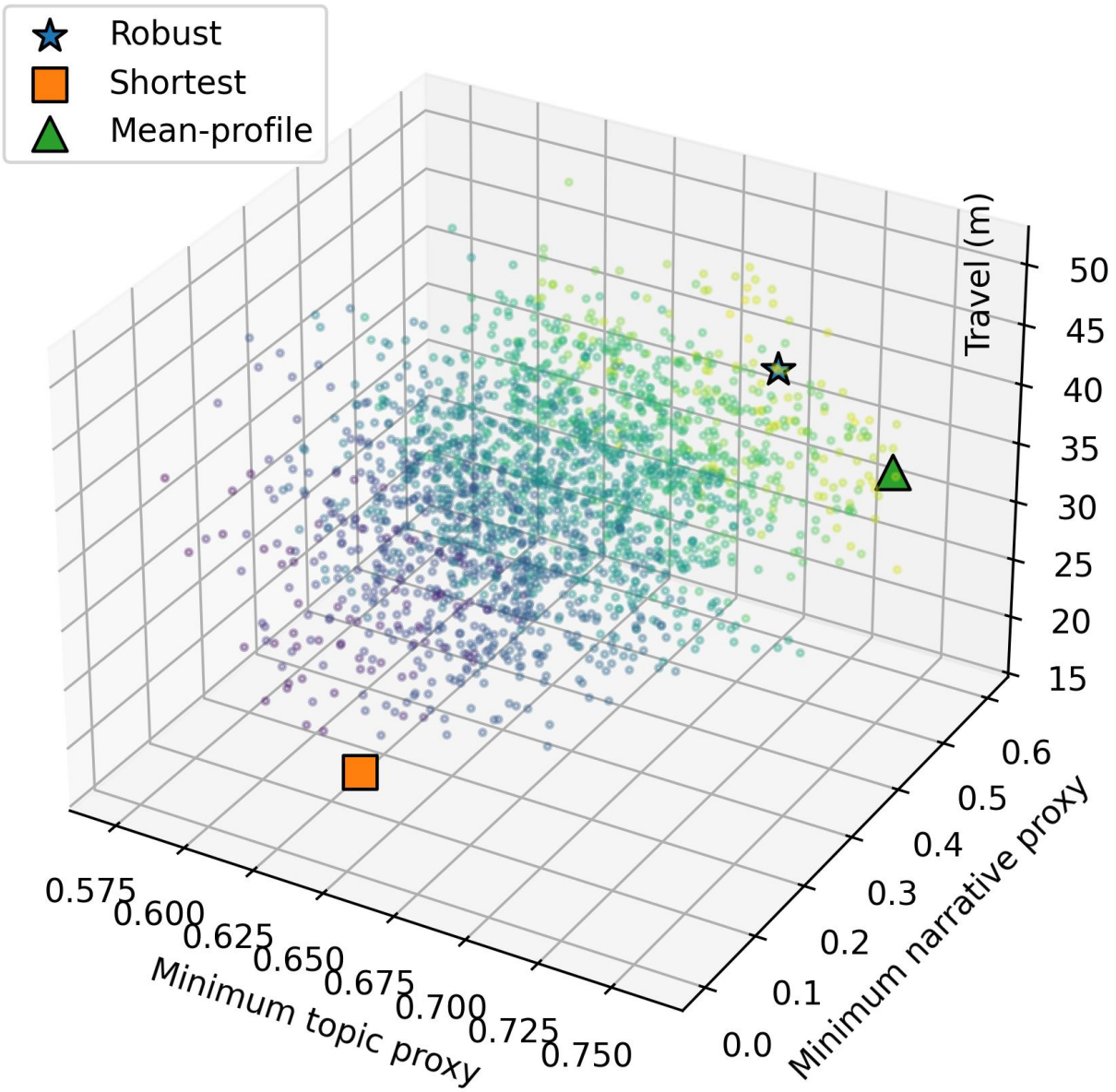

Fig. 6 Three-dimensional diagnostic for the worked gallery: minimum thematic proxy, minimum narrative proxy and route distance. Every twentieth feasible lexicographic route is shown for background clarity; all 40,320 routes were optimized. Highlighted policies use exact selected values. The plotted minima may belong to different profiles, so their average is not necessarily the minimum combined score. Data: worked_tours.npz

The three-dimensional plot exposes a potential interpretive error. A route can improve the minimum thematic component but damage narrative completion, or vice versa. Moreover, $\min_s[\theta C_s+(1-\theta)N_s]$ is generally not $\theta\min_s C_s+(1-\theta)\min_s N_s$, because different profiles may attain the component minima. Figure 6 is therefore a diagnostic projection, not a graphical reconstruction of the objective. Table 2 reports the actual combined scores used for selection.

This distinction matters when a professional-looking chart is used to justify a recommendation. A Pareto-like cloud should not conceal the population definition, the normalization of objectives, or the role of a subjective weight. The model retains per-profile values in its worked arrays so that the active constraint and the cost of protection can be inspected.

6.4 Timing misspecification and delivered prefixes

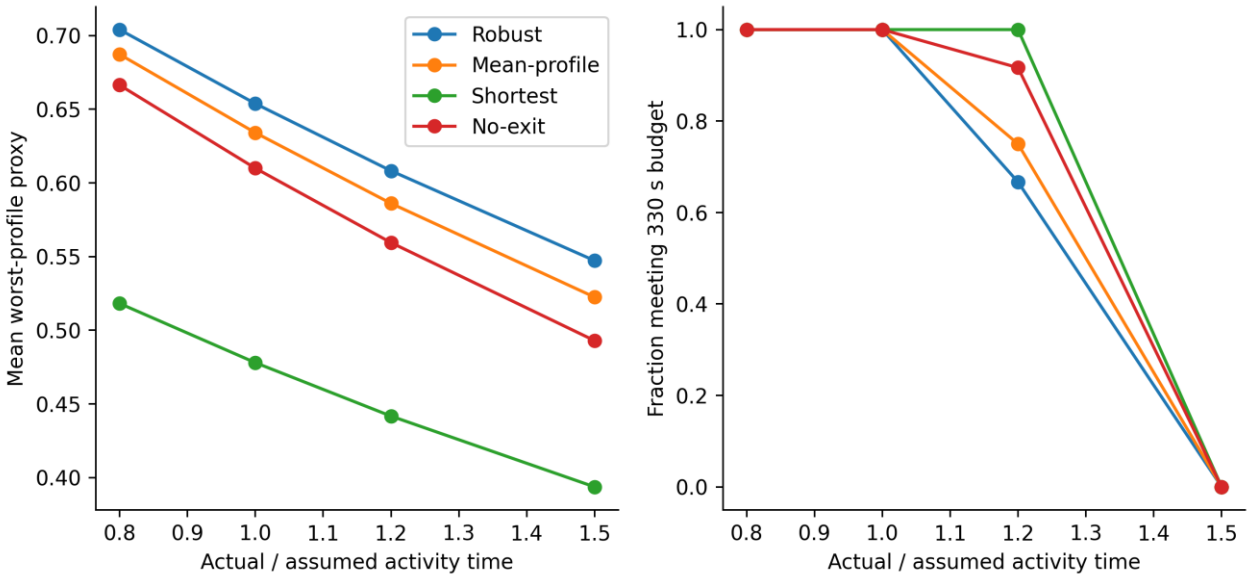

Fig. 7 Fixed policies under common activity-time misspecification. Left: mean worst-profile score evaluated without a hard cutoff. Right: fraction of the 12 layouts whose complete itinerary still meets 330 s for all planning profiles. These are different outcomes; an attractive untruncated score does not certify feasibility. Deadline-censored scores are separately released in deadline_censoring.csv

At 20% longer activity durations, only 8 of 12 robust routes remain full-tour feasible, versus 9 mean-profile routes, 11 no-exit routes and all 12 shortest routes. At 50% longer durations, none of these complete routes remains feasible. This adverse result limits the interpretation of robust planning: robustness over preference profiles is not robustness over every uncertain parameter. A session cap must be checked under plausible duration errors or addressed through an explicit recourse model.

