



Munich University of Applied Sciences

Hochschule Muenchen

Department of Computer Science & Mathematics

Fakultaet fuer Informatik und Mathematik

Quality-Driven Next-Best-View Planning for Target-Conditioned Egocentric 3D Reconstruction

Qualitätsgetriebene Next-Best-View-Planung für zielkonditionierte egozentrische 3D-Rekonstruktion

Master's Thesis / Masterarbeit

for obtaining the academic degree Master of Science (M.Sc.)

submitted to
Munich University of Applied Sciences
Department of Computer Science & Mathematics

Submitted by: Jan Duchscherer
Email: j.duchscherer@hm.edu
Program: Informatik - Master
Specialization: Visual Computing and Machine Learning
Matriculation Number: 00807819

First Examiner: Prof. Dr. Markus Friedrich
Second Examiner: Prof. Dr. David Spieler
Supervisor: Prof. Dr. Markus Friedrich
Submission Date: 7. Jan 2027
Last Updated: 22. Oct 2026

Transparency in the Use of AI Tools

Generative-AI tools supported literature-note organization, consistency checks, language revision, implementation, and figure development.

Tool	Model / Version	Date(s)	Writing role	Purpose	Scope
OpenAI ChatGPT and Codex	Session-specific OpenAI models	Thesis preparation period	Mixed	Drafting, revision, coding, and validation	Manuscript, figures, and supporting implementation

The author selected the research questions, verified sources and technical claims, implemented and evaluated the system, and remains responsible for the submitted work.

Acknowledgements

Acknowledgements are omitted from this version.

Abstract

Next-Best-View (NBV) planning addresses the challenge of autonomous viewpoint selection in active 3D reconstruction, aiming to maximize acquisition quality under a limited capture budget. We study quality-driven, target-conditioned view selection based on predicted reconstruction improvement.

A synthetic data-generation pipeline supplies training experience from egocentric indoor trajectories in Aria Synthetic Environments (ASE). It generates a new finite set of target-relative camera poses at each rollout step, renders candidate depth observations from ground-truth meshes, and measures their effect on target reconstruction. The training reward adapts Relative Reconstruction Improvement (RRI) by expressing each reduction in point-mesh error relative to the target’s initial error, so summing the rewards gives a return measuring improvement across the acquisition sequence.

The finite-horizon scorer predicts the expected cumulative improvement from acquiring a candidate pose first and then selecting further views over a specified horizon. Its inputs comprise previously acquired observations, candidate poses, and the remaining acquisition budget. We evaluate how pretrained spatial features from Egocentric Foundation Model 3D (EFM3D) and other available modalities help predict reconstruction improvement. A proposed volumetric representation combines these historical features with geometry from newly acquired depth observations, accommodating the different modalities available for recorded and generated acquisitions.

The evaluation design examines generated training experience, compares representations and learning methods, and tests both predicted values and reconstruction achieved through scorer-driven view selection on unseen scenes. Improved reconstruction on unseen scenes has not yet been established.

Zusammenfassung

Next-Best-View-Planung adressiert die autonome Wahl von Kamerastandpunkten für die aktive 3D-Rekonstruktion und soll bei begrenztem Aufnahmebudget die Aufnahmequalität maximieren. Wir untersuchen qualitätsorientierte, zielkonditionierte Ansichtsauswahl auf Grundlage der vorhergesagten Rekonstruktionsverbesserung.

Eine synthetische Datenerzeugungspipeline liefert Lernerfahrung aus egozentrischen Aufnahmesequenzen in den Innenräumen von ASE. Sie erzeugt bei jedem Rollout-Schritt eine neue endliche Menge zielrelativer Kameraposen, rendert Tiefenbilder aus Ground-Truth-Meshes und misst deren Wirkung auf die Zielrekonstruktion. Die Trainingsbelohnung baut auf RRI auf und drückt jede Verringerung des Punkt-Mesh-Fehlers relativ zum anfänglichen Fehler des Zielobjekts aus, sodass die Summe der Belohnungen den Return ergibt, der die Verbesserung über die Aufnahmesequenz misst.

Das Modell mit endlichem Planungshorizont sagt die erwartete kumulierte Verbesserung voraus, wenn zunächst eine Kandidatenpose erfasst wird und anschließend weitere Ansichten über einen festgelegten Horizont gewählt werden. Seine Eingaben umfassen bereits erfasste Beobachtungen, die Kandidatenposen und das verbleibende Aufnahmebudget. Der Repräsentationsvergleich prüft, ob vortrainierte räumliche Merkmale aus EFM3D die Vorhersage der Rekonstruktionsverbesserung unterstützen. Eine vorgeschlagene volumetrische Repräsentation verbindet diese historischen Merkmale mit Geometrie aus neu erfassten Tiefenbeobachtungen und berücksichtigt so die unterschiedlichen verfügbaren Modalitäten aufgezeichneter und generierter Aufnahmen.

Der Versuchsaufbau untersucht die erzeugte Lernerfahrung, vergleicht Repräsentationen und Lernmethoden und prüft sowohl die vorhergesagten Werte als auch die Rekonstruktion durch modellgesteuerte Ansichtsauswahl in unbekannten Szenen. Eine verbesserte Rekonstruktion auf unbekannten Szenen ist noch nicht belegt.

Contents

List of Figures	32
List of Tables	35
1 Introduction	1
1.1 Learning from simulated acquisitions	1
1.2 Contributions and evidence	2
1.3 Research Questions	3
1.3.1 RQ1 — Data generation and measurement	3
1.3.2 RQ2 — Representations and scorer design	4
1.3.3 RQ3 — Multi-step value learning	5
1.3.4 Scope	5
1.4 Thesis structure	6
2 Foundations and Related Work	7
2.1 Choosing the Next Observation	7
2.2 Reconstruction Objectives and Target Scope	7
2.3 Candidate-View Support and Motion Feasibility	9
2.4 Sequential Decision-Making under Partial Observability	12
2.5 Representations for Egocentric NBV	18
2.6 Related-Work Synthesis and Research Gap	21
3 Oracle and Data Generation	24
3.1 Dataset and Ecosystem Choice	24
3.2 State as an Information Contract	26
3.2.1 Provenance and Access Roles	27
3.2.2 Historical Observation Substrate	28
3.2.3 Selected Counterfactual Prefix and State Profiles	30
3.2.4 Oracle Evaluation Context	31
3.2.5 Visibility, Feasibility, and Utility	32

3.3	Task and Action Construction	34
3.3.1	Target Selection	35
3.3.2	Target population and sampling priority	37
3.3.3	Observed-Target Admission	37
3.3.4	Candidate View Generation	38
3.4	Measurement and Outcomes	44
3.4.1	Operational Reconstruction Error	47
3.4.2	Reliability under the Frozen Protocol	51
3.4.3	Construct Validity	52
3.4.4	RRI, Training Reward, and Endpoint Gain	53
3.5	Collection and Proposal Diagnostics	57
3.5.1	Candidate-Support and Geometric-Coverage Diagnostics	57
3.5.2	Rollout Branch Sampling and Dataset Impact	58
3.5.3	Proposal support and behavior distribution	60
3.6	Dataset and Storage Semantics	61
3.6.1	Immutable Sources and Causal Replay	62
3.6.2	Missingness and Derived Views	64
4	Method	65
4.1	Current and Intended Actor State	68
4.1.1	Sufficiency and the local-evidence limit	71
4.2	Target, Candidate, and History Encoding	75
4.2.1	Logical data roles	77
4.2.2	Target-conditioned query	77
4.2.3	Relative candidate geometry	78
4.2.4	Strictly causal pose history	81
4.2.5	Shared state tokens	82
4.3	Candidate-to-State Interaction	83
4.3.1	Candidate independence and consistency	86
4.3.2	Geometry, information, and horizon	87
4.4	Target-Conditioned Finite-Horizon Value	88

4.4.1	Conditional scorer	88
4.5	Representation Comparisons	89
4.6	Pose-history order and observed content	95
4.7	Spatial Readout and Interaction Comparisons	95
4.7.1	Planned interaction comparisons	96
4.8	Finite Action and Causal Replay	96
4.8.1	Immediate labels and one factual successor	97
4.8.2	Current replay transition	99
4.8.3	Scientific target transition	100
4.9	Value Definition and Learning	101
4.9.1	Ideal value and learned representation	101
4.9.2	Scalar requested horizon	104
4.9.3	Direct continuous objective	105
4.9.4	Selected decoder and matched control	106
4.9.5	Fitted-Q recursion	107
4.9.6	Checking the two-step training target	110
4.9.6.1	Oracle continuation and actor-achievable value	111
5	Experimental Design	113
5.1	Study Population and Evidence Gates	113
5.1.1	Candidate support and paired profile comparison	114
5.1.2	Task population and information available to each policy	115
5.1.3	Representation commissioning before horizon claims	115
5.1.4	Oracle opportunity and the learned comparison	117
5.2	Replay Eligibility and Finite-Horizon Learning Gate	118
5.2.1	Prediction support and independent backup checks	121
5.2.2	DDQN-style and behavior-return controls	122
5.2.3	Horizon-balanced replay and evaluation	124
5.3	Policy Comparison and Statistical Protocol	126
5.4	Distribution and support diagnostics	129
5.4.1	Target exposure under a fixed evaluation population	129

5.4.2	Proposal opportunity versus selection regret	129
5.4.3	Ranking, prediction error and selected-action regret	131
5.4.4	Collection coverage and the information boundary	132
5.4.5	Conditional capacity check	133
5.5	Interpreting the optional oracle comparison	133
6	Results	135
6.1	Does candidate cropping preserve the measured transition?	136
6.1.1	Bounded generation-runtime attribution	137
6.2	Historical collection in context	138
6.3	What development data completed and qualified?	138
6.4	What supervision survives across families and acquisitions?	141
6.5	Did the nominated collector changes improve supervision?	145
6.6	Where do the selected cameras move?	146
6.7	How continuous are the selected poses?	148
6.8	Does the one-step learner clear the frozen practical-effect screen? . .	151
6.9	Do action choice and prediction error agree in the larger one-step analysis?	152
6.10	Did additional trajectories improve the matched Mac learner? . . .	154
6.11	What did the continuing-value comparison establish?	155
6.12	Historical proposal-support and profile diagnostics	156
6.13	Target-Task Coverage	156
6.14	Candidate-Profile Development Diagnostic	157
6.15	Outstanding evidence	157
6.15.1	Representation and selected-observation comparisons	157
6.15.2	Independently evaluated finite-horizon predictions	157
6.15.3	Policy comparison	157
6.16	Confirmatory Report Interface	158
7	Discussion	159
7.1	What the available evidence establishes	159
7.2	What the unresolved comparisons must distinguish	161

7.3 Data collection before action-space replacement	161
8 Conclusion	164
Reader-state ledger	166
Maintenance contract	170
Development roadmap	171
Author review and professor feedback	171
TL;DR	172
Critical path	172
Milestones	173
Blockers	173
Promotion queue	174
Maintenance contract	174
HM/FK07 release gate	175
Freeze and submission	175
9 Development Pilots	176
9.1 Candidate-Family Support	176
9.2 Target-Frame $\mathcal{S}^2$ Diagnostics	177
Development Register: Method Alternatives	179
Target exposure and horizon-weighting alternatives	179
9.3 Ordered Pose-History Control	180
9.4 Optional local target shell	182
9.5 Current admission and proposed allocation: integrated illustration .	182
9.6 Failure-driven representation ladder	185
9.7 Conditional decoder comparison	185
9.8 Optional capacity and data pilot	186
9.8.1 Capacity and data pilot	186
State and history alternatives	186
Interaction alternatives	187
Objective and estimator alternatives	190
Ranked scientific-core work	192

Additional conditional hypotheses	194
Optional target-allocation rationale	195
9.8.2 Reliability and causal observation support	195
9.8.3 Training allocation without changing the evaluation question	196
Bibliography	197
A Reproducibility Record	204
A.1 Bounded Phase-A experiment contract	206
B Development Diary	208
B.1 Decisions Retained	208
B.2 Failures That Changed the Plan	209
B.3 Rejected Scope	209
B.4 Directional-Memory Hypothesis	209
B.5 Next Evidence Gates	210
C Architecture Atlas	211
C.1 Point–mesh metric validity	218
C.2 Recorded Replay Compatibility	219
C.3 Generation re-analysis contract correspondence	219
C.4 Replay lineage and stored relations	220
C.5 Replay batching and backup validation	221
C.5.1 Training unit, causal batching, and prediction support	221
C.5.2 Scalar requested-horizon targets	222
C.6 Expanded replay update and seed lineage	224
C.7 Measurement and evaluation details	225
C.8 Applied sampler construction	225
C.8.1 Target population and sampling priority	227
C.8.2 Sibling-heading guard	228
C.9 Historical pruning controls	229
C.10 Shared label and commit procedure	230
C.10.1 Ordered fusion and point capping	230
C.11 Alternative reconstruction measurements	231

C.12	Secondary proportional-error diagnostic	231
C.13	Secondary proposal geometry diagnostics	232
C.13.1	Candidate-support populations and aggregation	233
C.14	Jitter compliance	236
C.15	Ranking, association and selected-action statistics	237
C.15.1	Ranking and calibration do not certify the selected maximum	237
C.15.2	Policy advantage is a different quantity	238
C.16	Extended literature comparison	239
C.17	Bounded runtime attribution	239
C.18	Selected-camera diagnostic definitions	240
C.18.1	Target-centred positional occupancy	241
C.18.2	Signed angular increments and whole-chain quantities	242
C.19	Historical candidate-profile diagnostics	244
C.19.1	Radius and motion-bound alignment	244
C.19.2	Three-scene development confirmation	245
C.20	Historical collection coverage	245
C.21	Historical no-render proposal audit	247

Glossary and Abbreviations

Dataset

ASE – Aria Synthetic Environments: Synthetic indoor dataset with simulated Aria observations, trajectories, and scene annotations. The thesis uses ASE reference meshes and rendered depth to measure target reconstruction improvement. Semidense actor evidence is a separate input, not the default oracle reconstruction source.

MPS – Machine Perception Services: Project Aria services that derive poses, mapping, and other perception outputs from recordings. MPS-style poses and semidense points describe available sensing products. They should not be confused with the ASE-depth reconstruction used by the current oracle evaluation.

Project Aria: Meta's egocentric research-device and tooling ecosystem for calibrated, time-aligned multimodal sensing. ARIA-NBV treats Project Aria and its MPS-style products as the actor-visible sensing contract, while ASE meshes remain offline supervision and evaluation assets.

Snippet: Short synchronized temporal window of Aria sensor data used as one EVL/VIN input sample. A snippet typically contains RGB or grayscale streams, poses, calibration, semi-dense points, and scene metadata that EVL lifts into a voxel grid.

Entity

Target of Interest: Object whose reconstruction is to be improved during a rollout. The requested object remains fixed while camera poses and observations change. The actor receives an observed or predicted target descriptor; the oracle uses matched reference geometry to measure improvement on that object.

Geometry

***DoF* – Degrees of Freedom:** Number of independent parameters describing a pose or permitted motion. A rigid camera pose has six degrees of freedom: three translation and three rotation. Candidate generation with fixed roll uses a five-degree-of-freedom parameterization while still producing rigid poses in $SE(3)$.

***Frustum*:** Camera viewing volume bounded by the field of view and near and far depth limits. For the pinhole camera model it is a truncated pyramid. Inclusion permits projection into the image but does not imply visibility: other surfaces can occlude a point.

***LUF* – Left-Up-Forward:** Camera coordinate convention whose x axis points left, y axis points up, and z axis points forward. LUF is one of the coordinate-frame conventions that must stay explicit when moving between Project Aria cameras, PyTorch3D cameras, and ARIA-NBV candidate poses.