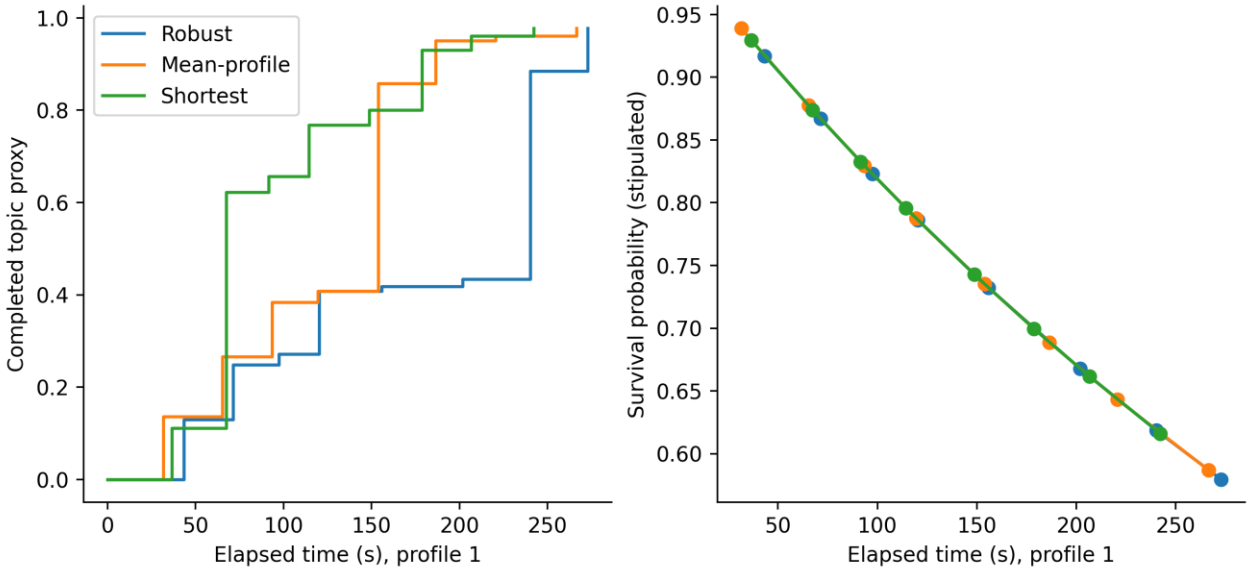

Fig. 8 Worked-gallery prefix interpretation for profile 1. Left: completed thematic proxy as a function of elapsed time; right: stipulated survival at the exhibit-completion events. Step curves depict available content, not measured knowledge. Their combination through Equation (5), together with ordered-edge completion, generates the expected score. Data: worked_tours.npz and worked_steps.csv

A completed exhibition can cover every theme well while exposing late themes only to visitors who remain long enough. Figure 8 separates the deterministic content schedule from the assumed survival law. A future empirical study could estimate each independently, then test whether the product-based thematic proxy predicts any meaningful response. The present calculation does not perform that calibration. Deadline-censored outputs further prevent a long but infeasible plan from being reported as fully deliverable.

6.5 Out-of-set failure and the exchange threshold

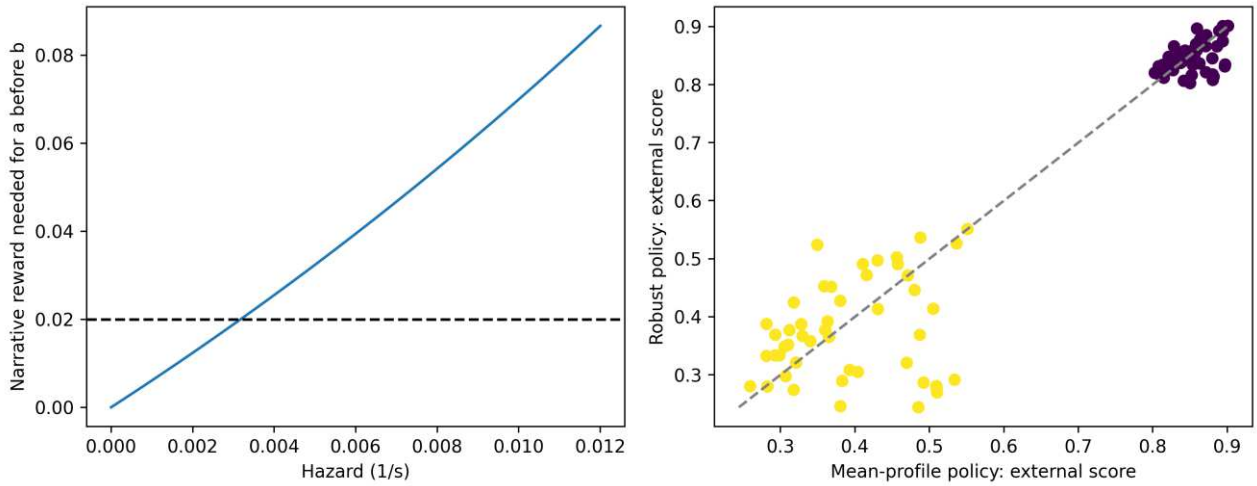

Fig. 9 Left: the exact two-item narrative-reward threshold in Equation (12), with v_a=.10, v_β=.30, d=30 s and reference reward .02. Right: fixed-policy scores in 96 out-of-set layout/preference/hazard conditions; the diagonal marks equality. Points below the diagonal favour the mean-profile policy. Inputs and outputs are in swap_boundary.csv and outside_profiles.csv

The robust policy scores below the mean-profile policy in 38 of 96 out-of-set conditions. This is an observed failure within a deliberately adverse computational test, not an estimated failure rate for museum visitors. The result demonstrates that protecting three profiles need not protect a fourth. The largest robust-minus-mean differences in the tested external grid are -0.2423 and 0.1744. A claim of universal dominance would contradict these outputs.

The local threshold explains a second failure of simple rules. Prerequisites should not always come first under an interruption-sensitive proxy: an important stand-alone theme may deserve earlier exposure when the narrative reward is too small. Conversely, greedily maximizing immediate theme coverage can sacrifice a valuable pair completion. The analytical identity does not decide the curatorial value of r; it states exactly how that declared value interacts with interruption and duration.

6.6 Transfer across survival shapes

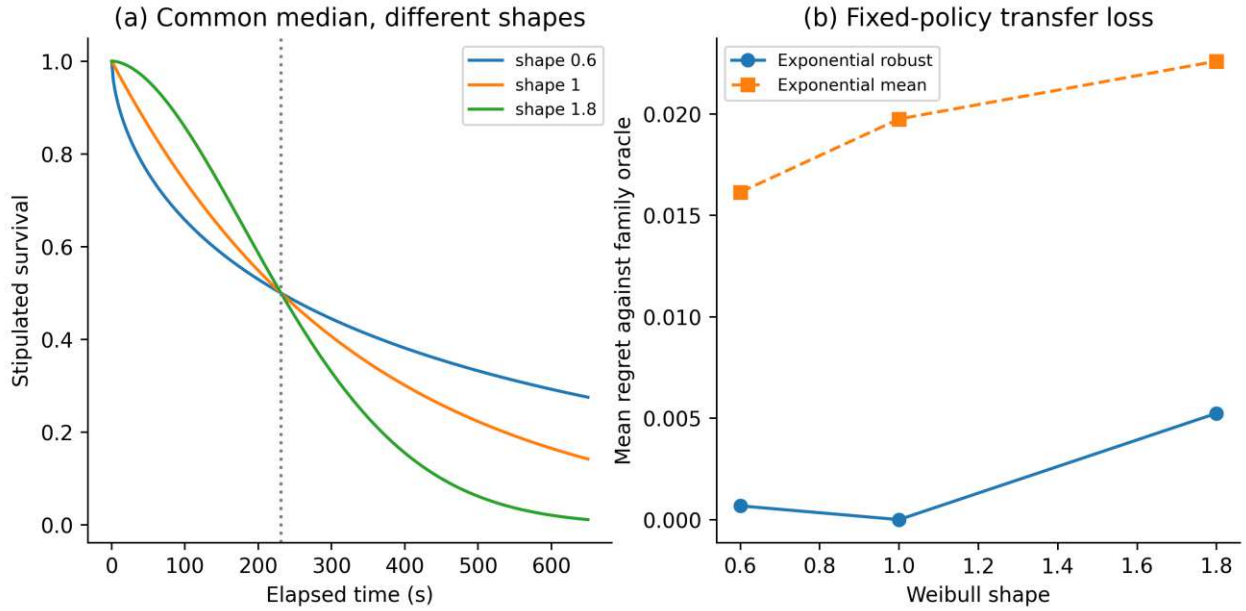

Fig. 10 Survival-shape sensitivity with a common median. Panel (a) illustrates the three stipulated laws for the second profile ($h=.003\ s^{-1}$). Panel (b) averages fixed-policy regret over the 12 layouts, relative to exact reoptimization under each shape, at $\theta=.5$ and $B=330$ s. Exponential robust and mean routes are frozen before transfer. Source: survival_transfer.csv; no participant survival data are used.

At shape .6, the fixed exponential-robust route loses .000678 proxy units on average relative to the matched-family oracle; at shape 1.8 the loss grows to .005243. The corresponding mean-policy deficits are .016145 and .022606. The exponential robust route necessarily has zero transfer regret at shape one. Unlike the broader factorial mean in Table 3, these figures use only the base hazard and 330 s cap; their numerical levels should not be compared as if the designs were identical.

Figure 10 shows why a single average-duration calibration is insufficient. Equal medians align survival at one time but leave early and late events weighted differently. Increasing-hazard termination rewards getting valuable pairs completed before the steeper part of the curve. Decreasing-hazard termination gives a long tail to visits that survive the early interval. The small transfer losses in this generator are conditional robustness evidence, not permission to ignore survival calibration in an actual collection.

Table 4 Additional computational tests and units of replication

Analysis	Configuration rows	Independent layouts	Interpretation
Survival transfer	108	12	Fixed routes vs matched-family oracle
Narrative credit	60	12	Normative assumption sensitivity
Beam exact comparison	120	12	Two budgets and five widths
Larger beam runs	48	12	Four layouts per size; optimum unknown

6.7 Search quality and feasible-route failures

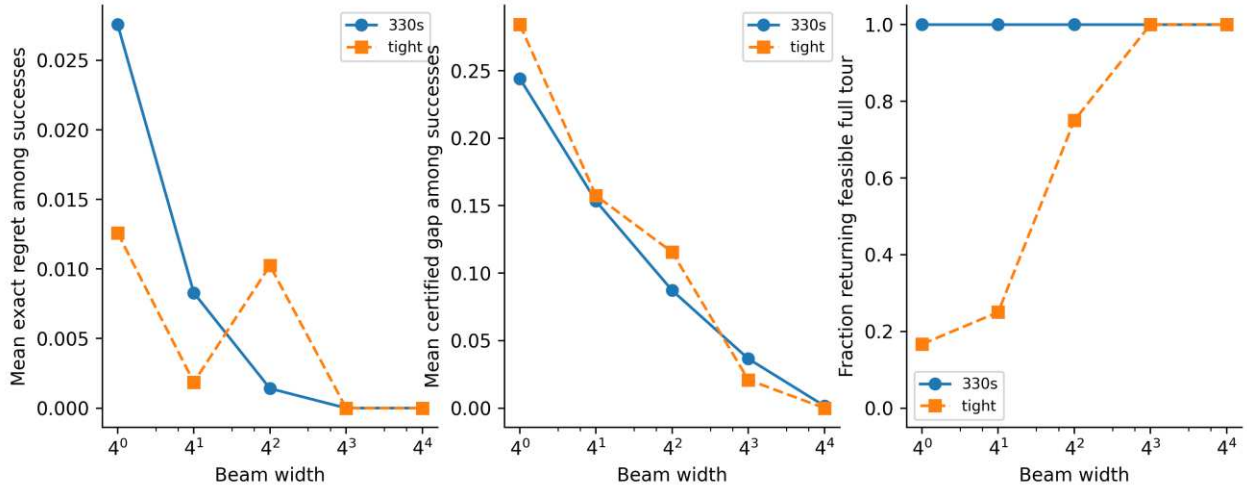

Fig. 11 Beam search against the exact eight-exhibit optimum. Left: mean realized regret among successful runs; centre: mean certified upper-minus-returned gap among successful runs; right: fraction of all 12 layouts with a returned feasible route. Width is logarithmic. Dashed lines denote tight caps; solid lines denote 330 s. A lower mean among few successes does not establish a better method. Source: beam_exact.csv.

At 330 s, every beam width returns a feasible route. Width one loses .027572 on average; width 16 loses .001408; widths 64 and 256 recover the exact optimum in all 12 layouts to numerical precision. Tight budgets expose a different failure mode: widths 1, 4 and 16 succeed in only 2, 3 and 9 layouts, respectively, while 64 and 256 succeed in all 12. In particular, the small conditional regret of width four hides nine failures. Figure 11 makes this selection effect visible rather than averaging failed runs away.

The certificate and realized regret answer different questions. At width 64 the 330 s realized regret is effectively zero, but the mean certificate gap remains .036533 because discarded prefixes retain optimistic bounds. Width 256 narrows it to .001567. Under tight caps, width 256 certifies all 12 optima with a zero gap, aided by feasibility pruning. Widening a beam is not assumed to monotonically improve every individual run; the reported trends are measured outcomes.

Table 5 Beam comparison with exact oracles; regret and gaps condition on successful runs

Budget	Width	Successes / 12	Mean regret	Mean bound gap
330s	1	12	0.02757	0.24410
330s	4	12	0.00828	0.15336
330s	16	12	0.00141	0.08713
330s	64	12	0.00000	0.03653
330s	256	12	0.00000	0.00157
tight	1	2	0.01259	0.28417
tight	4	3	0.00186	0.15736
tight	16	9	0.01024	0.11554
tight	64	12	0.00000	0.02092
tight	256	12	0.00000	0.00000

6.8 Larger collections and remaining optimization uncertainty

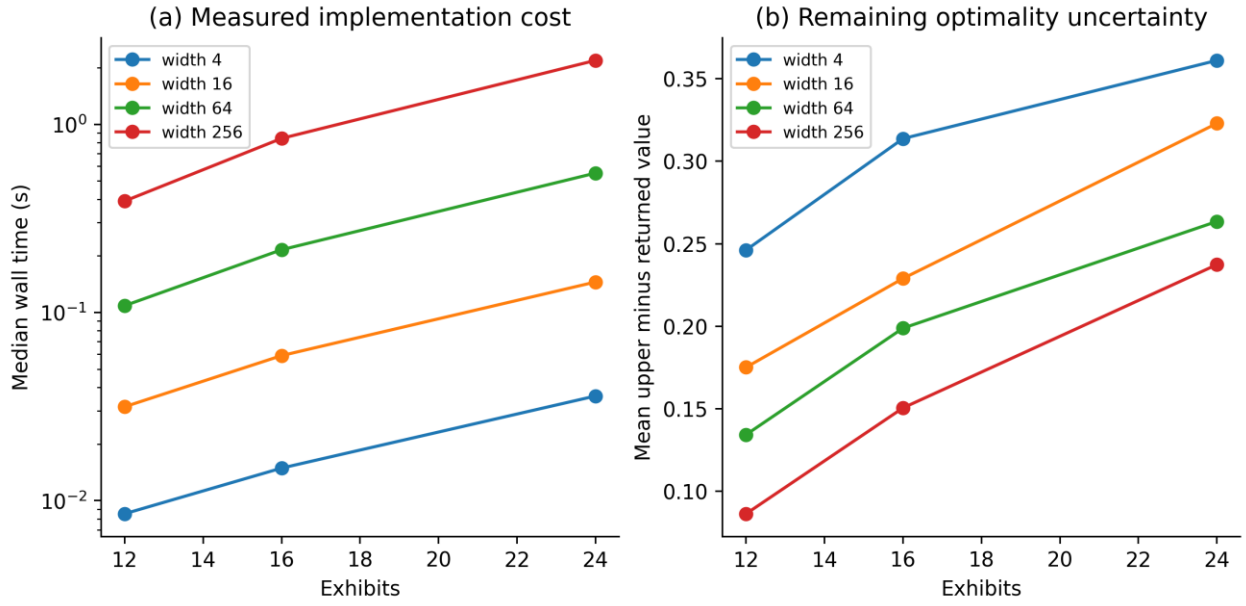

Fig. 12 Larger synthetic-gallery search. Panel (a) shows median wall time over four layouts per size; panel (b) shows the mean admissible optimality-gap bound. The 12–24 exhibit cases use edges i→i+4 and budget 40n s. No exact optimum is claimed. Timings are single-run measurements in a shared environment, not interactive VR latency. Source: beam_large.csv and large_inputs.json.

All 48 larger runs return feasible full tours. At width 256 the median times for 12, 16 and 24 exhibits are .391, .841 and 2.186 s. The corresponding mean certificate gaps are .08645, .15058 and .23736. The increasing gap is as important as the runtime: a quickly generated route can remain far from being certified optimal. The bound becomes permissive when many rewards are delayed but the bound credits them at current survival. A tighter time-aware reward bound is a concrete algorithmic research direction.