***OBB* – Oriented Bounding Box:** Box described by a center, orientation, and three extents. It approximates an object extent and defines a local target frame. Semantic class and confidence can accompany the detection but are not geometric properties of the box itself.

Metrics

***CR* – Coverage Ratio:** Fraction of a specified target surface or region treated as observed under a stated visibility or distance rule. Coverage is a diagnostic here. The primary reconstruction objective uses root-normalized error reduction, whose value also depends on the accuracy of reconstructed geometry.

***GT* – Ground Truth:** Reference geometry or annotations used to define supervision and evaluation. In this thesis, ASE meshes and object annotations provide this reference. Their reference role does not make them actor-visible, and the resulting measurements still depend on rendering, sampling, and reconstruction-error conventions.

Policy-Relative Endpoint-Error Reduction: Reduction in endpoint reconstruction error relative to a named comparison policy. Both policies use the same target task, acquisition budget, reference geometry and reconstruction preparation. Divide the reference-policy error minus the evaluated-policy error by the reference-policy error; the ratio is undefined at zero reference error and unstable near zero.

Target Reconstruction Error: Discrepancy between prepared reconstructed geometry and the fixed reference target mesh. The implemented point-mesh error sums squared point-to-mesh accuracy and equal-face mesh-to-point completeness terms, in square metres. Equal-face weighting depends on mesh tessellation. Error reduction measures reconstruction improvement.

RRI – Relative Reconstruction Improvement: Reduction in reconstruction error relative to the error before one acquisition. Target-specific RRI measures this change on the selected object. Oracle RRI obtains the measurement from privileged reference geometry. Unlike the root-gain reward, its denominator changes with the current reconstruction.

Root-Normalized Endpoint Gain: Reduction from initial to endpoint target error, divided by the initial error plus a stabilizer. For complete, consistently measured chains it equals the undiscounted sum of root-normalized rewards. Relative endpoint-error reduction against another policy uses a different denominator.

Root-Normalized Reward: Signed target-error reduction from one acquisition, divided by the fixed initial target error plus a stabilizer. For a complete chain measured consistently, its undiscounted sum equals the endpoint gain on that fixed scale. State-relative RRI instead divides by the error immediately before each acquisition.

Model

EFM3D – Egocentric Foundation Model 3D: Egocentric 3D perception framework used to extract spatial evidence from observations. Its EVL

architecture supplies frozen features and geometric predictions to the candidate-value model; the perception backbone and the learned planning head serve different roles.

EVL – Egocentric Voxel Lifting: Architecture that lifts synchronized egocentric image evidence into a gravity-aligned voxel feature volume. EVL belongs to the EFM3D framework. Its features and prediction heads provide observed spatial evidence; they do not by themselves supply observations from unvisited candidate poses.

Q_H – Finite-Horizon Q Function: Expected finite-horizon return after taking a specified candidate action first. The continuation policy determines later choices. The ideal optimal value maximizes over policies using the admitted actor information; a learned scorer approximates it from a finite representation. A sampled return, the fitted training target and the predicted value are distinct. The Q function specifies neither encoder architecture nor training algorithm.

Target-Conditioned Scorer: Model that predicts candidate utility from scene evidence, a proposed camera pose, and a target descriptor. The predicted quantity can be immediate target reconstruction improvement or return over a specified number of future acquisitions.

Value Calibration: Agreement of numerical return predictions with the expected return specified by a declared continuation policy and reference population. It requires suitable independent outcome evidence. Candidate ranking or useful policy selection can succeed without accurate numerical values; a privileged search reference need not equal the actor-achievable expectation.

VIN – View Introspection Network: Learned model that estimates the utility of a proposed view before acquiring its observation. The one-step baseline predicts reconstruction improvement from available evidence; the finite-horizon extension predicts return after selecting the candidate and continuing for the requested horizon.

Planning

Acquisition Budget: Limit on the number of acquisitions allowed for a task.

The remaining budget counts acquisitions still available. Asking for a shorter prediction horizon does not change that budget.

Acquisition Cost: Resources consumed by acquiring observations. The core comparison counts acquired views. An acquisition budget limits expenditure; it is not the cost already incurred. Motion distance and elapsed time are separate possible cost measures.

Action Support: Candidate row indices admitted for selection under the declared action-validity rules. Training-evidence support can be narrower because some admitted actions lack usable labels or recorded successors. A materialized row, an admissible action and a supported learning target are different predicates.

Actor Information: Information permitted to influence action selection under a specified observation protocol. Observation and target-source profiles specify the admitted historical evidence, earlier selected observations, and target descriptors. The oracle uses reference geometry to evaluate outcomes; exposing that geometry to a privileged control actor does not make it sensor-observed.

Candidate Table: Finite collection of proposed camera poses and their aligned row data at one decision. A candidate shell need not lie on a sphere. The admissible action set contains only the row indices retained by the validity mask.

Candidate View: Proposed rigid camera pose that could be selected for the next acquisition. The pose belongs to $SE(3)$; the action is its integer index in the current candidate table. Selection makes that prospective pose the next realized pose. Its utility may include immediate improvement or longer-horizon return.

Collection Policy: Selection rule used to choose the acquisitions recorded for training. It determines which candidate observations become selected

successors. A scorer predicts values; a policy chooses actions, potentially using those values. The collection policy may use privileged oracle scores and need not match the learned policy evaluated later.

Continuation Policy: Rule for choosing later actions after a specified first acquisition. Its information access and action choices determine which expected return a candidate value describes; the collection policy need not be the continuation being learned.

selected-action rollout transition: Update obtained by selecting a candidate action and advancing a simulated or replayed trajectory. The selected pose becomes the current pose, history and remaining budget advance, and a new candidate table is generated. Observation updates depend on the state profile: pose-only replay retains root evidence, whereas selected-observation profiles also incorporate the admitted measurements. Oracle reconstruction updates are evaluated separately. Factual identifies the selected recorded branch in this simulated process, not physical deployment.

Finite Candidate Action Set: Set of admissible integer indices into the candidate-pose table at a decision state. Each row proposes a rigid camera pose; choosing an index selects that pose. The action mask excludes inadmissible rows independently of their predicted utility.

Finite-Horizon Return: Sum of acquisition rewards over a specified prediction horizon, with discounting when specified. The core objective sums signed root-normalized rewards without discounting. A recorded return describes one realized continuation; an action value is an expected return under a specified continuation rule.

***S1-surface* – Selected-Surface Privileged Control:** Privileged control that augments root evidence with surface measurements rendered at previously selected camera poses. The implemented S1 control retains selected depth with camera calibration, poses, masks, and provenance. It reveals additional geometry to the control actor but does not establish that the same observations would be available to a deployable actor.

Historical Observations: Observations available from the logged trajectory before a counterfactual rollout begins. Images, poses, calibration, semidense points, and derived features form the initial evidence where available. Later selected synthetic observations extend this history only under profiles that admit them; ground-truth meshes and boxes remain separate oracle inputs.

Implemented Oracle: Concrete privileged evaluator and collection or search procedure used to generate labels and trajectories. It may evaluate many immediate candidates while retaining one or several continuations. Its policy, proposal support, search depth and computational budget determine what it explores; privilege alone does not make it optimal.

Causal Observation State: Proposed actor representation that updates from observations at selected poses while retaining relevant historical evidence. It should distinguish observed surface, observed free space, and unknown space, together with measurement support and uncertainty. This is a design objective, not a claim of proven minimality or Markov sufficiency.

NBV – Next-Best-View: Selection of a sensor viewpoint to improve reconstruction under a limited acquisition budget. A myopic selector ranks immediate improvement; a finite-horizon planner also values views that enable useful later acquisitions.

Operational Feasibility Teacher: Recorded action-mask labels on materialized candidate rows used to supervise an auxiliary feasibility prediction. The teacher reflects the operational validity decisions recorded by the pipeline; it does not certify complete geometric safety. Teacher support is distinct from value-label and successor support.

Optimal Oracle Value: The highest expected finite-horizon return achievable by a continuation policy admitted to privileged oracle information under the declared decision protocol. The candidate-generation and execution rules still constrain actions. This privileged optimum differs from an actor-information optimum and is not established by a bounded search or one recorded trajectory.

Oracle Evaluation Context: Privileged information used to render hypothetical observations and measure their reconstruction outcomes. Reference meshes, target crops, and candidate labels support supervision and bounded oracle lookahead. The oracle is distinct from actor information, and a bounded comparator need not attain the globally optimal return.

***S0-pose* – Pose-Only Counterfactual State:** Planning representation that retains initial observations and records subsequently selected camera poses and remaining budget. Selecting another pose changes the pose history and available candidates but adds no new surface observations to the actor representation.

Prediction Horizon: Number of future acquisition rewards included in a value query. It includes the candidate acquired first and is bounded by the remaining acquisition budget; recorded trajectory length is a separate limit.

Recording Cutoff: End of data recording before the underlying acquisition task has necessarily ended. Later rewards are unobserved, not known to be zero. A completed prediction horizon and genuine task termination have different meanings.

Rollout Root: Initial decision state from which a rollout and its comparison budget begin. It contains the historical observations available up to the reference pose. The root reconstruction error fixes the normalization for root-gain rewards; the current camera pose subsequently changes while the root remains fixed.

Rollout State: Representation of the information carried from one planning decision to the next. The rollout starts from a root with historical observations. Each selected action advances the camera pose, pose history, and budget; the chosen observation profile determines which new measurements also enter the actor representation. This representation is distinct from the physical environment and the privileged oracle context.

Sampled Return: Realized sum of acquisition rewards along one trajectory over a stated horizon. It is one outcome under the named continuation policy,

not that policy's expected value or the best achievable value. Shared-prefix branches are dependent, and retaining only the largest observed returns changes their sampling interpretation.

NBV MDP – Target-Conditioned NBV MDP: Decision-process formulation for target-conditioned acquisition from generated candidate views. The acquisition protocol specifies observations, admissible actions, transitions and rewards. Finite-horizon queries value a bounded number of acquisitions; a recording cutoff need not end the underlying task. A finite actor representation is not automatically a sufficient Markov state.

Validity Mask: Boolean indicator whose true entries identify admissible candidate actions. It constrains selection separately from utility prediction. An invalid-action mask has the opposite polarity. Label availability and stored-successor availability require separate indicators.

Reconstruction

PC – Point Cloud: Collection of sampled three-dimensional points representing geometry. Point coordinates may coincide while remaining distinct samples. The oracle concatenates observation clouds before shared fusion, cropping, and capping; treating this operation as set union would erase multiplicity.

SLAM – Simultaneous Localization and Mapping: Joint estimation of sensor motion and a map from observations. Logged poses and semidense mapping products provide actor-side evidence in this thesis. The default oracle reconstruction measurement instead uses ASE rendered depth and reference meshes.

VIO – Visual-Inertial Odometry: Pose-estimation method that combines visual measurements from cameras with inertial measurements from IMUs. Project Aria pose and mapping products build on visual-inertial estimation, and ARIA-NBV depends on those pose streams for snippet state and candidate frame contracts.

Representation

Composite Support Mask: Voxelwise indicator that at least one of the specified source-support conditions is satisfied. The implemented volume combines root feature-projection and occupancy inputs with semidense and selected-depth votes. A true entry alone does not establish measured surface or free space.

Descriptor: Quantity or collection of quantities describing relevant properties of a scene, target, history or candidate. Descriptors may summarize geometry, semantics or acquired support. A descriptor inventory does not imply that every field is available or used; an encoding maps selected descriptors into a model representation.

Learned Encoding: Model-produced representation of selected inputs. Its numerical features depend on learned parameters. It is distinct from the descriptor inventory and from a deterministic coordinate transformation.

Observation Modality: Type of acquired measurement, such as an image or a depth map. Poses and calibration locate measurements; learned features and support summaries are derived descriptors. Availability, source and actor access must be stated separately.

Occupancy Grid: Spatial grid used to represent evidence about occupied and free space. Unknown space has not been sufficiently observed and is distinct from observed free space. A low occupancy prediction alone does not establish free-space evidence; interpretation depends on the measurement and representation model.

Representation Aliasing: Mapping acquired histories with different action consequences to the same scorer representation. This discards information already acquired. Partial observability also concerns scene information never acquired; the two limitations can coexist.

Volumetric Fusion: Integration of spatially aligned information from several observations or sources in a common voxel grid. The admitted sources and

update rule determine its meaning. Combining acquired measurements and learned features does not make every grid cell a measured surface.

Supervision

Bootstrap Target: Fitted training target combining measured immediate reward with estimated continuation value. The continuation depends on the declared horizon, action support and successor evidence. An unavailable successor does not establish zero future reward. An exact measured continuation control and a learned backup provide different references.

Label Availability: Presence of a usable measured training label for a candidate at a decision state. An admissible action may lack a label. A one-step label does not establish that its successor or full continuation was recorded.

Successor Availability: Presence of a recorded successor linked to a selected acquisition. It supports learning from that transition; longer-horizon targets also require the observations and labels needed by their continuation rule.

List of Symbols

Symbol	Meaning	Key
Oracle and Reconstruction		
RRI	Relative Reconstruction Improvement score for a candidate view.	oracle.rri
$\mathcal{P}$	Reconstruction point set; source and access role are specified by its observation or evaluation context.	oracle.points
$\mathcal{P}_q$	Candidate-view point contribution used in oracle or counterfactual updates.	oracle.points_q
$\mathcal{P}_j^{\text{sel}}$	Raw strided world-frame samples from selected acquisition j, concatenated before reconstruction preparation.	oracle.selected_points_j
$\mathcal{P}_{t,i}^{\text{acq}}$	Raw strided world-frame samples rendered for candidate i at state t, before reconstruction preparation.	oracle.acquired_points_ti
$\mathcal{Q}$	Finite candidate-view set considered by the planner.	oracle.candidates
$\mathbf{c}$	World-space camera-center vector; subscripts identify a rollout root or candidate pose.	oracle.center
D_q	Oracle-rendered or candidate-specific depth map for a proposed view.	oracle.depth_q
$D_{P \rightarrow M}$	Point-to-mesh directional error component.	oracle.dist_pm
$D_{M \rightarrow P}$	Mesh-to-point directional error component.	oracle.dist_mp
D	Aggregate point-mesh reconstruction error used by RRI definitions.	oracle.err
ASE Assets		
$\mathcal{M}^{\text{GT}}$	Ground-truth scene mesh used for oracle labels and evaluation.	ase.mesh
$\mathcal{M}_e^{\text{GT}}$	Ground-truth mesh crop for target e .	ase.mesh_target
$\mathcal{F}^{\text{GT}}$	Ground-truth mesh faces used in point-mesh distance calculations.	ase.faces
$T_{R_t}^W$	Native logged Project Aria rig pose in the world frame at time t.	ase.traj
$T_{R_T}^W$	Final rig pose used as an anchor for gravity-aligned local fields.	ase.traj_final
Entity and Target		
RRI ^e	Target-specific Relative Reconstruction Improvement for entity e.	entity.rri_e

Symbol	Meaning	Key
$RRI_{t,i}^e$	State-relative marginal target RRI for candidate i at rollout step t .	entity.target_rri_marginal
$C_t^{RRI,e}$	Running sum of selected state-relative target RRIs.	entity.target_rri_cumulative
J_t^e	Running cumulative target gain normalized by rollout-root error.	entity.target_root_gain_cumulative
Δ_t^e	Target-specific reconstruction error at rollout step t .	entity.target_error
φ_e	Target/entity descriptor inventory; selected fields can condition target-specific value prediction.	entity.target_desc
p_e^W	World-space center of the selected target entity e .	entity.center
$J_e^{(H)}$	Endpoint reconstruction gain for target e over horizon H .	entity.endpoint_gain
$J_{e,\log}^{(H)}$	Log-scale endpoint reconstruction gain for target e .	entity.log_gain
Δ_{look}	Paired normalized endpoint-gain difference between oracle lookahead and oracle greedy.	entity.lookahead_headroom
$R_e^{(H)}$	Fraction of a named reference policy's remaining endpoint error removed by the compared policy.	entity.relative_reduction
η_Q	Fraction of the learned-myopic-to-oracle-lookahead endpoint gap closed by the Q policy.	entity.q_recovery
ℓ_e	Full target OBB side lengths in target-frame x, y, z order; half-extents are one half of this vector.	entity.extents
$\mathcal{P}_t^e$	Target reconstruction point set after the shared preparation operator at decision step t .	entity.prepared_points
$\Delta_{t i}^e$	Hypothetical target reconstruction error after acquiring candidate i ; the selected outcome becomes Delta at $t+1$.	entity.candidate_error
ε_{err}	Positive additive reconstruction-error stabilizer, with the same squared-metre units as target error.	entity.error_stabilizer
Planning and Rollout		
s	Abstract planning state.	rl.s
a	Action index selecting a candidate row.	rl.a
π	Selection policy over available actions. A collection policy selects the acquisitions recorded for training; it need not match the evaluated policy.	rl.pi
r	Scalar reward or immediate gain.	rl.r

Symbol	Meaning	Key
$\mathcal{H}_t$	Actor-visible observation and action history available before decision step t .	<code>rl.history</code>
$\mathcal{J}_t^{\text{act}}$	Causal actor information: historical actor observations plus the strictly selected counterfactual action-observation prefix.	<code>rl.actor_information</code>
$\mathcal{J}_t^{\text{oracle}}$	Privileged offline-oracle information, including actor information, privileged historical references, selected privileged observations, and current candidate outcomes.	<code>rl.oracle_information</code>
$\mathcal{Y}_t^{\text{oracle}}$	Current all-candidate counterfactual observations and labels available to the offline oracle.	<code>rl.oracle_candidate_outcomes</code>
j	Factual rollout-chain index used to preserve common trajectory heritage across selected steps.	<code>rl.rollout_index</code>
Ξ	Frozen decision protocol comprising generator, target source, state map, action-support provenance, reward, discount, and horizon.	<code>rl.decision_protocol</code>
$\mathcal{M}_{\text{NBV}}$	Target-conditioned finite-candidate NBV decision process.	<code>rl.mdp_nbv</code>
z_t^σ	Representation produced from actor-visible history by state-construction protocol sigma.	<code>rl.representation</code>
φ_σ	State-construction map from actor-visible history to the scorer representation.	<code>rl.representation_map</code>
$\mathcal{A}(s_t)$	Valid action set available in state s .	<code>rl.action_set</code>
$\mathcal{T}$	State-transition operator for selected candidate actions.	<code>rl.transition</code>
$s_t^{\text{S0-pose}}$	Implemented fixed-root pose-history state used by an actor pose-history control.	<code>rl.s_pose</code>
κ_t^{oracle}	Privileged oracle evaluation context for labels and diagnostics; a bounded search result is not automatically an upper bound.	<code>rl.s_oracle</code>
$\mathbf{H}_t^{\text{pose}}$	Strictly causal prefix of previously selected poses available at decision step t .	<code>rl.selected_pose_prefix</code>
r_t^e	Target-specific immediate reward at rollout step t .	<code>rl.target_reward</code>
$G_{t,e}^{(h)}$	Return for target e from step t over requested horizon h .	<code>rl.return_h</code>
Q_H	Finite-horizon candidate-value function Q_H .	<code>rl.qh</code>
$\ell_{t,i}^{\text{feas}}$	Physical or observed feasibility logit emitted by the scorer's feasibility head.	<code>rl.feasibility_logits</code>

Symbol	Meaning	Key
$\hat{Q}_{\theta,h}^{\sigma,\Xi}$	Learned predictor evaluated as $\hat{Q}_{\theta,h}^{\sigma,\Xi}(z_t^\sigma, e, q_{t,i})$ conditional on candidate feasibility, with hard action masking outside the scorer.	rl.learned_q
γ	Return discount factor.	rl.gamma
H	Number of acquisitions in the declared fixed-horizon comparison; recorded trajectory length is separate.	rl.H
$H_{\max}$	Maximum residual horizon admitted by a scorer/data contract.	rl.H_max
h	Requested residual horizon h , constrained by $1 \leq h \leq b_t \leq H_{\max}$.	rl.requested_horizon
$m_{t,i}^{\text{act}}$	Authoritative physically valid action support.	rl.action_mask
$m_{t,i}^{Q,h}$	Availability of a finite factual value target for candidate i at requested horizon h.	rl.q_label_mask
m_t^{succ}	Availability of an actual recorded successor state for transition t; finite hard-valid backup-label support is a separate condition.	rl.successor_mask
e	Selected target entity, fixed along one rollout.	rl.target
b_t	Remaining acquisition budget at rollout step t.	rl.budget
$C(\tau)$	Acquisition cost of a selected trajectory.	rl.acquisition_cost
$\mathcal{A}_t$	Finite feasible action set at rollout step t.	rl.action_set_t
$\mathcal{Q}_t$	Indexed candidate collection $\mathcal{Q}_t$ of endpoint poses; row permutations only permute aligned metadata and scorer outputs.	rl.candidate_table
$q_{t,i}$	Candidate endpoint pose $q_{t,i}$ in SE(3) at row i and step t .	rl.candidate_qti
$Q_{h,e}^{\star,\text{oracle},\Xi}$	Ideal privileged-oracle action value: maximal expected return over oracle-admissible continuations under the declared protocol.	rl.ideal_oracle_q
π_{oracle}	Concrete implemented oracle collection or search policy; privilege and finite search do not establish optimality.	rl.oracle_policy
$G_{t,e}^{(h,\text{oracle})}$	Realized finite-horizon reward sum along one trajectory produced by the named implemented oracle policy.	rl.sampled_oracle_return
$y_t^{(h,e)}$	Fitted target for the selected action at target e and requested horizon h; continuation may be estimated.	rl.fitted_target
Coordinate Frames		

Symbol	Meaning	Key
R^{native}	Native logged Project Aria rig frame, before the fixed local-Z sampling-basis correction.	frame.r_native
R_0	Fixed rollout-root reference after the rig-basis correction.	frame.root
R_t	Corrected sampling-reference frame at decision step t.	frame.rt
C_t	Camera frame realized at decision step t.	frame.ct
$\tilde{C}_{t,i}$	Prospective camera frame of candidate pose i at decision step t; selection realizes it as the next current camera frame.	frame.cq
E_e	Target-object frame for entity e.	frame.e
Spatial Geometry		
$T_{\tilde{C}_{t,i}}^{R_t}$	Transform from candidate frame $\tilde{C}_{t,i}$ into reference frame R_t .	spatial.ref_candidate_transform
$h_{t,i}^{\text{pose}}$	Relative/local candidate pose descriptor derived from a reference pose rather than raw world coordinates.	spatial.candidate_pose_feat
$h_{t,e i}^{\text{rel}}$	Candidate-target relation descriptor in the candidate/query local frame.	spatial.candidate_target_rel_feat
$e_{a i}^{\text{rel}}$	Query-local relative positional embedding for target, history, support, or candidate relations.	spatial.relation_rpe
r_e	Geometric-mean target-proxy scale derived from the target OBB semi-axis lengths.	spatial.target_obb_scale
$\hat{\delta}_{j,t}^e$	Unit factual selected-camera displacement expressed in target-object coordinates for rollout j and step t.	spatial.target_frame_motion_direction
$\hat{v}_{j,t}^e$	Unit selected-camera forward optical axis expressed in target-object coordinates.	spatial.target_frame_view_direction
M^{dir}	Second directional moment used to summarize selected target-relative approach directions.	spatial.dir_moment
$\mathcal{F}_{j,t}^e$	Front-facing target-proxy surface directions that project inside the calibrated selected-camera image.	spatial.target_frame_frustum
$\kappa_{j,t}^e$	Fraction of the target-centred proxy sphere geometrically supported by one selected calibrated frustum.	spatial.target_frame_frustum_fraction
τ_{side}	Neutral-band tolerance for target-relative lateral coordinates in the dimensionless normalized proposal-support frame; fixed at 1e-9 in the diagnostic.	spatial.side_tolerance

Symbol	Meaning	Key
$\Omega_{j,t}^{\text{FOV}}$	Intrinsic solid angle of one calibrated selected-camera pinhole field of view.	spatial.frustum_solid_angle
$T_{\tilde{c}_{t,i}}^W$	Candidate-camera endpoint pose mapping camera coordinates to world coordinates.	spatial.candidate_endpoint
Observation and Actor State		
$F_t^{\text{DINO@pt}}$	Visibility-gated logged DINO feature bank attached to semidense or fused world points.	obs.dino_point_bank_t
X_t^{pt}	Point-token set combining geometry, support, uncertainty, history, and optional compressed logged descriptors.	obs.point_tokens_t
D	Depth observation sequence.	obs.depth
n	Per-point or per-face normal observation.	obs.face_normal
I^{rgb}	RGB image observation.	obs.img_rgb
$\mathcal{P}^{\text{cf}}$	Counterfactual point observation.	obs.points_cf
$\mathcal{P}^{\text{semi}}$	Semi-dense observed point set.	obs.points_semi
$\mathcal{P}_t$	Point set available at rollout step t .	obs.points_t
Scene Representation		
M_t^{ray}	Sparse actor-visible ray-aware memory separating occupied, free, and unknown support at step t .	scene.ray_memory_t
Φ_t^{scene}	Composite actor-visible scene-memory descriptor queried by the finite-candidate value model.	scene.scene_memory_t
$E_0^{\text{EVL-local}}$	Root local EVL evidence field used for target/candidate support reads.	scene.evl_local
$\omega_{t,i}^{\text{EVL}}$	Candidate or target-query fraction supported by the root EVL voxel extent.	scene.evl_support_frac
$g_{t,i}^{\text{EVL}}$	Pooled local EVL evidence token for a target, candidate frustum, or target-candidate intersection query.	scene.evl_support_token
g_e^{tgt}	Pooled target-support descriptor for the selected target hypothesis.	scene.target_support_pool
$g_{t,i}^{\text{fr}}$	Candidate-frustum support descriptor pooled from actor-visible scene memory.	scene.frustum_support_pool
$g_{t,e,i}^{\text{cap}}$	Pooled descriptor over the intersection between target support and candidate frustum support.	scene.target_frustum_pool
$g_{t,i}^{\text{ray}}$	Candidate-conditioned ray query over occupied, free, unknown, hit, target-support, and uncertainty summaries.	scene.ray_query_ti

Symbol	Meaning	Key
RenderQuery	Abstract query operator that summarizes scene memory in a candidate camera/frustum without introducing fresh counterfactual RGB features.	scene.render_query
M_t	Deterministic multi-channel evidence on the fixed root-aligned voxel grid; source availability and support are separate channels.	scene.evidence_volume
A_t	Boolean voxelwise composite support indicator; positive support does not by itself identify measured surface evidence.	scene.composite_support
Candidate-Value Model		
h_e^{tgt}	Learned selected-target token built from root-relative target pose and metric extents.	model.target_token
$h_{t,i}^{\text{phys}}$	Candidate-local physical token from root/current-relative candidate poses and compact root evidence.	model.candidate_physical_token
$x_{t,i}$	Candidate-local conditional-value query formed by adding candidate-from-target pose encoding to the physical token.	model.candidate_row
$p_{t,k}^{\text{hist}}$	Previously selected pose k encoded from the current camera at decision state t.	model.history_pose_feature
$a_{t,k}^{\text{hist}}$	Normalized relative age of selected pose k at decision state t; the immediate predecessor has age zero.	model.history_relative_age
h_t^{hist}	Fixed-width causal selected-pose history token supplied to scorer state fusion.	model.history_token
F_t	Learned spatial features on the evidence grid, recomputed by the volume encoder; distinct from image-feature carriers.	model.encoded_volume
VIN Scorer		
r	Target or scene RRI value used as a model target.	vin.rri
$\hat{r}$	Predicted RRI value emitted by the learned scorer.	vin.rri_hat
$\mathcal{L}$	Training loss for scorer or value-model objectives.	vin.loss
m	Candidate validity indicator used by scorer diagnostics.	vin.cand_valid
F_v	Multi-channel scene feature field assembled from EVL evidence; not a scalar action-value field.	vin.field_v
Shape and Size		
N_q	Number of candidate views in a candidate table.	shape.Nq

Symbol	Meaning	Key
K	Number of ordinal bins or discrete levels when used in scorer losses.	<code>shape.K</code>

List of Figures

Figure 1	Semidense geometric support in an ASE image.	10
Figure 2	Candidate-family geometry.	40
Figure 3	Candidate position and gaze.	42
Figure 4	Shared reconstruction preparation.	46
Figure 5	Historical measured target-error transition.	50
Figure 6	Selected transitions and bounded lookahead.	59
Figure 7	History compression and value aliasing.	66
Figure 8	From observed evidence to the compact scene carrier.	73
Figure 9	Proposed causal state composition.	74
Figure 10	Candidate-query architecture.	76
Figure 11	Physical and target-conditioned pose encoding.	80
Figure 12	Five shared contexts.	83
Figure 13	Candidate-to-state attention and decoding.	85
Figure 14	Row consistency of the independent scorer.	87
Figure 15	Implemented volumetric scorer.	91
Figure 16	Privileged selected-surface residual.	94
Figure 17	Counterfactual supervision and factual replay serve different roles. .	99
Figure 18	Evidence used by finite-horizon fitted-Q learning.	109
Figure 19	Paired full/crop lifecycle costs.	137
Figure 20	Candidate feasibility and accumulated gain.	140
Figure 21	Selected proposal families through acquisition.	140
Figure 22	Candidate-family yield and opportunity.	143
Figure 23	Supported choices across acquired prefixes.	145
Figure 24	Selected camera positions and orbital direction.	147
Figure 25	Turning frequency and angular travel.	148
Figure 26	Pose changes at each acquisition.	149
Figure 27	Movement accumulated over a chain.	150
Figure 28	One-step action choice and prediction error.	154

Figure 29	Scheduled development gates through the 30 November 2026 full-draft freeze. Arrows show planned order; policy endpoints and oracle headroom are not prerequisites for Q1 prediction.	172
Figure 30	Attempted candidate centres from a two-scene, two-state real-data method audit. Each state is centred on its factual expansion pose, yaw-aligned to the oracle-evaluation target with world Z-up, and divided by its current three-dimensional target distance. The cross marks the expansion/root pose, the stars mark the two oracle task target centres, colour denotes proposal family, and marker shape denotes actor-valid or selected status. The figure diagnoses proposal geometry under the <code>v0_gt_input</code> target protocol; it is neither a population estimate nor evidence of actor-visible target discovery or policy quality.	176
Figure 31	Target-frame $\mathcal{S}^2$ movement directions from the frozen pilot report. This view records how the camera moved; it is not a visibility measurement.	177
Figure 32	Target-frame $\mathcal{S}^2$ selected-camera optical-axis directions from the same frozen pilot report. This view records where the camera pointed; it is distinct from both camera motion and calibrated surface support.	178
Figure 33	Calibrated selected-frustum support on the target-centred proxy sphere. The surface heat map is the complete equal-solid-angle coverage field; overlaid incidence samples preserve rollout and time heritage. This is not measured target-mesh visibility.	178
Figure 34	Current admission and proposed allocation.	183
Figure 35	Development-only directional-memory fixture.	210
Figure 36	Compact scorer interfaces.	212
Figure 37	Candidate and selected-view target-proxy footprints.	213
Figure 38	Orientation-only fields of view.	214
Figure 39	Volume encoder and candidate readout.	215
Figure 40	Expanded QH computation with shared operations shown once. .	216