Table 6 Larger-instance results at width 256; four synthetic layouts per row

Exhibits	Median time s	Mean returned value	Mean certificate gap
12	0.391	0.6097	0.0864
16	0.841	0.5904	0.1506
24	2.186	0.5473	0.2374

These cases establish executable extension beyond eight exhibits, while preserving a visible uncertainty interval for the objective. They do not establish scalability to full museum collections, parallel users, dynamic scenes or device-constrained deployment. In particular, the n-dependent budget and longer prerequisite chains change the problem composition; cross-size score differences are not a controlled causal effect of adding exhibits.

6.9 Retrospective meaning and recommendation stability

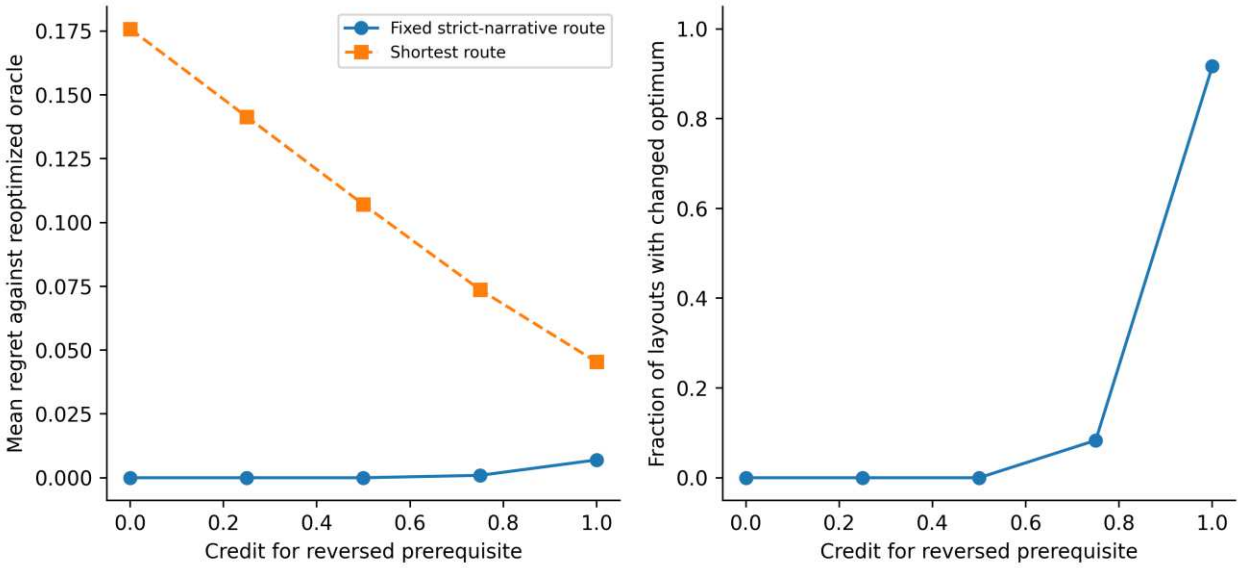

Fig. 13 Sensitivity to retrospective narrative credit. Left: the fixed strict-narrative route and the shortest route are evaluated against the exact optimum at each η. Right: fraction of the 12 layouts whose selected optimum changes from η=0. All tests use base hazards, θ=.5 and B=330 s. Source: retrospective_credit.csv; η expresses an interpretation rule, not an estimated psychological parameter.

The strict-route optimum remains unchanged for every layout through η=.5. At η=.75, one layout changes and the mean fixed-route regret is .000946. At η=1, eleven layouts change and mean regret reaches .006992. This contrasts a visible route instability with a relatively small average score penalty. A curator might care about which object appears first even where the numerical objective is nearly indifferent. Recommendation stability and objective stability should therefore be reported separately.

The mechanism is specific: increasing η releases the prohibition on reversed prerequisites while retaining the need to complete both objects. Shorter or thematically balanced alternatives can then recover pair value. This experiment directly interrogates a content assumption rather than adding another geometric rendering of the same result. It also shows why “narrative coherence” cannot be treated as a universally agreed scalar without curatorial justification.

6.10 An author supplied artwork application with traversable geometry

Identified as A01–A08. The source filenames, pixel dimensions and SHA-256 hashes are recorded in author_artworks.json and collection_asset_manifest.json. No formal artwork titles, creation dates, physical sizes or calibrated colour measurements were supplied. Study identifiers and short visual descriptions are not official titles. No public-domain or CC0 status is asserted. The virtual exhibition is a designed computational case, not a documented physical installation.

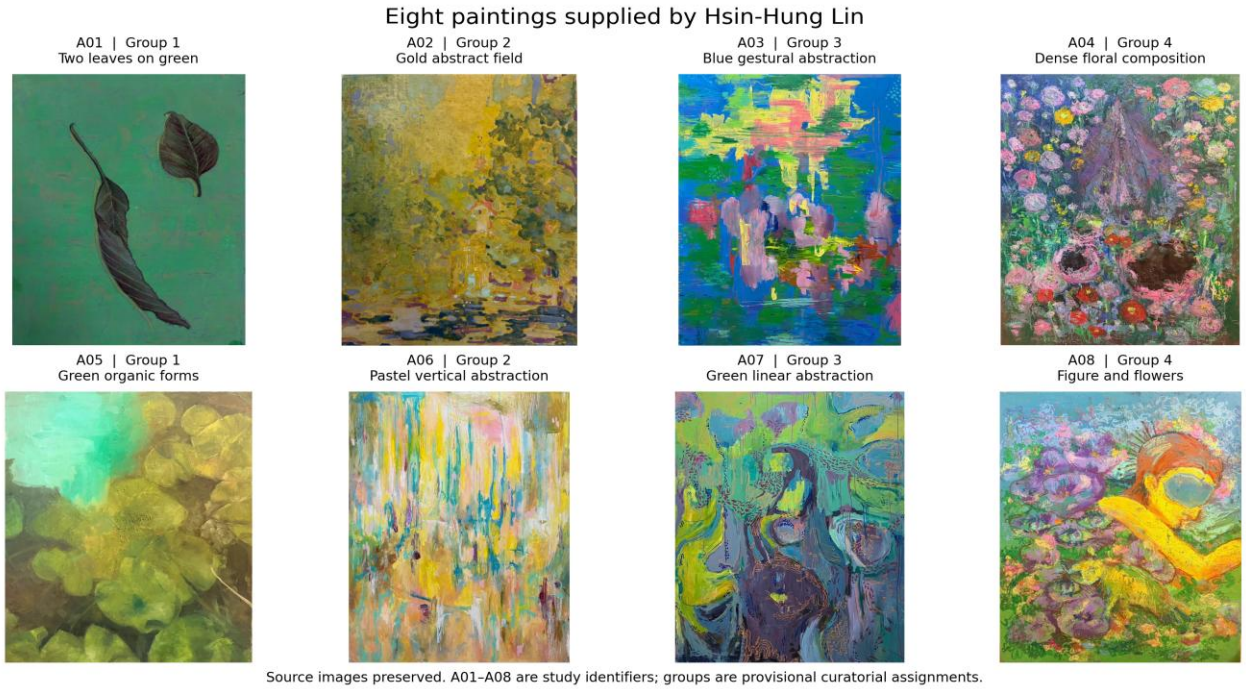

Fig. 14 Author-supplied paintings for RQ5. A01–A08 identify the exact source images. The four provisional groups pair A01/A05, A02/A06, A03/A07 and A04/A08; assignments are curatorial assumptions rather than validated labels. Source pixels are preserved; thumbnails are not physical-size measurements. Source: author_artworks.json, collection_asset_manifest.json and the eight supplied JPEG files.

The four provisional groups organize visual comparisons: botanical forms (A01/A05), layered colour fields (A02/A06), gestural abstraction (A03/A07), and floral imagery (A04/A08). These assignments describe visible features and require curatorial validation; they are neither artist categories nor independent ratings. The directed pairs 1→5, 2→6, 3→7 and 4→8 are stipulated viewing sequences, not chronology or learning prerequisites. Exposure coefficients remain .8 and base dwell 24 s, with the three uncalibrated profiles in Table 1. Pixel content is not used to estimate these quantities.