Figure 41	Point-mesh validity and tessellation.	218
Figure 42	Offline rollout store relation.	220
Figure 43	Applied directional sampling.	226
Figure 44	Label evaluation and selected commit.	230
Figure 45	Positional occupancy at every acquisition.	242
Figure 46	Within-chain pose-continuity summaries.	244
Figure 47	Historical candidate-family validity.	247
Figure 48	Candidate family phase a support.	248

List of Tables

Table 1	Geometric roles in the applied 60-candidate development profile. . . .	39
Table 2	Architecture roles and evidence boundaries.	67
Table 3	Independent status distinctions used in the development draft.	68
Table 4	Current baseline and scientific targets.	70
Table 5	Logical data roles at the dataset–model boundary.	77
Table 6	Research questions, comparisons and evidence.	118
Table 7	Required replay evidence by learning target.	120
Table 8	Claim-specific evidence requirements.	120
Table 9	Matched policy comparison.	127
Table 10	Chapter-wide evidence by research question.	135
Table 11	Completed development cohorts.	138
Table 12	Generation posthoc coverage.	141
Table 13	States with no training-supported candidate in the indicated family.	142
Table 14	Evidence availability in the loaded report bundle.	158
Table 15	Human review and feedback targets; dates do not certify completed review or empirical evidence.	171
Table 16	Evidence-gated schedule from the completed pilot infrastructure to submission. Dependencies are checked against the adjacent roadmap data.	173
Table 17	Conditional representation interventions.	185
Table 18	Direct and ordinal decoder controls.	185
Table 19	Development-only architectural feature-integration ladder. Status records implementation maturity, while promotion remains conditional on a named scientific failure and a one-factor comparison.	188
Table 20	Development catalogue of value-model alternatives by scientific role and estimand. Executability, selection, and evidence are distinct; activation requires a diagnosed failure and a one-factor comparison.	191

Table 21	Loaded development-fixture provenance. The fixture verifies the document interface only and is not scientific evidence.	206
Table 22	Generated Phase-A experiment contract.	208
Table 23	Recorded replay compatibility.	219
Table 24	Protocol and runtime correspondence for the bounded generation re- analysis.	220
Table 25	Scientific roles of the normalized replay relations.	221
Table 26	Alternative reconstruction measurements.	231
Table 27	Candidate-generation diagnostics and their aggregation populations.	236
Table 28	Extended comparison of related methods.	239
Table 29	Recorded historical continuation support.	246
Table 30	Candidate family phase a support.	248

1 Introduction

Active perception begins from a geometric premise: what can be reconstructed depends on what a sensing action makes visible. [NBV](#) planning chooses where a sensor should acquire its next observation to improve a reconstruction within a limited [acquisition budget](#). The choice depends on what has already been observed and what the reconstruction is intended to capture. Scanning an isolated object, reconstructing one object among occluders, and mapping an environment pose different problems [1], [2], [3]. This thesis studies a requested object within a synthetic indoor scene. Its objective is to reduce that object’s reconstruction error over several acquisitions.

The scorer receives a proposed camera pose before the corresponding image or depth observation has been acquired. It must predict the benefit of taking that observation, and of subsequent acquisitions, from the evidence already available. A good immediate view need not start the best sequence. When the allowed movement is too short to reach a revealing viewpoint directly, a preparatory move can enable a better later observation despite adding little geometry itself. This motivates testing whether multi-step predictions improve view selection in the evaluated scenes.

1.1 Learning from simulated acquisitions

[Project Aria](#) provides the wearable sensing setting: research glasses record images and sensor streams; calibration and [Machine Perception Services](#) provide camera geometry, estimated trajectories and partial three-dimensional maps [4]. [Aria Synthetic Environments](#) supplies simulated indoor observation sequences and annotations [5]. Reference meshes available through [EFM3D](#) allow an offline evaluator to render proposed observations and measure their effect on the reconstruction [6]. These measurements provide supervision for prediction before acquisition.

The actor is the system that selects a view from the information available at decision time. An [implemented oracle](#) is a privileged evaluator and collection policy that can use reference scene geometry. We use it to evaluate proposed poses

and collect trajectories. The generator determines which poses can be chosen; the [collection policy](#) determines which later states are recorded. Consequently, the training data reflects both the scene and the acquisition procedure. Evaluating many immediate alternatives still leaves unrecorded continuations. The actor uses only information admitted by the observation protocol: hypothetical candidate outcomes cannot enter its input before selection. When a simulated control receives selected reference-mesh depth, that input remains privileged.

The available types of observation ([observation modality](#)) can also change along a generated trajectory. Recorded history can contain images and pretrained features, whereas generated acquisitions may supply depth and derived geometry. [Egocentric Voxel Lifting](#) combines image features and geometric evidence in a local grid [6]. It motivates studying a shared volumetric representation of the observations acquired so far. Which scene and target descriptors improve value prediction, and whether pretrained features help beyond geometric inputs, remain empirical questions.

VIN-NBV demonstrates learned ranking by immediate [RRI](#) in object-centred reconstruction [1]. Instance-aware planning focuses an information score on a requested object in clutter [2]. GenNBV and Hestia instead predict camera actions for coverage objectives [7], [8]. ARIA-NBV asks whether a scorer can predict target-specific reconstruction improvement over several acquisitions while accounting for previously acquired observations. It ranks a finite [candidate table](#); learning to propose continuous poses remains future work.

1.2 Contributions and evidence

The study connects three contributions: a target-specific measurement and data-generation pipeline; scene and target representations that combine available observations; and finite-horizon candidate scoring with separate tests of prediction and view selection. Numerical agreement with independent outcomes tests what the predicted values mean. Comparing reconstruction after the same initial task and view budget tests whether selecting by those values is useful. A scorer can

rank candidates well while misestimating the size of their benefits; neither test alone establishes unbiased training experience.

The present draft reports generation diagnostics and bounded one-step learning results. Independent multi-step prediction and learned-policy reconstruction outcomes remain to be evaluated. Conclusions concern the declared synthetic acquisition setting; performance on unseen scenes in this setting would not by itself establish physical transfer or realistic wearer motion.

Research TODO: RRC12 — Establish the physical relevance of simulated observations, reconstruction updates and feasible motion, or limit final conclusions to the declared synthetic setting.

Source: Synthetic-to-real limitations in Foundations; EFM3D’s detection evidence does not evaluate the learned acquisition policy.

Gate: State the intended use and external-validity evidence, or retain the bounded synthetic claim.

1.3 Research Questions

This thesis asks whether a scorer can predict the expected reconstruction improvement from choosing a candidate view and then acquiring further views of a requested object. Its prediction uses observations already acquired and a specified total number of acquisitions, including the candidate view. The primary outcome is the fraction of the target’s initial reconstruction error removed over the acquisition sequence.

Two histories ending at the same camera pose can leave different cabinet surfaces unseen. A useful scorer must distinguish them when predicting whether another side view would improve reconstruction. The three research questions address the generated training experience, its representation, and learning to predict and use multi-step values.

1.3.1 RQ1 — Data generation and measurement

How do reconstruction measurement, candidate generation and the collection policy affect the reliability and coverage of the training experience?

Candidate generation proposes camera poses. Rendering supplies their synthetic observations, and reconstruction evaluation measures the resulting error reduc-

tions used as rewards. The [collection policy](#) chooses which views to acquire. Their observations contribute to the recorded training sequences and, when admitted by the model’s input design, to the evidence used to score later views.

We examine whether measurement choices change the gains and candidate rankings, and which acquisition choices and subsequent observations the generated data cover. A pipeline can produce many labels while rarely recording a move needed to reveal an occluded surface. Comparisons of measurement settings, candidate generators and collection policies assess these dependencies; RQ3 tests the resulting learned predictions and decisions on unseen scenes.

A stopped recording can leave later outcomes unobserved. [Chapter 3](#) develops the measurement and collection procedure; [Chapter 5](#) specifies its evaluation.

1.3.2 RQ2 — Representations and scorer design

How do scene and target representations, observation fusion and scorer architecture affect reconstruction-improvement prediction and candidate ranking?

Scene and target descriptors describe the observed geometry and the requested object. We study which information to retain and how to represent it so that the scorer can relate a candidate view to what remains to be reconstructed. With [volumetric fusion](#), observations are combined in a common spatial grid. Historical views may supply images and pretrained features, whereas simulated acquisitions may supply only depth and derived geometry; the model must accommodate this modality difference.

The core study updates the scorer’s input with newly acquired observations. The scorer predicts a view’s benefit before that view is acquired. Its rendered observation can supply an offline training label, but cannot enter the scorer’s input for that decision.

Representation comparisons give models the same observations and vary how they encode and combine them, for example by pooling geometric evidence or retaining its spatial arrangement. An optional pretrained-feature comparison identifies its feature sources and computational cost. Separately, we test the benefit

of adding newly acquired observations to the model input. When several model components differ, the comparison measures their combined effect. [Chapter 4](#) develops the representations, fusion operations and scorer architectures.

1.3.3 RQ3 — Multi-step value learning

How accurately can a finite-horizon scorer predict target reconstruction improvement, and does selecting views by its scores improve reconstruction on unseen scenes?

A recorded sequence shows what happened after particular collection choices. The learning aim is the expected improvement from acquiring a candidate first and then choosing later views as well as the acquired information and available candidate views permit. Foundations defines this expected value; [Section 4.9.1](#) specifies the learning aim.

Immediate rewards can be measured for many candidate views. Multi-step learning must also account for later observations and choices, although successors are recorded only for selected views. One-step prediction provides a reference for assessing what changes when learning over longer, supported horizons. [Chapter 4](#) develops finite-horizon fitted Q-learning from these transitions. Fitting recorded returns offers a possible learning-method comparison, but targets outcomes under the collection policy rather than the best admissible continuation.

Independent measurements are needed to assess the numerical predictions. We also use the learned scores to select successive views and compare the achieved reconstruction with named reference policies on unseen scenes. This tests whether learning yields useful decisions beyond fitting collected labels. Good decisions alone do not establish accurate numerical values or remove bias from data generation. [Chapter 5](#) separates these prediction and acquisition comparisons.

1.3.4 Scope

The main study uses finite generated candidate views in synthetic indoor scenes. Generalization concerns unseen scenes under fixed candidate-generation, observation and reconstruction-measurement conditions. The declared comparison

may change the view-selection policy. Any use of ground-truth targets, action constraints or observations must be identified; success in that setting does not establish real-device performance. Privileged rendered observations can support representation diagnostics but require separate interpretation from actor-visible acquisition results. A separate comparison with fixed scorer weights may isolate the effect of observation fusion. Continuous-action control and real-device transfer remain future work. Reward definitions, learning algorithms and statistical procedures are developed in the subsequent chapters.

1.4 Thesis structure

[Chapter 2](#) explains the objectives, observations and sequential values needed to formulate the problem. [Chapter 3](#) defines the target measurement and generated experience. [Chapter 4](#) constructs the scorer and its learning targets, and [Chapter 5](#) specifies how they are tested. [Chapter 6](#) presents the evidence; [Chapter 7](#) interprets its limitations and implications; [Chapter 8](#) states the contribution.

2 Foundations and Related Work

View selection couples three choices: what to reconstruct, how to measure improvement, and which observations to acquire. This chapter first separates those choices, then explains what geometric evidence reveals and how later acquisitions affect a candidate’s value. The final sections connect these requirements to representations and prior learning-based methods.

2.1 Choosing the Next Observation

A **NBV** system updates its reconstruction from acquired observations, chooses a new camera pose, acquires an observation there, and repeats. With a finite **candidate table**, these decisions separate into proposal, scoring and selection. Proposal determines the available poses; scoring compares their anticipated benefits; selection commits to one pose. PB-NBV uses this decomposition to rank proposed views from the current geometric map [9]. The scorer cannot choose a useful pose that the generator has not offered.

A policy can instead predict the camera action directly. GenNBV predicts pose parameters [7]. Hestia first predicts a look-at point and then a camera position conditioned on that point; their geometric relation determines pitch and yaw [8]. This separates where to look from where to place the camera. Its internally proposed look-at point is different from a requested object whose reconstruction defines the task.

2.2 Reconstruction Objectives and Target Scope

Views can be valued for extending surface coverage, reducing uncertainty in a reconstruction model, or lowering reconstruction error. These objectives measure different consequences of observing the same region. **Coverage** rewards newly observed surface. SCONE predicts surface-coverage gain, MACARONS learns to anticipate coverage while reconstructing online, and Hestia rewards newly visible voxel faces [3], [8], [10]. Their scores value extending what has been seen. They do not directly measure how accurately an already observed surface has been reconstructed.

Information gain values a view for reducing uncertainty in a reconstruction model. FisherRF uses the sensitivity of rendered images to model parameters to estimate how informative a candidate observation would be [11]. A view can constrain uncertain parameters without exposing much new surface. Whether that reduction in uncertainty also improves geometric accuracy depends on the model and its relation to the reconstruction metric. Next Best Sense uses depth-sensitive Fisher information during online robotic Gaussian reconstruction [12]. Updating a reconstruction after each acquisition and evaluating its depth error do not make the selection score a direct measurement of error reduction.

Reconstruction improvement measures the change in a specified reconstruction error. 3D-NVS trains a selector through supervised voxel reconstruction loss, choosing a second image from fixed object viewpoints [13]. VIN-NBV instead defines **Relative Reconstruction Improvement** from the reduction in point-cloud Chamfer distance after adding an observation. It learns to predict this improvement for sampled poses before acquiring their images [1]. The supervised quantity therefore concerns the reconstruction outcome rather than coverage or parameter uncertainty. This thesis uses that principle with a target-specific point-mesh error; the two measurement operators are not interchangeable.

The **reconstruction scope** is a separate choice. VIN-NBV and 3D-NVS evaluate object-centred capture. Hestia’s main benchmarks also use individual objects; its predicted look-at location is not a requested object identity [1], [8], [13]. MACARONS addresses large environments [3]. Object-Aware Guidance selects objects to scan within an indoor scene, using retrieved shape hypotheses to guide local views [14]. Jeong et al. instead condition information gain on a requested instance among cluttered objects [2]. Their shared map retains the other objects, but training masks remove background; the reported target conditioning also depends on supplied instance masks. These settings test different target-selection and surrounding-scene assumptions.

Target-conditioned reconstruction requires a way to specify the requested instance and associate its observations across views [2]. A supplied object label

or reference box and an object detected from acquired images provide different information. [Chapter 3](#) separates these target sources and defines how the target is maintained during collection. Surrounding geometry remains relevant: it determines occlusion and reachable camera poses even where it contributes no target reconstruction error. The next section explains the geometric evidence needed to reason about those views.

2.3 Candidate-View Support and Motion Feasibility

A camera pose locates and orients a camera in three-dimensional space; calibration relates image locations to viewing rays. Posed images, reconstructed points and depth maps provide different evidence along those rays. A point projecting inside the image can still be hidden behind a nearer surface. [Figure 1](#) shows semidense points projected into a recorded image. Their sparse support differs from a dense depth map: absence of a point does not identify an empty part of the scene [\[6\]](#).