Table 7 Provenance and inferential scope of the application case

Layer	Evidence supplied	What it does not establish
Artwork assets	Eight author-supplied images, source filenames, pixel sizes and hashes	Physical sizes, dates, validated groups, viewer preferences or open licensing
Spatial model	Designed 14×10 m room; three door positions; complete graph paths	Actual museum architecture, body clearance or accessible locomotion
Behavioral model	Declared weights, speeds, dwell and interruption hazards	Estimated visitor behavior or psychological outcomes
Software	Executed canvas, callbacks, route export and independent rescoring	Headset integration, frame rate, usability or user effectiveness

The designed room has a partition at x=7 m, with a 2 m opening at y∈[2,4], [4,6] or [6,8] m. Walkable point nodes lie on a .5 m four-neighbor grid; stations and the entrance are fixed in collection_case.json. Each graph has 499 nodes. All 81 ordered station-pair paths per door are stored, and 243 paths are checked for blocked nodes and agreement between segment-summed length and the distance matrix. Refining the grid to .25 m gives 1,965 nodes and identical optimum values and routes at the base parameters in these three layouts. This aligned-grid result is not general convergence evidence or validation of continuous locomotion.

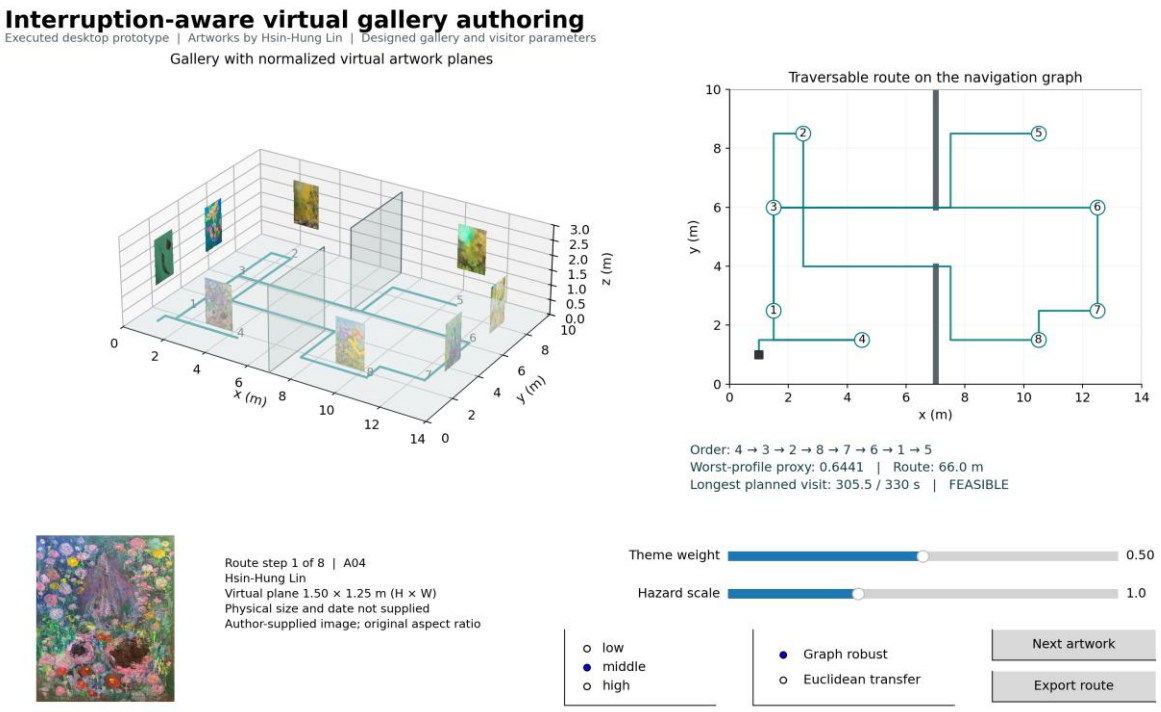

Fig. 15 Executed desktop authoring canvas for RQ5. Author-supplied artwork images appear at a stipulated height of 1.5 m with preserved pixel aspect ratio in the designed 3D gallery; the linked plan shows graph-traversable paths. Controls recompute θ, hazard scale, door location and planning model; the artwork panel follows the exported sequence. This state uses the middle door, θ=.5, base hazards and B=330 s, giving worst-profile proxy .644126 and maximum duration 305.486 s. The image is captured from the running Matplotlib canvas, not a participant photograph or headset recording. Source: prototype.py, prototype_event_log.json and the author-supplied assets.

A complete factorial design crosses three door positions, three hazard multipliers {.5,1,2}, three theme weights {.25,.5,.75} and two caps. The first cap is 330 s; the tight cap is ceil(min$_\pi$ max$_s$T$_{\pi_s}$+5 s), computed separately for each geometry. The 54 cells produce 432 method rows. They are repeated evaluations of one designed collection case, not 54 independent exhibitions or participant observations. We report conditional differences and ranges without significance tests or population confidence intervals.

Eight methods share the actual graph-based evaluation: exact robust, exact mean, exact shortest, Euclidean-transfer, beam widths 16/64/256, and eight-start single-item relocation. Euclidean-transfer optimizes on straight distances, then replays the returned sequence on the graph without repair. The relocation baseline optimizes the same robust objective and feasibility rule as beam search. It starts from nearest-neighbor ordering and seven seeded random permutations; at each pass it tests every single-item relocation, minimizing positive cap violation first and then maximizing worst-profile value. It stops at an unchanged route or 30 passes per restart. The seed is 20261103 plus condition index. It uses no exact oracle. This comparator tests algorithmic increment without confounding the reward definition; it is not a reproduction of the distinct systems in [6–8].

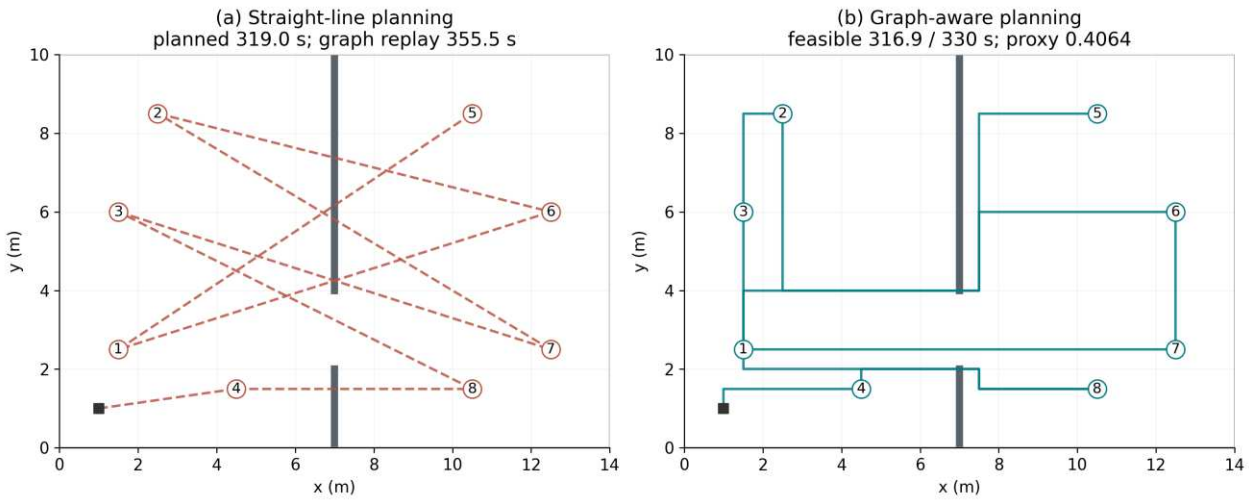

Fig. 16 Geometry-transfer failure under the low door, hazard multiplier 2, θ=.25 and B=330 s. Panel (a) draws the straight-line plan, including crossings of the partition; its time is recomputed on navigable graph paths. Panel (b) optimizes directly on the graph and respects the cap. The route overlay is a deterministic program output, not recorded movement. The figure addresses RQ5 and is generated from collection_results.csv, station coordinates and stored shortest paths.

Euclidean-transfer exceeds the actual cap in 8/27 cells at 330 s and all 27 tight-cap cells, for 35/54 failures overall. Failed overruns range from 1.77 to 31.20 s. Graph-aware methods remain feasible in all 54 cells. This does not imply every straight plan fails: 19/27 looser-cap plans remain feasible. Nor is the discrepancy attributable solely to the wall; four-neighbor directional constraints also increase distance. The result isolates the danger of transferring a budget certificate between different cost models.

Table 8 Application results against exact robust planning; mean/max regret uses only feasible cells

Method	330 s feasible	330 s mean / max	Tight feasible	Tight mean / max
Exact-robust	27/27	0.00000 / 0.00000	27/27	0.00000 / 0.00000
Exact-mean	27/27	0.01886 / 0.04199	27/27	0.00125 / 0.00849
Exact-shortest	27/27	0.02905 / 0.07379	27/27	0.00215 / 0.01248
Euclidean-transfer	19/27	0.00110 / 0.00630	0/27	N/A (no feasible route)
Beam-16	27/27	0.00074 / 0.01088	27/27	0.00065 / 0.00849
Beam-64	27/27	0.00000 / 0.00000	27/27	0.00000 / 0.00000
Beam-256	27/27	0.00000 / 0.00000	27/27	0.00000 / 0.00000
Relocation-8	27/27	0.00005 / 0.00128	27/27	0.00036 / 0.00982

At 330 s, mean-profile and shortest routing have mean robust regrets .01886 and .02905. Under the tight cap these shrink to .00125 and .00215, reflecting a more restrictive feasible set rather than better calibration. Beam widths 64 and 256 recover all 54 exact optima. Width 16 remains feasible but has maximum regret .01088. The relocation baseline is competitive: its mean regret is .00005 at 330 s and .00036 under tight caps, with maximum .00982. Thus a wider beam is not essential for useful solutions in this small case; its distinctive benefit is maintaining an explicit bound, whose tightness must be reported separately.

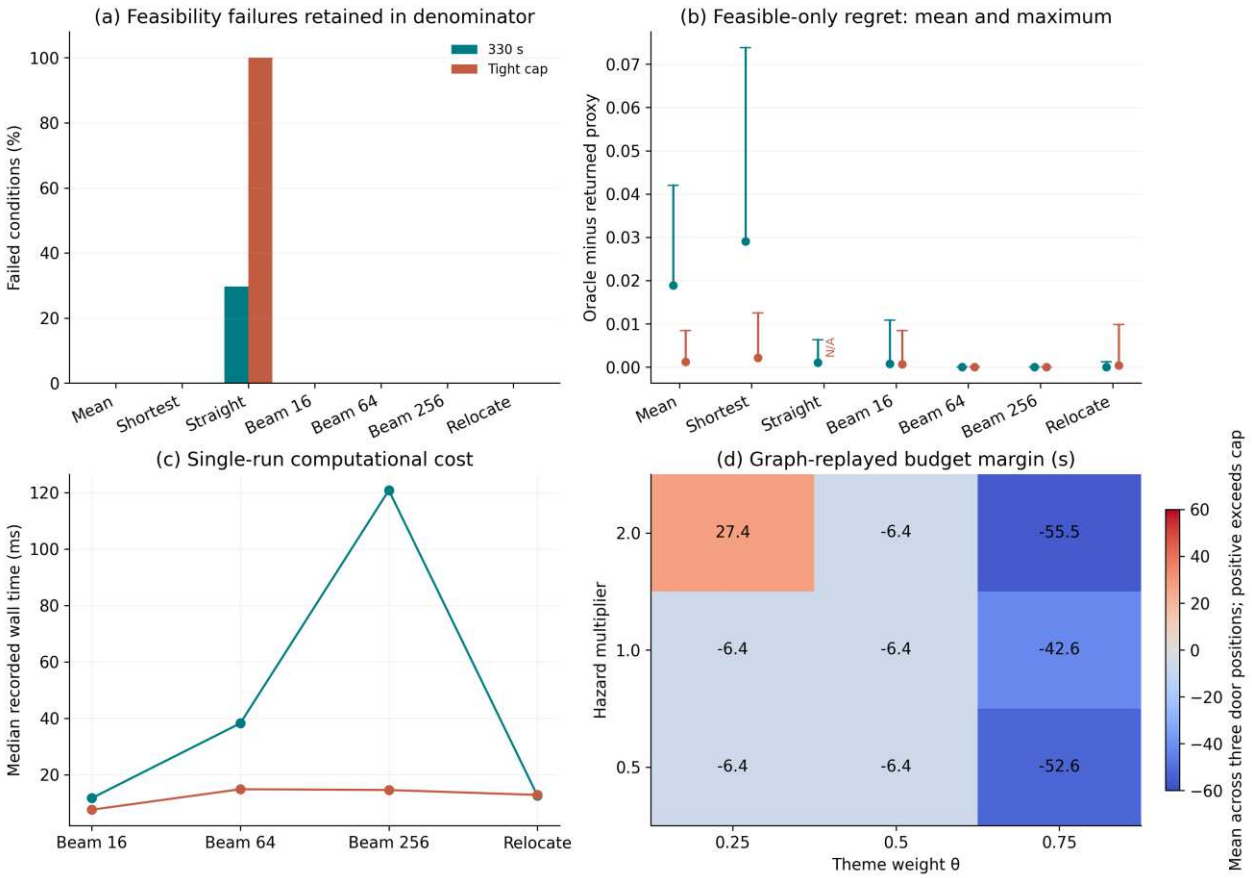

Fig. 17 Matched-evaluation comparison for RQ4–RQ5. Panel (a) retains all 27 cells per budget in the failure denominator. Panel (b) shows mean (dot) and maximum (cap) feasible-only regret; these are descriptive ranges, not confidence intervals. Panel (c) reports medians of single-run recorded search times; exact-enumeration timings were not separately recorded and are omitted. Panel (d) shows graph-replayed duration minus the 330 s cap for Euclidean-transfer, averaged over the three door variants. Positive entries exceed the cap. Source: collection_results.csv; no visitor measurements.

At 330 s, recorded median search times are 11.69, 38.20 and 120.77 ms for beam widths 16, 64 and 256, and 12.64 ms for relocation. These single-run shared-environment measurements exclude geometry construction, rendering and export; they do not establish interactive or headset latency. The 432 returned routes were independently rescored for 1,296 profile scores and 1,296 final times. Maximum discrepancies are 3.33×10^{-16} proxy units and zero seconds, respectively. Across the 27 unique geometry/hazard/weight cells, Equation (24) holds for 3,265,920 route-profile evaluations; repeated budget copies are not counted twice. Maximum fixed-route score loss from geometry transfer is .07169. These checks establish internal consistency, not external validity.

The desktop prototype implements linked 3D artwork display, route preview, theme/hazard sliders, door/method selection, next-artwork inspection, over-budget feedback and JSON export. Eight programmatic functional checks exercise these paths and compare scores/times with the independent scalar evaluator. The largest score discrepancy is 2.22×10^{-16}. The supplied adverse-state canvas records the budget warning. Verification used the noninteractive Agg backend to render the same application canvas; a GUI backend is required for manual use. No person used the interface in this evaluation, and no headset, tracked movement, performance frame counter or usability study is claimed.

7 Discussion

7.1 Implications for virtual curation

The appropriate use of the model is as an authoring audit. A curator can inspect which themes are delayed, which pairs are vulnerable to early termination, and which audience assumption makes a recommended order attractive. An interface could expose θ and profile weights alongside a route preview, rather than presenting a single opaque best route. Because the exact benchmark is small, it is especially suitable for a focused exhibition chapter or a short guided segment. It is not a complete navigation system for a large, changing collection.

The findings distinguish narrative continuity from thematic balance. An uninterrupted complete sequence rewards good ordering of all prerequisites. An interrupted visit values the timing of those rewards. Repeated-theme exhibits are not redundant in an absolute sense: they can offer additional coverage, reinforce a dependency, or be placed differently for spatial reasons. Nevertheless, the multiplicative saturation assumption embeds diminishing marginal returns and must be debated with curators. Alternative models that reward contrast, contradiction or delayed revelation could reverse the computed preferences.

7.2 What the numerical improvements establish

Exact enumeration eliminates search error at n=8. It does not eliminate specification error. The small mean-policy regret suggests that protecting a finite profile set can yield a modest gain over an already competent average-score planner, whereas comparison only with shortest routing would exaggerate the practical increment. The adverse-profile and timing tests show that even this modest guarantee has a domain. Reporting those tests is more informative than attaching visitor-like sample sizes or significance tests to deterministic scenario evaluations.

The quantity called quality in the input file is best understood as an exposure coefficient. It has no sensory units, no validated relationship to gaze duration, and no direct interpretation as learning. Similarly, the survival hazard is not a cybersickness probability. The model deliberately avoids physiological or educational benefit claims. Its strongest defensible conclusion is that specific planning criteria produce different exposure schedules under declared assumptions, and that their trade-offs can be exactly audited.