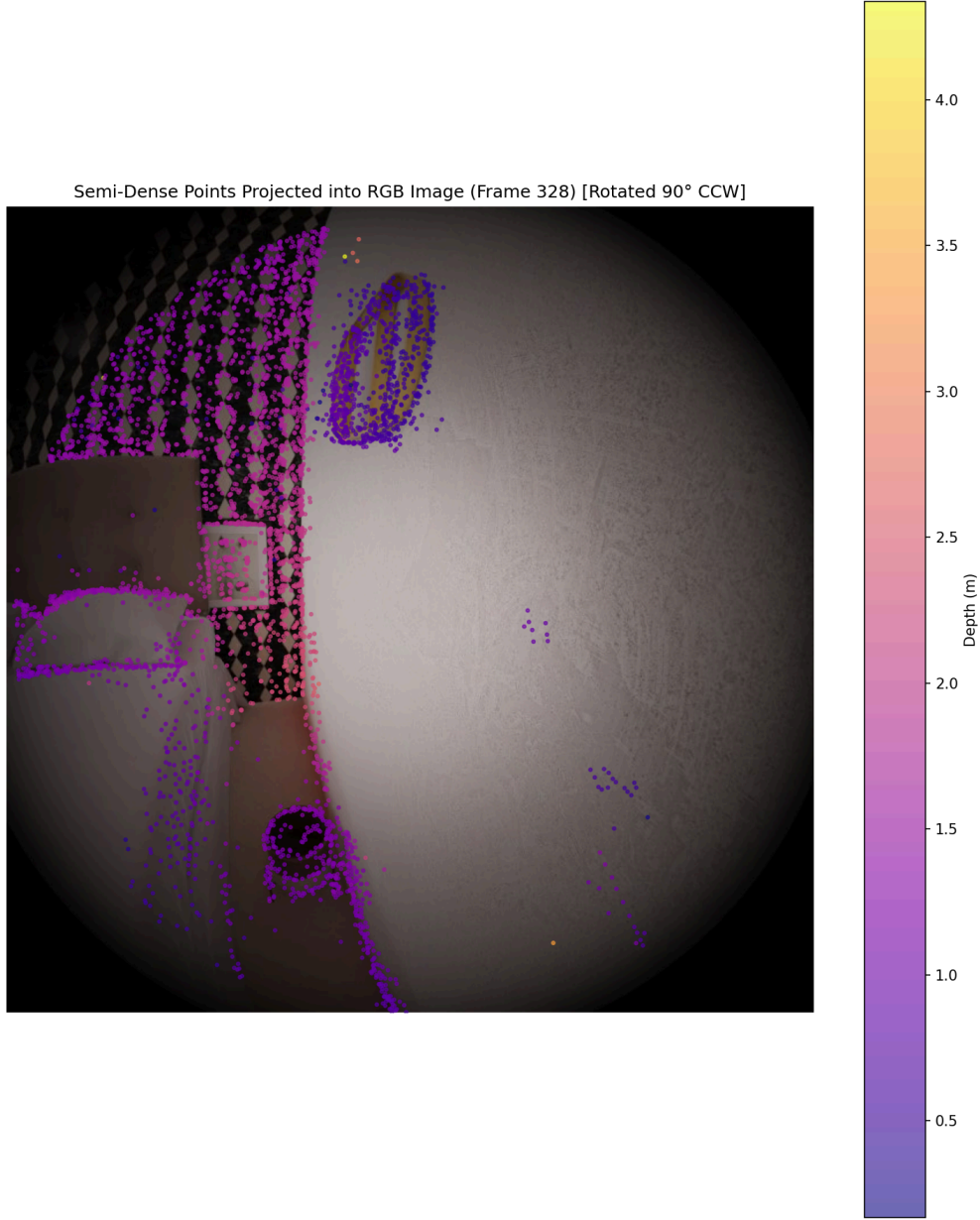


Figure 1: Historical ASE illustration of semidense points projected into an RGB frame. Point colour encodes depth according to the image’s scale. This illustrates partial geometric support, not dense depth or a measurement of reconstruction accuracy.

A depth measurement places a surface along a viewing ray. Under an ideal opaque-surface model, the segment between the camera and that surface is free; space behind it remains unknown from this measurement. EFM3D uses calibrated rays to semidense points to construct separate surface and free-space masks [6]. A voxel grid can retain these distinctions as evidence channels: surface-supported,

traversed free space, or not observed. A learned occupancy prediction is a further estimate, not another observation of an unseen surface. The representation section returns to their use in a grid.

A finite-candidate policy first constructs endpoint poses and then selects an admissible row index [15, Secs. 3.1 and 3.5, pp. 47–53 and 58–62]. We use

$$\begin{aligned} q_{t,i} &:= \mathbf{T}_{\tilde{\mathcal{C}}_{t,i}}^{R_0} \in \text{SE}(3), \quad \mathcal{Q}_t = \{q_{t,i}\}_{i=1}^{N_q}, \\ \mathcal{A}_t &= \{i \in \{1, \dots, N_q\} : m_{t,i}^{\text{act}} = 1\}, \quad a_t \in \mathcal{A}_t \end{aligned} \tag{1}$$

Here $q_{t,i}$ is a rigid endpoint pose in $\text{SE}(3)$: its transform maps coordinates from the prospective candidate camera frame $\tilde{\mathcal{C}}_{t,i}$ into the fixed rollout-root rig frame R_0 . The root frame remains fixed as the camera moves. $\mathcal{Q}_t$ contains N_q proposed poses, and the action a_t is the integer row index selected from $\mathcal{A}_t$. The indicator $m_{t,i}^{\text{act}}$ admits row i only when the declared hard constraints are satisfied. PB-NBV uses a reachability- and camera-conditioned partial hemisphere of target-facing candidates [9]. Proposal geometry therefore bounds what even a perfect scorer can select.

Target-relative proposal geometry concentrates candidates around useful regions. PB-NBV points them toward the object centre, while Hestia conditions position on a predicted look-at point [8], [9]. An orbit can therefore offer views from several sides without ensuring that every side is visible or that the intervening motion is executable.

Geometric admissibility has at least two levels. Hestia shifts a proposed camera position to its nearest collision-free endpoint, whereas Next Best Sense obtains a list of feasible candidate views but still falls back to the next-ranked view when inverse kinematics or trajectory planning fails [8], [12]. A collision-free endpoint therefore does not by itself establish that the transition from the current state is executable. For the thesis, the geometric checks include mesh-clearance queries at finitely sampled positions along a camera movement segment. This is a discrete camera-path surrogate: it neither certifies the whole continuous path nor

models the volume of a wearer or robot body. Chapter 3 defines the checks and their admission role.

Geometric admissibility does not establish resemblance to wearer motion. Deliberate scanning and everyday egocentric activity can differ in how poses follow one another [4], [6]. The new trajectories also depend on the candidate generator and [collection policy](#), even when their initial recordings come from ASE. The experiments therefore separate geometric support and pose-continuity diagnostics from claims about human movement or physical deployment.

The utility or value function ranks only the admitted candidates; a rejected candidate is unavailable rather than merely low-value [9], [12]. Separating proposal, admission, and ranking makes the state dependence of available actions explicit [15, Secs. 3.1 and 3.5, pp. 47–53 and 58–62] and prepares the sequential decision model introduced next.

2.4 Sequential Decision-Making under Partial Observability

Histories that end at the same camera pose can contain different earlier observations. Their next views can consequently have different value. An updated geometric map retains some of that history; learned temporal memory is another possible construction. Updating a map with an acquired observation is distinct from predicting an observation that has not yet been obtained [16]. The core study examines acquired-observation updates; a static representation omits newly acquired observations, providing a control for the information incorporated by the update.

Partial observability concerns scene information that has not been acquired. Representation aliasing occurs when acquired distinctions are discarded: for example, two histories collapse to the same stored state even though they contain observations of different target surfaces. In a partially observable Markov decision process, a belief state is a sufficient statistic of the action–observation history for policy choice [16]. ARIA-NBV does not attempt general belief inference each

finite representation is a hypothesis about which history distinctions suffice for predicting target improvement.

For a static scene, fusion of fixed, calibrated observations with order-independent weights can produce the same geometric evidence regardless of their insertion order. This does not make the decision independent of current pose, reachable next actions, remaining budget, or changing scene conditions. Nor does it allow an observation to enter the state before it is acquired. The relevant question is which temporal distinctions affect future consequences, rather than whether every ordering must have a distinct encoding [16].

The preceding section defined the admissible choices as the state-dependent row-index set $\mathcal{A}(s_t)$ after proposal and feasibility checks. That set can change as the current pose, observations, and proposal geometry change [9]. Sequential value must account for which candidates are available now and which may become available after an acquisition.

A first acquisition’s value also depends on the subsequent choices it enables. Figure 6 illustrates this relation with a constructed two-step comparison. The following return makes the horizon and reward scale explicit [17].

Let $\mathcal{H}_t$ collect the actor-visible observations and previously selected actions available before decision step t . The symbol Ξ denotes one frozen decision setting: it fixes how candidates and their support are formed, how targets and states are constructed, how rewards, transitions, and termination are defined, and which discount and maximum horizon apply. Within that setting, row index i identifies the physical candidate pose $q_{t,i}$, while the event $a_t = i$ states that this row is selected. For target e and requested horizon h , the finite return and corresponding conditional value are

$$G_{t,e}^{(h)} = \sum_{k=0}^{h-1} \gamma^k r_{t+k}^e, \quad 1 \leq h \leq b_t \leq H_{\max} \quad (2)$$

$$\begin{aligned}
Q_{h,e}^{\star,\Xi}(\mathcal{H}_t, q_{t,i}) &= \sup_{\pi \in \Pi^{\text{act}}} \mathbb{E}_{\pi, a_t := i} \left[G_{t,e}^{(h)} \mid \mathcal{H}_t \right], \\
i &\in \mathcal{A}_t, \quad 1 \leq h \leq b_t \leq H_{\max}, \\
Q_{0,e}^{\star,\Xi}(\mathcal{H}_t, q_{t,i}) &= 0
\end{aligned} \tag{3}$$

Here r_{t+k}^e is the target reward from acquisition $t+k$, and γ weights later rewards: unit discount gives them equal weight. The return $G_{t,e}^{(h)}$ sums at most h such rewards. Each $\pi \in \Pi^{\text{act}}$ is a [continuation policy](#) that uses only the actor’s acquired information and selects admissible candidates. The expectation averages return over possible outcomes given the acquired history, with $a_t := i$ prescribing the first action and π choosing later actions; sup denotes the best expected return over those continuations. A recorded trajectory realizes one sequence of actions and observations. For a fixed continuation policy, the expectation accounts for the possible outcomes given the acquired history, including uncertainty about the unseen scene and any randomness in acquisition or action selection. It is not a uniform average over candidate poses: the policy specifies which later actions are chosen. The supremum then compares these policies. A return sampled under the collection policy is therefore not itself the optimal expected value [15, Secs. 3.3–3.6, pp. 54–67].

$Q_{h,e}^{\star,\Xi}$ compares physical candidate poses under the same decision setting rather than treating a row number as a camera pose. These equations specialize the standard discounted-return and action-value constructions to a requested target and finite horizon [15, Secs. 3.3–3.6, pp. 54–67]. Fixed-Horizon TD defines horizon-indexed predictions with the boundary $Q_0 = 0$ and relates horizon h to the shorter horizon $h-1$ [17]. When a prediction reaches its requested horizon, it has no rewards left to count. This boundary says nothing about whether reconstruction is complete. The optimal value states the consequence of taking a candidate now and then using the best admissible continuation under the frozen decision protocol; it is not an intrinsic property of the camera pose alone.

An [optimal oracle value](#), $Q_{h,e}^{\star,\text{oracle},\Xi}$, is a separate privileged reference. It maximizes expected finite return over admissible continuation policies that may use

the declared oracle information [15, Sec. 3.6, pp. 62–67]. The candidate-generation and execution protocol still bounds their available actions. The actor value above instead restricts continuations to acquired actor information; privileged optimal continuation need not be achievable by that actor. Extending the action space to unrestricted continuous poses would change the decision problem and remains future scope.

An **implemented oracle** uses a concrete collection or search policy, π^{oracle} , with finite computational resources. Its chosen trajectory yields the **sampled return** $G_{t,e}^{(h,\text{oracle})}$: a realized sum of rewards under that policy, not automatically the policy’s expected value or the privileged optimum. Knowing immediate rewards does not establish which sequence has the highest expected return. Multiple suitably sampled trajectories under a fixed policy can estimate that policy’s expected return; optimizing over admissible policies is a different operation [15, Secs. 3.5–3.6, pp. 58–67]. Chapter 3 states which branches the implemented collector actually records, and Chapter 5 separates return estimation from bounded search evidence.

The **prediction horizon** h states how many future acquisition rewards the query represents; the remaining **acquisition budget** b_t records how many acquisitions are still available. This thesis restricts model queries to $1 \leq h \leq b_t \leq H_{\max}$, where $H_{\max}$ is the largest supported horizon. A shorter query $h < b_t$ changes how far ahead the model predicts without changing the available budget. For example, with three acquisitions remaining and a supported maximum of at least three, queries for one or two future acquisitions fit the budget; a query for four does not. Formal validity alone does not establish that the model has adequate training evidence for every admissible query. An early terminal outcome contributes no later rewards; it does not authorize queries beyond these bounds. Universal Value Function Approximators establish the general possibility of conditioning one approximator on an additional task variable, while Fixed-Horizon TD shares representations across horizon-indexed predictions [17], [18]. Neither result licenses extrapolation beyond the horizons actually supported and evaluated.

The **recorded length** is the number of acquisitions retained in a data segment. It is an evidence limit, distinct from the prediction horizon and any real task budget. A complete two-acquisition recording supplies a two-step collector return even if scanning could continue. It does not supply an observed three-step return. A genuine end under the task’s rules contributes no later rewards; a recording that stops earlier leaves those rewards unknown. Pardo et al. distinguish task time limits from cutoffs used only during training [19]. Bootstrapping at a cutoff requires evidence of the successor from which continuation is estimated; the distinction alone cannot supply an absent observation.

Finite-horizon control chooses the next action using a shorter remaining prediction: h , then $h - 1$, down to zero [17]. Repeatedly asking for the next h acquisitions instead defines a rolling lookahead controller. Both can act in an ongoing environment, but they can choose different continuations. A value comparison must specify which continuation it predicts and executes.

Open decision: RRC03 — Choose countdown within a real acquisition budget or renewed bounded lookahead, and specify the role of acquisition count, time or motion cost.

Source: Finite-horizon control [17]; reward/continuation synthesis, 2 October 2026.

Gate: Name the execution clock and supported budget-horizon queries; align training and independent policy evaluation.

Greedy selection nevertheless has a principled role. Consider fixed sets of surface elements visible from each candidate. Coverage of several views is the union of those sets: as that union grows, an additional view can contribute fewer new elements, never more. Occlusion can determine the fixed visible sets without invalidating this diminishing-returns property. Lauri et al. study a related submodular multi-sensor NBV utility under their specified constraints [20]. Their guarantee does not transfer to an arbitrary reconstruction metric or motion-constrained sequence.

The preparatory move described in the Introduction changes future access: it makes a revealing second view available. A different source of delayed value is complementarity, where observations jointly constrain geometry better than their individual gains suggest. These mechanisms require separate reasoning; an occluded surface alone does not prove that greedy coverage is inadequate. Likewise,

repeated view selection can remain myopic even when the policy maintains an observation history [20].

When step gains are reductions of the same error on a fixed scale, summing undiscounted gains cancels intermediate errors and gives the initial-to-final reduction. Step-dependent normalization or discounting changes that relation. Chapter 3 owns the thesis’s exact reward and endpoint definitions [15].

Learning such values from a fixed data collection introduces an epistemic constraint in addition to the mathematical horizon. Batch-constrained and conservative offline reinforcement-learning methods are motivated by the error that arises when a learned policy assigns high value to actions outside the data distribution [21], [22]. These methods differ in how they control that error, but they share the warning relevant here: a longer horizon does not create evidence for transitions that the data never resolves.

A measured transition can nevertheless support a longer-horizon learning target: add its immediate reward to an estimate of the successor’s value for $h - 1$ acquisitions. The successor estimate may choose a different action from the collector’s recorded next choice. This is the off-policy mechanism of fitted Q-learning [23]; it does not require the recorded trajectory already to follow the best continuation. Its reliability still depends on the available experience and on whether the representation retains the information that determines those consequences. A complete observed h -step return and a target with estimated continuation are different evidence.

Together, partial observability, state-dependent choices, and finite-horizon return make candidate value explicitly relational [16], [17]. A camera pose has no context-free value. Its value depends on the causal information state, requested target, currently admissible support, horizon, state-update rule, and continuation policy. The next dependency is therefore representational. The actor state must preserve the distinctions on which those future consequences depend.

2.5 Representations for Egocentric NBV

The preceding section showed why future view value depends on acquired history as well as the candidate pose. A causal actor input excludes future and unselected observations. Among the distinctions available in its acquired history, the representation should preserve those that can change return for an admitted target, action, or horizon, while ignoring arbitrary descriptions of the same situation. Changing the coordinate origin leaves the acquired evidence unchanged; acquiring another view can change which observation is useful next [24].

Consistent geometric descriptions. Consider a **passive** change of world coordinates. Let $A = \mathbf{T}_{C_t}^W$ map the current camera frame into the world chart and $B = \mathbf{T}_{E_e}^W$ map the target frame into that same chart. Changing the chart by the same rigid transform G on both poses leaves their relative transform unchanged: $(GA)^{-1}(GB) = A^{-1}B$. Both sides map target coordinates into the current camera frame. The same change must re-express all scene, sensor, observation, and gravity quantities consistently. Scalar utility then agrees across descriptions, while geometric vectors and fields follow their stated transformation laws [25], [26]. The preceding section’s candidate pose still uses its fixed rollout-root rig as destination; this generic world-chart example does not redefine it. Changing only the target’s physical orientation, scene scale, camera calibration, gravity relative to the scene, or occlusion changes the sensing problem and need not preserve a score. Coordinate augmentation uses this distinction to describe stored evidence in another frame; it creates neither a new acquisition nor an independent scene.

A second description change merely reorders candidate rows. If two candidate views exchange positions in a table, their predicted scores must exchange positions with them. Pooling can produce row-invariant **context**, but a row-aligned score vector must be **permutation equivariant**. Deep Sets supplies invariant aggregation and Set Transformer permits equivariant candidate interaction [27], [28]. Independent candidate queries can also preserve row alignment; attention is not required. Removing a candidate from the current table changes which row can be selected. A surviving candidate’s value need not change under the same

continuation protocol, but changing which later actions become available after taking it can change its multi-step value.

Task-relevant information. Coordinate and row consistency concern equivalent descriptions. **Predictive sufficiency** asks whether the causal actor state retains the acquired distinctions needed for the task; causal admissibility alone does not establish it. **Task-relevant abstraction** then asks which physically different histories may be represented alike. Histories with different previously observed surfaces can require different next views, even when the current images are similar. Merging those histories can therefore discard information needed for action choice. Preserving every possible future observation is a stronger demand than predicting value for declared targets and candidates, but compression must retain any difference that changes their expected return. Adequacy therefore depends on the target, admitted actions, and prediction horizons. Ni and colleagues distinguish value-relevant, self-predictive, and observation-predictive history abstractions [24]. Their self-predictive condition uses expected reward and the **distribution** of the next latent state; matching only its mean is weaker. None of these theoretical conditions follows from using an egocentric voxel field.

For an observation-updated representation, encoding the extended history should agree with updating its previous encoding using the selected action and newly acquired observation. This **update consistency** alone does not establish adequacy. An encoder that maps every history to a constant can update to that constant forever, yet merge histories with different useful next views. Ni and colleagues use this collapse to explain why next-latent prediction needs task grounding such as reward prediction [24]. The task test is whether a representation retains differences needed for target-specific value.

Representation choices and information loss. Relevant geometry may never have been observed, lie outside the map’s extent, disappear in a coarse voxel, or be present in the input but discarded by encoder pooling. Extending map extent cannot create an unacquired observation. Finer voxels cannot recover geometry never sensed, and a decoder cannot generally undo many-to-one pooling.

EFM3D’s finite local support and surface/free-space input masks make the first three distinctions concrete [6]. A voxel supported by neither mask is unobserved, not measured empty space; learned geometry there is a prediction rather than acquired evidence.

These failures motivate the choice of **carrier**. Point representations retain irregular samples and local relations; sparse voxel fields provide structured spatial processing at a declared extent and resolution [29], [30]. Equivariant message passing encodes a selected geometric symmetry [31]. No carrier supplies missing observations or guarantees that useful distinctions survive compression. Transforming input geometry before fusion should ideally agree with transforming the fused representation, but finite grids may break exact agreement through interpolation, clipping, or revoxelization.

EFM3D provides a concrete egocentric carrier: calibrated poses lift image features into a gravity-aligned local voxel field with learned 3D processing and explicit geometric input evidence [6]. It freezes its image feature extractor while learning the upsampler, 3D U-Net, and task heads. Pretraining plays different roles elsewhere: 3D-NVS initializes a view classifier from ImageNet, MACARONS uses pretrained image features in mapping, and Next Best Sense uses learned priors for scene construction while retaining an analytic selector [3], [12], [13]. Whether an egocentric field improves candidate value prediction over matched geometric controls remains an empirical question.

Together, these criteria distinguish three questions about a representation: does it preserve the acquired evidence needed for view choice, behave consistently under equivalent descriptions, and retain useful distinctions despite finite extent and compression? These questions bring the geometric, task-predictive, and spatial perspectives together [6], [24], [25]. They guide comparisons between representations without prescribing one architecture. The Method chapter specifies the state and scorer. RQ2 examines how representations incorporate newly acquired observations; the comparison must declare which modalities are available at each acquisition.

Transfer across domains. Representation quality may also change across domains. A **distribution shift** means that the joint distribution of inputs and outcomes differs between training and evaluation or use. A **domain gap** or **domain shift** describes source-to-target differences; the **sim-to-real** or **reality gap** is the synthetic-to-physical instance. **Sim-to-real transfer** concerns performance carried across it. EFM3D reports reduced object-detection performance from ASE to real egocentric data and discusses appearance, placement, and taxonomy differences [6]. These identify relevant mismatch mechanisms, not NBV policy transfer.

For view selection, changed observations alter the actor’s evidence, while changed candidate and motion distributions alter which consequences it can learn and encounter. The motion-realism question in Section 2.3 complements the scene and sensor gap. EFM3D also describes reconstruction artifacts caused by the moving wearer [6]. Held-out ASE scenes test generalization within that synthetic setting; they cannot by themselves establish transfer to real indoor wearable use. The next section compares methods along these representation and evidence distinctions.

2.6 Related-Work Synthesis and Research Gap

The preceding sections distinguished target utility, candidate support, acquired evidence, and delayed value. The reviewed methods combine these ingredients in different ways [1], [9], [10], [11]. The appendix comparison in Table 28 records what each method values, what evidence it uses, which actions it can consider, and how it decides over time. These are setting differences, not a ranking of reported performance.

The utility matters because extra surface coverage, information about a scene model, and measured reconstruction improvement can rank the same views differently [1], [10], [11]. The target scope matters because a score for a requested object must value its geometry, while still using surrounding scene evidence to predict occlusion and access. A method’s objective and target scope must therefore be known before its sequential decisions can be compared with this thesis.

Sequential execution also needs to be distinguished from planning for delayed benefit. GenNBV uses history to choose successive views under a coverage reward [7]. Hestia likewise updates directional coverage, but its arXiv v4 version sets the reward discount to 0.1 expressly for a greedy-like objective [8]. Active Perception Network instead uses an incrementally updated map and a perception roadmap to plan non-myopic view sequences for exploration and surface coverage [32]. It establishes a non-learned planning counterpoint, not a result about target-region reconstruction error. Repeated view choice, learned memory, and non-myopic planning thus answer different questions.

Object difficulty itself also depends on the evaluation protocol. The ObjView-Bench preprint separates an estimated externally observable surface fraction, the number of sampled directions needed before coverage saturates, and the number of complementary views needed within a fixed candidate set [33]. These quantities distinguish hidden surface, slow revelation, and view-set selection in that protocol. They neither measure this thesis’s target-error lookahead headroom nor establish a universal limit for indoor acquisition. Here, the candidate generator, view budget, and geometric endpoint must be declared before an apparent planning advantage is interpreted.

The actor–oracle distinction established in the Introduction has a further consequence for learning: labels for every current candidate do not supply a complete tree of acquired successors. More labels at the current state cannot replace missing continuation experience. Even complete labels need not make all oracle advantage recoverable: the same acquired history can be compatible with hidden target surfaces requiring different best actions [16].

The reviewed methods establish direct-quality prediction, target-aware scoring, and sequential decisions under differing objectives and supports [1], [2], [7]. Their results leave open how to learn target-specific, multi-step reconstruction improvement from the observations available before choosing a view. In this setting, that question links the experience generated for training, the representation of acquired evidence, and the learning of future value. Chapter 3 specifies the target

tasks, candidate views, collected continuations, and privileged measurements that supply the learning problem. The motion-realism and transfer limitations established above bound the interpretation of the study's results.

3 Oracle and Data Generation

The preceding chapter established what a target-conditioned planning state must preserve. This chapter constructs the controlled decision process in which those requirements can be tested. Starting from a logged egocentric snippet and a requested target, the system proposes candidate observations, removes geometrically infeasible actions, and uses privileged scene geometry to measure the hypothetical reconstruction consequence of each remaining candidate.

For one acquisition, the system admits a target, proposes and filters camera poses, then renders the remaining candidates. It prepares the current reconstruction and each hypothetical updated reconstruction in the same way; the decrease in target error supplies a candidate reward. A collection policy chooses an action, whose observation extends only that branch’s acquired history. The successor retains the target and initial normalization, consumes one acquisition, and receives a newly generated candidate set.

This process supplies many immediate labels but only selected factual successors. Unselected renders never become actor observations. The following sections explain the observation sources, task and proposal construction, measurement, and collection in that order; Chapter 4 uses the resulting relations to define model inputs and learning targets.

3.1 Dataset and Ecosystem Choice

Choosing another view of the partially observed cabinet requires images and geometry acquired so far. Measuring the value of a hypothetical view also requires a reference for surfaces that remain hidden. An offline study therefore needs both a calibrated egocentric observation model and scene geometry reserved for supervision. The Aria ecosystem supplies complementary parts of this setting.

Project Aria provides the sensing reference: calibrated, time-aligned streams and tools for relating images, poses, and partial three-dimensional structure. Its Machine Perception Services produce trajectories, calibration, and semi-dense point clouds with observation tracks [4]. These products make geometric

correspondence inspectable. Their presence in a recording does not establish decision-time availability: the protocol must still restrict evidence to the permitted observation history and account for processing that uses later information.

Aria Synthetic Environments (ASE) supplies procedurally generated indoor scenes and simulated Aria observations, including RGB images, ground-truth trajectories, depth, instance labels, and synthetic semi-dense geometry [5]. EFM3D adds oriented object annotations and provides reference meshes for its simulated validation setting [6]. For the cabinet example, the reference geometry supports measuring the reconstruction consequence of a candidate observation; generating that observation requires a separately defined rendering or acquisition process. Ground-truth geometry remains in the evaluator; an observation-derived target description remains distinct from the ground-truth object used to score it. The subsequent sections specify how hypothetical views become supervision and which selected observations may extend the actor state.

EFM3D also supplies a relevant representation baseline: Egocentric Voxel Lifting combines posed, calibrated images and geometric priors in a local three-dimensional representation [6]. ATEK supplies preprocessing, WebDataset stores, and task evaluators that connect compatible data to such models [34], [35]. This shared tooling provides reusable interfaces for constructing the offline study. Target-conditioned view selection, feasible actions, and reconstruction-improvement supervision remain part of the thesis protocol; the upstream detection and surface-reconstruction tasks do not establish their validity.

The scale of the source corpus and the scope of an experiment are separate. ASE reports 100,000 generated scenes [5]. The scenes admitted to a particular experiment must be specified by its own inventory and scene-level split. Additional overlapping windows from a scene provide more observation histories, while sharing the same environment; they therefore cannot be counted as additional independent scenes.

Together, calibrated egocentric observations and reference geometry make the offline comparison possible. Synthetic appearance and trajectories limit what can

be inferred about real sensor noise and human motion. A renderable camera pose still needs a separate feasibility rule, and results in this setting do not establish on-device operation. The resulting substrate makes a controlled offline question possible: how much can a selected view improve reconstruction when the actor has access only to its permitted history?

3.2 State as an Information Contract

Consider a cabinet observed from the recorded Aria trajectory. Its frontal RGB, poses, semi-dense points, and derived EFM3D field are historical evidence. A synthetic side view is different: before selection it is a hypothetical oracle query; after selection, only the observation produced at that endpoint can join the causal rollout prefix. The actor and the offline oracle can receive different quantities from the same acquisition. Both when evidence becomes available and who may use it therefore matter when constructing actions, labels, and replay records.

Even within the synthetic setting, recorded and generated observations can have different density, noise and surface coverage. Historical RGB-derived features and newly rendered depth therefore need not follow the same input distribution. This source difference is separate from who may access the observations and from transfer to a physical sensor. Exact depth samples can reduce evaluated error on an unrepresented wall without requiring image texture. That gain describes the simulated acquisition; transfer to a physical sensor also depends on how it would observe and reconstruct the surface.

The distinction specializes the POMDP model introduced in [Section 2.4](#). The physical scene has a latent system state, while a policy acts from observations and their action history. A belief state is a sufficient statistic of that history in the classical model; ARIA-NBV instead studies finite information-state representations whose sufficiency must be tested [\[16\]](#). Consequently, **state space**, **observation space**, and **stored record** are not interchangeable terms.

3.2.1 Provenance and Access Roles

Two axes determine whether a quantity may influence a decision. **Acquisition provenance** distinguishes historical observations, selected counterfactual observations from earlier rollout steps, and current hypothetical candidate outcomes. **Access role** distinguishes actor inputs from oracle-only information. The resulting causal information sets are

$$\mathcal{J}_t^{\text{act}} = \left(\mathcal{O}^{\text{hist,act}}, \left(a_\tau, o_{\tau+1}^{\text{cf,sel,act}} \right)_{0 \leq \tau < t} \right) \quad (4)$$

Here $\mathcal{O}^{\text{hist,act}}$ contains the historical modalities admitted to the actor under the declared protocol. The ordered pairs $\left(a_\tau, o_{\tau+1}^{\text{cf,sel,act}} \right)$ contain only earlier selected actions and their admitted observations. Privileged information along the same causal branch is tracked separately:

$$\mathcal{J}_t^{\text{priv}} = \left(\mathcal{O}^{\text{hist,oracle}}, \left(o_{\tau+1}^{\text{cf,sel,oracle}} \right)_{0 \leq \tau < t} \right) \quad (5)$$

At the current decision, the oracle may additionally evaluate every candidate:

$$\mathbf{y}_t^{\text{oracle}} = \left\{ \left(o_{t,i}^{\text{cf,oracle}}, y_{t,i}^e \right) \right\}_{i=1}^{N_q} \quad (6)$$

The complete offline-oracle context is therefore

$$\mathcal{J}_t^{\text{oracle}} = \left(\mathcal{J}_t^{\text{act}}, \mathcal{J}_t^{\text{priv}}, \mathbf{y}_t^{\text{oracle}} \right) \quad (7)$$

The oracle context therefore contains $\mathcal{J}_t^{\text{act}}$ as well as privileged evaluation data. The actor receives only the quantities admitted by its observation protocol; an unselected candidate render never enters its next information set.

The access tag describes who consumes a field; its source role describes how the field was obtained. A mesh-rendered depth map can therefore be exposed to a control actor while remaining privileged and non-deployable. Actor-visible does not by itself mean sensor-observed. Every **observation modality** must retain both its access role and source provenance.