7.3 Empirical and technical validation needed next

A follow-on implementation would map actual exhibits to themes and prerequisite edges through documented curator judgments, assess agreement and contested interpretations, and record timestamps at entry, arrival, dwell completion, skipping and exit. A prospective comparison could randomize route suggestions while preserving the option to decline them. Primary outcomes should be specified independently of the optimization score, for example balanced theme recall or perceived narrative coherence, with appropriate consent, ethics determination and handling of early exits. These procedures are proposed validation steps, not completed experiments.

The application translates obstacle-aware navigation costs into an inspectable authoring workflow, but transfer to an immersive system still requires the actual locomotion method. Teleportation needs target-selection and orientation costs; continuous motion needs clearance, turning and collision handling. Display legibility, gaze direction and accessibility may affect dwell and exposure coefficients. Empirical calibration should retain a held-out exhibition or audience group rather than evaluate the planner on its fitting data.

7.4 Translating the computations into visual authoring evidence

The framework and gallery views support complementary inspections. Figure 1 locates a claim within its computation and validation stage. Figure 2 links geometry to coefficients and directed relations; Figures 3 and 8 expose route and prefix consequences; Figure 10 locates sensitivity in event timing; Figures 11–12 separate search success, objective quality and unclosed optimality gaps. This chain is more useful than rendering a

photorealistic scene that conceals arbitrary weights. Source coordinates, coefficients and route identifiers remain available for every computational panel.

The executed prototype now makes theme weight, interruption intensity, door geometry and the planning model adjustable before route export. Figure 15 documents that software state; Figures 16–17 connect it to feasibility and algorithmic evidence. This closes an implementation-demonstration gap without closing the visitor-validation gap. Published systems and perceptual work [3,10,11] require distinct measurements of rendering/interaction performance and human response; the present callback checks establish neither.

8 Limitations

The exact benchmark contains 12 synthetic eight-exhibit layouts, complemented by 12 larger synthetic layouts and one author-supplied artwork case with three designed door variants. Exhaustive enumeration remains factorial, and larger-instance beam certificates remain wide. The collection case has fixed stations, all-exhibit permutations and point-agent four-neighbor movement. Omission, revisits, replanning, crowding, wheelchair clearance and actual museum architecture are absent. Author-supplied images improve traceability and application concreteness; they do not validate themes, narrative direction, dwell, exposure or hazards. Three modeled profiles do not span audience diversity. The survival and retrospective-credit sensitivity tests relax assumptions without estimating their real-world values.

The base resampling ranges characterize a stipulated generator; application-cell ranges characterize the stated parameter grid. Neither is a population confidence interval. No new participant data, headset experiment, eye-tracking measurement, expert consensus exercise or ethics approval is claimed. The targeted literature review now includes complete-method comparisons with the three closest museum systems [6–8], but it is not systematic evidence of uniqueness. A desktop canvas and eight functional checks do not establish VR integration, accessibility or curatorial effectiveness. Those remain material limits on a submission to a visual-computing journal.

9 Conclusion

Interrupted visits turn virtual curation into a prefix-sensitive sequencing problem. Exact comparisons reveal the cost of profile protection; analytical bounds separate search uncertainty from specification error. A author-supplied artwork application and executed authoring prototype make the computation inspectable. Geometry transfer invalidates 35 of 54 planned budget certificates, while a competitive relocation baseline limits claims of algorithmic superiority. The contribution is an auditable connection between content timing, navigable geometry and design decisions. Visitor learning, presence and aesthetic benefit remain empirical questions.

Online Resource 1 contains the original and extended synthetic benchmarks, exact/beam/relocation analysis, author-supplied artwork files and provenance, prototype source and regenerated canvas outputs, route exports, tabular results including failures, literature matrix and independent checks. Artwork replacement retains the same modeled group structure, stations and coefficients. All 432 application rows were recomputed; their routes, scores, bounds and feasibility match the previous configuration, excluding variable execution times. This establishes computational invariance under asset substitution, not evidence that the new artworks evoke the same visitor response.

References

1. Huang, N., Goswami, P., Sundstedt, V., Hu, Y., Cheddad, A.: Personalized smart immersive XR environments: a systematic literature review. The Visual Computer 41, 8593–8626 (2025). https://doi.org/10.1007/s00371-025-03887-9

2. Lemonari, M., Charalambous, P., Pettré, J., Chrysanthou, Y.: CrowdImprint: decomposing context-aware interactions. The Visual Computer 42, 128 (2026). https://doi.org/10.1007/s00371-025-04329-2

3. Zhou, M., Aburumman, N., Li, Z., Raisamo, R.: VR-Deform: real-time visual and haptic interaction with deformable bodies using XPBD in virtual reality. The Visual Computer 42, 346 (2026). https://doi.org/10.1007/s00371-026-04570-3

4. Bashar, M.R., Barrera Machuca, M.D., Stuerzlinger, W., Batmaz, A.U.: The effect of visual depth on the vergence–accommodation conflict on 3D selection performance within virtual reality headsets. The Visual Computer 41, 9645–9661 (2025). https://doi.org/10.1007/s00371-025-03990-x

5. Tomasetig, G., Maravita, A., Laera, C., Pizzi, S., Paulesu, E., Sacheli, L.M.: Aesthetic experience in the physical and immersive virtual space. Virtual Reality 30, 123 (2026). https://doi.org/10.1007/s10055-026-01372-9

6. Rubio, C., Barriga, N.A., Ingram, B., Luna-García, H., Besoain, F.: Exploring Procedural Content Generation of Environments for Virtual Museums: A Mixed-Initiative Approach. Heritage 8(4), 134 (2025). https://doi.org/10.3390/heritage8040134

7. Zhu, S., Li, Y., Chen, J., Cao, X.-M.: Virtual Museum Navigation Path and Visual Interaction Optimization Based on Intelligent Space Algorithm. ISPRS International Journal of Geo-Information 15(10), 474 (2026). https://doi.org/10.3390/ijgi15100474

8. Qian, T., Zhu, S., Li, Y., Tang, Y.: Intelligent Optimization Algorithms for Model Display and Spatial Layout in Huizhou Carving Virtual Exhibition Halls. Applied Sciences 16(21), 10589 (2026). https://doi.org/10.3390/app162110589

9. Ferone, D., Festa, P., Pastore, T.: Sim-approaches for the stochastic clustered team orienteering problem. Annals of Operations Research (2026), online 9 September. https://doi.org/10.1007/s10479-026-07411-7

10. Jiménez-Navarro, D., Serrano, A., Malpica, S.: Minimally disruptive auditory cues: their impact on visual performance in virtual reality. The Visual Computer 41, 5059–5073 (2025). https://doi.org/10.1007/s00371-024-03707-6

11. Makled, E., Gerhardt, C., Schwandt, T., Weidner, F., Broll, W.: Evaluating behavioral realism in AR and VR: a comparison of single-point IK and full-body motion capture virtual humans. The Visual Computer 41, 11721–11733 (2025). https://doi.org/10.1007/s00371-025-03934-5

12. Nemhauser, G.L., Wolsey, L.A., Fisher, M.L.: An analysis of approximations for maximizing submodular set functions—I. Mathematical Programming 14, 265–294 (1978). https://doi.org/10.1007/BF01588971

Appendix A Worked computation

For the first layout, the robust route is 4–3–7–2–6–8–1–5. Table A1 reports profile-1 completion times, survival probabilities and unweighted thematic increments. To reproduce C_1, multiply each increment by the weight of its theme and its survival, then sum. To reproduce N_1, retain edges whose prerequisite appears first and add one quarter of survival at each dependent completion. Combine these using $\theta=.5$. The minimum over all three profiles, not only profile 1, produces .666733.

Table A1 Step-by-step values from worked_steps.csv

Position	Exhibit	Theme	Time s	Survival	Increment
1	4	4	43.346	0.916959	0.864776
2	3	3	71.390	0.866946	0.790476
3	7	3	97.469	0.822885	0.155771
4	2	2	120.257	0.786224	0.906260
5	6	2	155.901	0.732127	0.069761
6	8	4	201.828	0.667874	0.108878
7	1	1	240.196	0.618541	0.818125
8	5	1	272.834	0.579454	0.168713

Appendix B Worked sensitivity and search bounds

A diminishing-returns example uses coefficient .7 for the completed exhibit and .8 for the next same-theme exhibit. Its new thematic contribution is (1−.7)×.8=.24, compared with .8 if the theme were initially absent. With hazard .003 s^{-1}, the common Weibull median is ln(2)/.003=231.049 s. At t=100 s, survival is .657457, .740818 and .857686 for shapes .6, 1 and 1.8. Equality at one median therefore does not imply equal rewards at an earlier exhibit.