A declared modality profile μ compresses the causal actor information into a finite information state. Candidate generation then attaches a finite action support:

$$\begin{aligned}
s_t^\mu &:= \sigma_\mu(\mathcal{J}_t^{\text{act}}, e, b_t), \\
d_t^\mu &:= (s_t^\mu, \mathcal{Q}_t, \mathbf{m}_t^{\text{act}})
\end{aligned} \tag{8}$$

This factorization separates the state representation s_t^μ from the decision context d_t^μ . The candidate table and hard mask are generated action support, not sensor observations. S0-pose, S1-surface, and S2-ray are therefore representation or control profiles with different admitted modalities and source roles; they are not different physical worlds.

ASE contributes both sensor-like streams and ground truth; EFM3D transforms only the logged sensor-side subset into local 3D evidence [5], [6]. Co-location in one adapted sample or replay row never makes those layers equally observable.

The information boundary is also temporal. A render from an unselected candidate is a counterfactual oracle query. After that candidate is selected, an observation from its pose may enter the next state, but only under a declared acquisition model: a real observation, an admitted sensor simulation, or a privileged mesh render. Actor visibility requires both factual selection and an admitted source role.

3.2.2 Historical Observation Substrate

The historical substrate $\mathcal{J}^{\text{hist}}$ groups the evidence available before any counterfactual acquisition:

$$\mathcal{J}^{\text{hist}} = (\mathbf{I}^{\text{rgb}}, \mathbf{X}, \mathcal{P}^{\text{semi}}, \mathbf{F}_v) \tag{9}$$

The tuple groups complementary evidence, rather than four interchangeable measurements. RGB images record appearance: edges, texture, and visual cues may help identify the requested object and distinguish its visible parts. Calibration and trajectory poses relate image locations to viewing rays and place observations in a common spatial frame. They let us compare evidence acquired from different viewpoints, but do not alone reveal the surfaces hidden behind the cabinet front.

Semi-dense points $\mathcal{P}^{\text{semi}}$ locate sampled surfaces in metric space. They can anchor visible geometry and indicate where reconstruction has support, while uncertainty records qualify the reliability of those locations. Their timestamps

and observing-camera lineage retain when and from where support was obtained. Camera-to-point rays additionally provide evidence about traversed free space; a missing point is simply missing surface support. This specializes the surface/free/unobserved distinction from [Section 2.5](#) [5], [6].

The EFM3D field $\mathbf{F}_v$ is a derived representation of these sources. It combines posed image features with point and ray evidence in a finite, gravity-aligned local volume [6]. Its appeal is to make appearance and geometry available at common 3D locations, potentially helping a scorer relate a candidate view to the requested target. Learned field values are not direct surface measurements, and locations outside the volume have no representation support. Whether this representation improves return prediction remains a question for the later model comparisons.

The historical tuple uses only the observations admitted at the rollout root; its image and pose streams cover the declared historical window, and point support retains its recorded time and source. It does not denote all frames in the source trajectory. Availability and support masks identify where a source is present or usable. A zero value without that context cannot distinguish a measured zero from an unavailable observation. The modality profile selects which sources and derived features enter the state; the tuple does not require every profile to consume them all.

The target e_t and remaining budget b_t are task variables rather than historical sensor observations. They make the value conditional, conceptually $Q(s_t, e_t, a_t)$, and distinguish two otherwise identical observation histories posed with different requested objects or acquisition budgets. The implemented [GT](#) control derives e_t from a selected reference box. In the observed-target route, the detected box supplies the target instruction used by the model and candidate generator; its uniquely matched reference supplies the evaluation crop without replacing that instruction. The next section specifies target construction and the remaining privileged dependencies.

3.2.3 Selected Counterfactual Prefix and State Profiles

After an action is selected, actor information may grow only with evidence generated by that factual choice:

$$\mathcal{J}_{t+1}^{\text{act}} := \text{Append}(\mathcal{J}_t^{\text{act}}, a_t, o_{t+1}^{\text{cf,sel,act}}) \quad (10)$$

The pose-only S0 control records the chosen camera poses but does not add the selected RGB or depth observations to the scorer’s state. This is an input restriction, not a claim that the acquisition produces no observation. A visual successor instead adds the selected observation through the sources admitted by its profile. This update turns a hypothetical endpoint into the next factual branch; outcomes for the unselected rows remain oracle-only.

The architecture-neutral actor information-state target is

$$s_t^{\text{cf0}} = (\Phi_t^{\text{scene}}, \varphi_e, \mathbf{H}_t^{\text{pose}}, b_t) \in \mathcal{S}^{\text{cf0}} \quad (11)$$

It combines scene memory, the target descriptor, the ordered selected-view history, and the remaining acquisition budget. These components specify what information a model could use; their encoding and candidate-conditioned readout belong to the Method chapter. Here the root field is immutable, while accumulated scene memory may grow only through selected observations. Candidate rows $q_{t,i}$ and the hard mask $m_{t,i}^{\text{act}}$ belong to the associated decision context d_t^μ . Invalid-reason codes explain exclusions as diagnostic fields outside the value-model input. This separation prevents the proposal table from leaking the consequences it is supposed to predict.

The implemented **S0-pose** baseline retains root evidence and selected poses

$$s_t^{\text{S0-pose}} = \left(\mathcal{S}_{\text{root}}^{\text{VIN}}, \mathbf{F}_v^{\text{root}}, \left(\mathbf{T}_{E_e}^{R_0}, \ell_e \right), \mathbf{H}_t^{\text{pose}}, b_t \right) \quad (12)$$

It carries root evidence, target context, the selected-pose prefix, and the remaining budget. Its decision context separately attaches the current candidate table and mask. This state represents the actor’s motion history. A geometry-enriched successor additionally retains what the selected observations revealed:

$$s_t^{\text{cf+}} = \left(s_t^{\text{cf0}}, (\mathbf{D}, \mathbf{V}, \mathcal{P}^{\text{cf}}, \mathbf{n})_{1:t} \right) \quad (13)$$

In this geometry-state description, $\mathbf{D}$ gives selected-view depth, and $\mathbf{V}$ records the associated visibility or observation-support cue. Backprojected points $\mathcal{P}^{\text{cf}}$ place valid depth samples in the shared frame; normals $\mathbf{n}$ describe local surface orientation when available. These cues could help distinguish newly revealed surfaces from repeated views of already supported regions. Visibility and normals are derived quantities whose availability and source must be specified, rather than extra sensor streams assumed to exist for every profile.

The implemented selected-depth control retains the intermediate carrier

$$s_t^{\text{CF-GT-carrier}} = \left(s_t^{\text{S0-pose}}, \left(\mathbf{D}_\tau, \mathbf{V}_\tau, \mathcal{K}_\tau, \mathbf{T}_{C_{\tau+1}}^{R_0} \right)_{0 \leq \tau < t}^{\text{sel}} \right) \quad (14)$$

It retains depth, its validity mask, camera calibration, and camera pose only for selected actions. Depth and calibration locate surfaces along viewing rays; valid returns provide surface evidence, while the traversed ray segment can support free-space evidence. Invalid pixels and space behind a return remain unknown. Together with the poses, these observations could reveal cabinet sides missing from the root evidence. Selection makes this carrier causal along the rollout; its mesh-rendered depth remains privileged, rather than a sensor measurement.

The broader state could accumulate surface and ray support, reliability, and appearance evidence from those selected observations. RGB updates require an admitted image stream; updating an EFM3D field requires its corresponding posed inputs. Depth alone does not provide colour or learned image features. These are possible information updates, not a requirement that every source be available or every representation be implemented. The Method chapter specifies which combinations are used. At transition t to $t + 1$, any such update may consume only the observation produced by a_t . s_t^{cf0} is an **information state** for value prediction. Its sufficiency for the physical process is an empirical question: it is adequate only if it preserves the distinctions needed to predict future target-specific return.

3.2.4 Oracle Evaluation Context

The oracle evaluation context augments the causal state with quantities needed to answer counterfactual questions:

$$\kappa_t^{\text{oracle}} = (s_t^{\text{cf}+}, \mathcal{M}^{\text{GT}}, \mathcal{M}_e^{\text{GT}}, \mathbf{D}_q, \mathcal{P}_q, \text{RRI}) \quad (15)$$

ASE provides per-frame metric **GT** depth aligned with RGB, per-pixel instance identifiers, class mappings, and the scene trajectory. The EFM3D release adds ASE OBB metadata and validation meshes for object-detection and surface-reconstruction supervision [5], [6]. EFM3D uses depth to supervise occupancy at sampled free, surface, and behind-surface points, while visible OBBs supervise centerness, class, and box geometry [6]. These uses establish label provenance; the information source determines actor observability.

For ARIA-NBV, the scene mesh $\mathcal{M}^{\text{GT}}$ supports candidate rendering, and the target crop $\mathcal{M}_e^{\text{GT}}$ fixes the surface on which error is measured. All-candidate depth $\mathbf{D}_q$ and backprojected points $\mathcal{P}_q$ instantiate the hypothetical observation associated with each candidate. The resulting RRI labels the consequence of an action and becomes available only to the oracle.

In the **GT** control, the reference OBB defines the target instruction and bearing as well as the evaluation identity and crop. The observed-target route instead uses detected geometry for the model’s instruction and candidate proposals, retaining the matched reference for scoring. This separates target information from its evaluation reference; it does not make the whole action protocol observation-only, since current feasibility checks still use the reference mesh.

Historical actor evidence is shared by rollouts from the same source. Selected observations may extend it step by step, according to the observation profile. The model represents that history as an information state and combines it with the generated candidate table to form the decision context. The oracle adds reference geometry and hypothetical outcomes for supervision. [Section 3.6](#) maps these relations to stored source, state, candidate, and successor records.

3.2.5 Visibility, Feasibility, and Utility

Three predicates apply to a candidate and answer different questions. **Visibility** asks whether a view supplies evidence about a spatial element. **Feasibility** asks whether the action satisfies the declared pose and movement constraints. **Utility**

asks how the target reconstruction changes after an admissible action. A candidate can be physically feasible yet outside the local EFM3D volume, or it can frame the target yet yield little reconstruction gain. These cases coexist because the predicates refer to different structures.

The hard action mask $m_{t,i}^{\text{act}}$ excludes geometrically invalid candidates from policy selection, stochastic normalization, and value supervision. When the optional feasibility auxiliary is enabled, materialized hard-invalid candidates provide negative feasibility labels; padding provides no teacher. Reward is defined on feasible actions and may be low or negative. Oracle evaluation can fail independently, for example because a target crop or render is unusable. The configured recipe then skips the row or table, or clears the label mask $m_{t,i}^{Q,h}$ and Q-training eligibility, while reporting the failure separately. This failure does not change the geometric reason code.

Value learning requires both feasible actions and supported labels: the selected-row loss and next-state backup use label support within the feasible action set. Thus an inadmissible action, a feasible action with unknown gain, and a feasible action with negative gain receive different treatment. Oracle-derived feasibility also remains distinct from deployable validity in an actor-visible target protocol.

Actor visibility is consequently an end-to-end property of the decision protocol. It includes how the target instruction is obtained, how candidate support is proposed, how the hard mask is computed, which selected observation updates the next state, and which fields reach the scorer. A scorer can be free of privileged tensors while still choosing from support oriented or pruned with privileged geometry. Such a protocol is a valid bounded oracle or control experiment, but its result cannot silently become a claim about independently actor-constructed actions.

3.3 Task and Action Construction

Comparing possible next views requires a description of the requested object, a set of camera poses, and reference geometry against which to measure the resulting reconstruction. The **target task** identifies the requested object e and its evaluation mesh $\mathcal{M}_e^{\text{GT}}$. The target **descriptor** brings together geometry, semantic evidence, and any available observation summaries:

$$\varphi_e = \left(\hat{\mathbf{B}}_e, \hat{\mathbf{y}}_e, \hat{\pi}_e, \mathbf{s}_e^{\text{supp}}, \ell_e^{\text{src}}, \mathbf{T}_{E_e}^{R_0}, \mathbf{T}_{E_e}^{C_t} \right) \quad (16)$$

Here $\hat{\mathbf{B}}_e$ is the estimated oriented box. Its centre anchors views around the requested object, its orientation relates those views to object sides, and its side lengths describe the spatial region to inspect. For the cabinet, these descriptors distinguish a side view of the requested cabinet from a similar camera move around another object. $\hat{\mathbf{y}}_e$ records its class, while $\hat{\pi}_e$ is detector confidence when provided by the target source. Confidence expresses the detector’s certainty, not how much reconstruction remains to be gained. The two transforms map target-frame coordinates into the fixed rollout-root rig frame and the current camera frame, respectively; they express the same target geometry relative to those viewpoints, using the frame convention introduced in [Section 2.5](#). The source label ℓ_e^{src} distinguishes an observed detection from a reference-derived control target.

The optional summary $\mathbf{s}_e^{\text{supp}}$ describes evidence already acquired for the target. Possible entries include projected-box and visible-mask fractions of image area, the number of acquired views, distinct observed surface elements, and the number of observed viewing-direction bins. These quantities answer different questions: projected box area describes image framing, whereas the visible mask accounts for occlusion. Repeated views measure exposure, but do not necessarily reveal new surfaces. Surface and direction summaries require associations and measurements from their producers; an unavailable measurement is distinct from a measured zero.

This tuple is an inventory of potentially useful descriptors, rather than a **learned encoding** or a requirement to use every field. The implemented target

instruction supplies class identity, box pose and side lengths, and a root-relative pose. Detected-target records additionally retain confidence and source for audit. Observation summaries depend on the available source measurements; listing them here does not imply that every target has them. The Method chapter selects the scorer inputs and their representation. The descriptor supplies the object estimate used to propose views. The reference mesh has a different role: it allows the oracle to evaluate the reconstruction, including surfaces absent from the observations. Its hidden surfaces and the resulting oracle gains are not descriptor features available to the actor.

At decision step t , the proposal mechanism produces a finite **candidate table** $\mathcal{Q}_t$, also called a **candidate shell**. Here shell denotes the set of proposed views, not a requirement that their positions lie on one sphere. Its row index i aligns endpoint pose $q_{t,i}$ with proposal provenance and support fields; those fields do not change the pose’s meaning. The hard mask $m_{t,i}^{\text{act}}$ then induces the feasible action set

$$\begin{aligned} q_{t,i} &:= \mathbf{T}_{\tilde{\mathbf{c}}_{t,i}}^{R_0} \in \text{SE}(3), \quad \mathcal{Q}_t = \{q_{t,i}\}_{i=1}^{N_q}, \\ \mathcal{A}_t &= \{i \in \{1, \dots, N_q\} : m_{t,i}^{\text{act}} = 1\}, \quad a_t \in \mathcal{A}_t \end{aligned} \tag{17}$$

Generation determines which views are offered; the hard mask identifies which may be executed. Rejected rows are retained so that failures can be traced to the proposal or feasibility step. Among feasible views, a measured low gain is a valid training example. A missing oracle label instead leaves the gain unknown. Learning a nonterminal continuation also requires a recorded successor, as explained below.

3.3.1 Target Selection

The implementation provides two ways to construct a target task. In the **observed-target route** (V1), detected object boxes provide the descriptors used by the actor and the candidate generator. Every non-padding descriptor is recorded, and each descriptor with exactly one qualifying ground-truth match is admitted. The match supplies the reference crop for evaluation; it does not replace the

detected box in the actor’s input. There is no target-count cap at this construction stage.

The **ground-truth control** (V0) instead chooses targets directly from the reference object table. It samples eligible objects uniformly without replacement up to a per-snippet cap. Selection is independent of measured reconstruction gain, so difficult, nearly complete, and negative-gain targets remain eligible.

For this control, the sampler uses the latest ground-truth OBB slice containing non-padding rows. A row is eligible when its serialized values are finite and its extents are strictly positive. The implementation calls this status `matched`; here it denotes eligibility, not an observed-to-reference match. Rows with invalid auxiliary values remain in the audit table but are not sampled. Invalid poses or extents that prevent descriptor construction raise an error before a row can enter that table.