For the original worked robust route, profile 1 has $\partial U/\partial h = -116.974$ s and $\partial^2 U/\partial h^2 = 23577.064$ s^2. A hazard increment .0001 s^{-1} gives a first-order change of −.011697 proxy units; adding the second-order term yields approximately −.011580. This is a local Taylor approximation of the fixed route, not a reoptimized policy change. If a survival-law error were bounded by .02 at every event, Equation (17) would bound exact-policy transfer regret by .04, without implying that the actual calibration error equals .02.

For retrospective credit η=.5, a reversed pair completed at 200 s under h=.003 earns .5×exp(−.003×200)/4=.068601 narrative units before the (1−θ) mixture weight. Forward ordering earns .137203 under the same final completion time. This isolates the interpretation convention from travel and thematic effects.

Table B1 gives the admissible bounds after the worked prefix 4→3, calculated from the released inputs. At profile 1, u=.110870 and current survival=.866946; the remaining undiscounted bound is .864431. Consequently UB_1=.110870+.866946×.864431=.860284. The robust prefix upper bound is the minimum across profiles, .810992. The maximum necessary completion-time lower bound is 251.475 s. Passing the 330 s cap at this stage does not certify a feasible continuation.

Table B1 Worked prefix 4→3 from worked_bound.csv; value terms are dimensionless and time is in seconds

Profile	Prefix time	Prefix value	Survival	Value UB	Time LB
1	71.390	0.110870	0.866946	0.860284	251.475
2	61.243	0.123983	0.832162	0.824239	221.572
3	53.517	0.270725	0.807292	0.810992	196.333

The full width-64 search in this layout returns L=.666733 with global upper bound .709290 and certificate gap .042556. Exact enumeration confirms that L is already optimal. The positive gap is therefore residual conservatism of the discarded-prefix bound, not measured suboptimality. This example explains why an algorithm can empirically find an optimum without certifying it from its own search state.

Appendix C Worked geometry perturbation

For the low-door case, θ=.5 and base hazards, Euclidean planning returns 4→8→3→2→6→7→1→5. For profile 1, graph replay changes the final time from 293.935 to 316.914 s and the score from .682940 to .666080. The .016861 score loss is the sum of the eight Equation (24) terms in Table C1. The largest event delay is 22.978800 s, so the upper bound is .002×22.978800=.045958. Both score loss and bound are dimensionless. This route is still feasible at 330 s: the bound is illustrated separately from the failing condition in Figure 16.

Table C1 Independent step-by-step geometry loss for profile 1; times and delays are seconds

Step / ID	b coefficient	Euclidean time	Graph time	Delay	Loss term
1 / 4	0.060	31.451	32.114	0.664	0.000075
2 / 8	0.137	66.422	68.514	2.092	0.000501
3 / 3	0.060	107.197	114.200	7.003	0.000673
4 / 2	0.060	137.443	145.600	8.157	0.000738
5 / 6	0.137	178.569	195.571	17.003	0.003205
6 / 7	0.137	209.969	226.971	17.003	0.003010
7 / 1	0.220	252.083	269.086	17.003	0.004443
8 / 5	0.169	293.935	316.914	22.979	0.004217

For example, step 2 completes the Cézanne pair and contributes b=.5×(.2×.8×.15)+.5×.25=.137 before survival discounting. Its loss is .137×exp(−.002×66.422191)×[1−exp(−.002×2.092094)]=.000501. Every term can be recomputed from geometry_worked_steps.csv; rounded table values are for display only.

Appendix D Prospective headset validation

This appendix specifies future validation, not an experiment performed for this manuscript. No participant records, actual experimental photographs, ethics determination or headset measurements are supplied. The computational findings in the main text remain unchanged. The companion execution pack provides editable procedures and blank data templates; it is not evidence of human effectiveness.

D1 Research question and controlled comparison

The proposed question is whether robust ordering increases observed thematic exposure relative to shortest-path ordering during a 330-second visit. A two-arm parallel randomized design will avoid carryover from repeated viewing. Both arms will use the same middle-door gallery, eight artworks, positions, hardware, locomotion mode, interface and instructions. Only the recommended sequence will differ. The route exports, tie-breaking rules and software build will be frozen before recruitment. Allocation will be concealed until identical familiarization is complete, and deviations from the assigned route will be retained.

The primary outcome will be the number of distinct provisional curatorial groups with at least one completed exposure divided by four. This interpretable endpoint differs from the optimized utility and must not be interpreted as comprehension or aesthetic quality. The proposed completion rule is at least 10 accumulated seconds with an artwork in the headset field of view and within 2 m. These are operational assumptions requiring a separate pilot and freezing before confirmatory recruitment. Head direction is not eye tracking. Five- and fifteen-second thresholds may be reported as prespecified sensitivity analyses.

D2 Procedure and implementation gate

After consent and baseline discomfort assessment, participants will complete 3–5 minutes of familiarization in a separate scene, receive their randomized assignment, and visit for at most 330 seconds. Participants may stop at any time. Workload and discomfort will be recorded afterward. Recruitment through teaching relationships requires safeguards against pressure. The applicable ethics determination and consent procedure must be established before collection; no approval from another project will be reused.

The present desktop prototype is not a headset deployment. An XR implementation must first demonstrate correct scale, synchronized monotonic event timing, route cues, safe operation, export and handling of interruptions. Device model, firmware, refresh rate, engine version, build hash, frame-time distribution and tracking losses will be logged. A continuous-locomotion implementation and a teleport implementation require different timing assumptions. No headset performance claim is made here.

Actual photographs must be captured with the relevant permission and linked to the session and original file. Staged demonstrations must be labeled as such. Generated planning illustrations are excluded from experimental evidence. A photograph documents an arrangement or action; effectiveness requires the linked observations and analysis.

D3 Sample-size sensitivity and analysis

An independent equal-variance noncentral-t planning calculation uses two-sided alpha .05 and target power .80. With n participants per arm and assumed standardized difference d, the degrees of freedom are 2n−2 and noncentrality is d sqrt(n/2). These assumptions are not inferred from the computational utility. Because the exposure endpoint is discrete and bounded, the pilot distribution and the smallest meaningful difference must inform the final preregistered design.

Table D1 Prospective sample-size sensitivity computed by human_validation/analyze.py --plan. Counts are recruitment targets, not observed participants.

Assumed d	Analyzable per arm	Calculated power	Total with 15% loss
0.3	176	0.801	416
0.4	100	0.804	236
0.5	64	0.801	152
0.6	45	0.804	106

For d=.5, 64 analyzable participants per arm are required in this approximation. Allowing 15% missingness gives ceil(64/.85)=76 per arm, or 152 total. None has been recruited through this work. Changing the assumed effect to .3 increases the analyzable target to 176 per arm, illustrating the uncertainty of planning without empirical calibration.

The analysis will report the observed-case mean exposure difference, robust minus shortest, with a within-arm percentile bootstrap 95% interval using 10,000 resamples and seed 20261103. Hedges g will be supplementary and undefined when the pooled variance is zero. The discreteness, ceiling effects and possible degenerate bootstrap distributions will be considered when interpreting intervals. These quantities will not be used to claim learning, presence or aesthetic benefits.

Voluntary termination with an intact log is an observed short visit, not an exclusion. Unrecoverable log loss is missing, not zero exposure. Counts and reasons will be reported by randomized arm. Extreme bounds will assign all missing robust outcomes zero and missing shortest outcomes one for the lower bound, with reverse assignments for the upper bound. Complete-case results will not be described as a complete intention-to-treat estimate.

Secondary observations will include completed artworks, stopping time and reason, virtual path length, unweighted six-dimension NASA TLX and discomfort change. The companion program currently implements the primary exposure analysis and missing-outcome bounds; raw-event aggregation and secondary analyses await calibrated headset integration. Null or adverse results, including increased workload or early stopping, must be retained.

D4 Evidence and reporting requirements

Before reporting human results, the record must include the final protocol and ethics determination, consent and image-use scope, participant flow, original logs and hashes, actual device/build information, deviations, deidentified outcomes and reproducible outputs. Source hashes establish file identity, not authenticity. With the current empty participant file, the analysis returns NOT ANALYZED and does not generate human findings. The application therefore remains a computational and desktop-authoring study.

Protocol resources: NASA Task Load Index, https://www.nasa.gov/human-systems-integration-division/nasa-task-load-index-tlx/; SciPy bootstrap documentation, https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.bootstrap.html. Accessed 3 November 2026. These methodological resources support planning, not participant outcomes.