The applied development collection further restricts this reference population to objects with measurable reconstruction support at the current rollout root. The prepared ASE-depth prefix must contribute at least one point inside the unexpanded target box under the same fusion and point-preparation procedure used for scoring. A valid cabinet box with no reconstructed points inside it is therefore excluded before sampling; it is not assigned zero improvement. The sampler draws uniformly without replacement from this support-eligible pool. Root support concerns acquired points, observed-target admission concerns a unique box association, and proposal eligibility concerns available camera motions. These tests define different populations. The sampling receipt records the root identity, each object’s support count and eligibility, and each selected object’s inclusion probability. An empty eligible pool or unavailable support inspection produces an exclusion.

Neither construction rule applies an additional threshold to finite confidence values. The resulting populations nevertheless differ: the control starts from known reference objects, whereas the observed route includes only detections that can be associated unambiguously with a reference object. Observed-target results

therefore require reporting association failures as well as performance on admitted targets.

Conflict: Freeze the target source and information available to the core RQ2 comparisons. Distinguish observation-derived targets from historical GT controls and the executable state and feasibility paths. Method’s privileged baseline does not establish an end-to-end actor-visible comparison.
Source: RQ2; Target Selection; Method, current and target actor state; Experimental Design, task population and policy information
Gate: For each comparison, freeze the target source, admission rule, actor state, proposal profile and feasibility regime. Align Method and the experiment manifest without relabelling historical GT evidence.

3.3.2 Target population and sampling priority

Detection confidence concerns the reliability of an object estimate, not how much of its surface remains to be reconstructed. Admission defines eligible targets; sampling determines their exposure in training. The value prediction conditions on the supplied target [18].

The implemented sampler selects a uniform subset of eligible reference targets; the observed route includes every uniquely matched descriptor. Repeated objects in overlapping snippets nevertheless receive repeated exposure, so neither route samples unique objects or scenes uniformly. Eligibility is fixed before measuring gain, and target identity and root normalization remain fixed along a rollout. Exact inclusion probabilities, execution exclusions and possible sampling alternatives are retained in [Section C.8.1](#).

3.3.3 Observed-Target Admission

The observed-target matcher implements the **V1** admission stage. Given an observed or predicted proposal $\hat{e}$, it compares only same-class reference OBBs using oriented three-dimensional IoU,

$$\mu_{\text{IoU}}(\hat{e}, e) = \text{IoU}_{3\text{D}}(\hat{\mathbf{B}}_{\hat{e}}, \mathbf{B}_e) \quad (18)$$

$$\tau_{\text{IoU}} = 0.20 \quad (19)$$

A reference row qualifies only when its IoU is strictly greater than 0.20. In the following count, $\kappa(\hat{y}_{\hat{e}}, y_e)$ equals one when the proposal and reference have the

same semantic class, and zero otherwise. The proposal is admitted exactly when one reference row qualifies:

$$n_{\text{qual}}(\hat{e}) = \text{card}(\{e \in \mathcal{E} : \kappa(\hat{y}_{\hat{e}}, y_e) = 1 \wedge \mu_{\text{IoU}}(\hat{e}, e) > \tau_{\text{IoU}}\}) \quad (20)$$

$$a_{\text{id}}(\hat{e}) = 1 \text{ iff } n_{\text{qual}}(\hat{e}) = 1 \quad (21)$$

Equality at the threshold, no qualifying row, and multiple qualifying rows leave the proposal unresolved. The rule associates an observed proposal with an unambiguous evaluation identity. Association confidence remains a separate diagnostic. The legacy GT sampler bypasses this observed-to-reference matching stage because it constructs its tasks directly from reference rows.

The floor-mat association illustrates why admission uses volumetric overlap: a 7.8 mm-high reference and a 100 mm-high detected box can share a footprint without sufficient three-dimensional overlap. The complete mixed illustration (Figure 34) is retained with the allocation alternatives; its containment, history-priority and weighted-exposure stages are unselected. Current admission remains the unique-reference rule above, followed by action selection for that fixed target.

3.3.4 Candidate View Generation

Candidate generation is a proposal distribution over a bounded local camera motion space. It determines the alternatives over which the NBV scorer ranks actions. At step t , it constructs the shell

$$\mathcal{Q}_t = \{q_{t,i}\}_{i=1}^{N_q}, \quad N_q = 60 \quad (22)$$

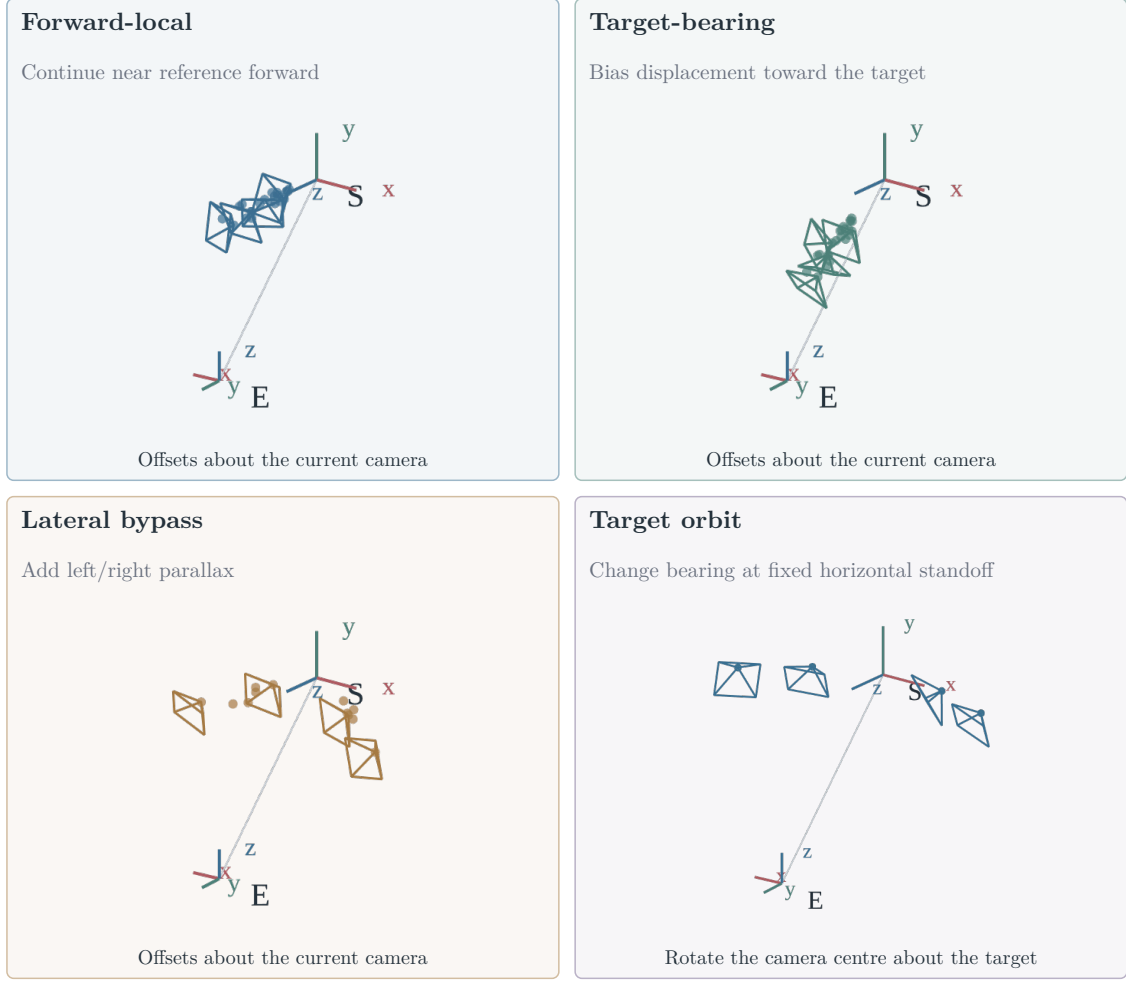
and records a family label $k(i)$ for every row. The generator’s default and earlier 60-candidate control allocate 24 forward-local, 24 target-bearing-local, and 12 lateral-target-bypass rows. The applied October development profile instead allocates 18 forward, 18 bearing, 12 bypass, and 12 orbit rows. The first three families perturb local translation; orbit proposals add paired angular offsets around the target. The table specifies the applied mixture:

Family	Geometric role	Profile support
Forward-local	Preserve egocentric motion continuity while perturbing the current forward direction.	18 rows; radius $[0.25, 1.1]$ m
Target-bearing-local	Move approximately along the current target bearing and orient the camera towards the target.	18 rows; radius $[0.4, 1.1]$ m
Lateral-target-bypass	Introduce side-step parallax while retaining a component towards the target.	12 rows; radius $[0.25, 1.1]$ m
Target-orbit	Move to either side around the target at the current horizontal standoff.	12 rows; distance-scaled signed offsets capped at 26° ; 1 m step budget

Table 1: Geometric roles in the applied 60-candidate development profile. Counts and bounds define this intervention; they do not establish individually optimal settings or wearer-motion realism.

Optional refinement, revisit, and target-shell families provide additional challenger profiles. Because a profile changes the represented action support, its family counts, radii, angular caps, and pruning parameters belong to the resolved run manifest. A comparison fixes that manifest across methods.

Figure 2 compares four position families around one measured root and target. Separating position generation from camera gaze makes their different motion hypotheses visible.



Sampling frame S and target frame E_e . Dots: centres; wireframes: calibrated views.

Illustrative unpruned proposals, not a collected 60-row shell.

First three panels share metric scale; orbit is fitted independently.

Figure 2: Generator-derived proposals around the measured ASE 83269/000056 root and observed chair. Four calibrated frusta illustrate each family. The local-family proposals are seeded simulations before feasibility filtering; the orbit panel uses illustrative signed angles rather than the applied distance-scaled twelve-row schedule. These are not archived cohort rows. All four panels share the sampling frame and measured target. The first three panels share metric scale; the orbit panel is fitted independently; the applied counts and bounds are specified separately in [Table 1](#).

Local families use bounded, forward-biased directional draws and a family-conditioned radius. Their maps reinterpret these draws relative to rig forward, target bearing or a lateral bypass direction. The exact remapping, radius law and seeded distribution illustration are retained in [Section C.8](#); these are applied sampler details, not an additional proposal family.

The base gaze of a forward-local row reuses the gravity-aligned sampling orientation. Target-looking rows use the supplied target centre $\mathbf{p}_e$ to construct an orthonormal look-at frame. In world coordinates, up is $\mathbf{e}_z^W$:

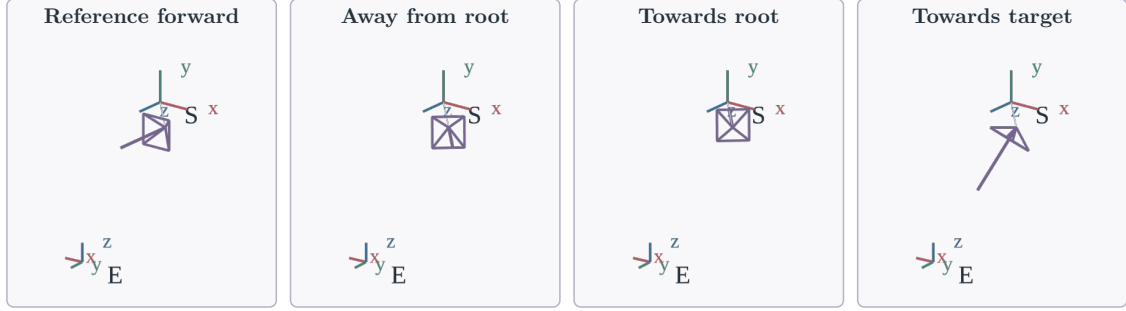
$$\begin{aligned}
\mathbf{z}_i^W &= \frac{\mathbf{p}_e - \mathbf{c}_i^W}{\|\mathbf{p}_e - \mathbf{c}_i^W\|_2}, \\
\mathbf{g}_i^{\text{up}} &= \mathbf{e}_z^W - (\mathbf{e}_z^{(W)^T} \mathbf{z}_i^W) \mathbf{z}_i^W, \\
\mathbf{y}_i^W &= \frac{\mathbf{g}_i^{\text{up}}}{\|\mathbf{g}_i^{\text{up}}\|_2}, \\
\mathbf{x}_i^W &= \mathbf{y}_i^W \times \mathbf{z}_i^W
\end{aligned} \tag{23}$$

The paired projections in [Figure 37](#) and [Figure 38](#) show why position and gaze must remain distinct: a calibrated field of view can exist while the same camera misses a target-centred proxy surface. Proxy footprints indicate potential framing, not measured visibility.

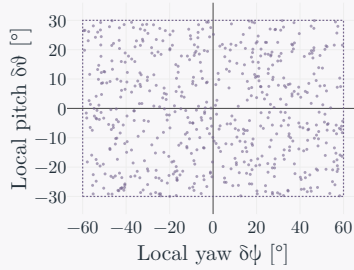
The final orientation combines this base gaze with the configured local rotation perturbation, or **view jitter**. With jitter disabled, a forward-local candidate retains the sampling-frame rotation and a target-looking candidate points at the target centre. With jitter enabled, the final gaze varies around that base. Framing and pruning therefore depend on both choices.

[Figure 3](#) holds the position fixed to isolate the base-gaze choice, then displays the local perturbation law separately.

1 Choose the base gaze at one camera centre



2 Perturb in the local camera frame



$$\delta\psi \sim \mathcal{U}(-60^\circ, 60^\circ)$$

$$\delta\theta \sim \mathcal{U}(-30^\circ, 30^\circ)$$

600 actual draws from the configured box law. Dotted boundary: yaw/pitch support; roll is zero.

Top: no jitter, common measured root/target and metric scale. Bottom: local perturbations before composition.

Figure 3: Position and orientation are separate choices. The upper panels compare four implemented base-gaze rules using the same measured root and target. The lower panel contains 600 draws from the campaign’s local yaw/pitch box; these rotations compose with the base camera frame and use zero roll.

The supplied centre comes from the reference box in the GT control and from the detected box in the observed-target route. The current feasibility checks still use the reference mesh. Thus an observed target centre removes one use of privileged geometry, but not its use in determining which actions are allowed.

An unselected target-centred shell construction is retained in [Section 9.4](#). It is separate from the applied orbit family and does not alter the development profile described here.

The target-orbit family in the applied development mixture makes this distinction explicit. It proposes paired signed angular offsets at the current horizontal target standoff, thereby testing bilateral partial-orbit coverage. Pruning can remove one side, so proposal symmetry and feasible symmetry are measured separately. The family adds target-relative angular coverage; its geometry does not establish human-motion realism.

Pruning evaluates each proposal independently of its oracle reward. A row is feasible only if it lies in the snippet support, stays clear of the GT mesh, passes the configured path-feasibility check, and satisfies the enabled local motion bounds. The applied profile retains a maximum step of 1 m and a maximum absolute height change of 0.25 m, together with snippet, mesh-clearance and path checks. It disables the optional backward-motion and heading-change caps. The earlier numerical control is retained in Section C.9. Changing these bounds changes available actions without establishing wearer-motion realism.

With the configured PyTorch3D backend, path feasibility is evaluated through mesh-clearance queries at finitely many positions along the movement segment, including both endpoints. A thin obstacle can therefore lie between sampled positions. Passing this check is an operational feasibility condition, not a certificate that the continuous path is collision-free. The test also concerns the camera path, not the clearance of a wearer or robot body.

The full shell retains each row’s family and feasibility results, including reasons for rejection. The hard mask produces $\mathcal{A}(s_t)$. A separate root-support check can reject tasks with too few remaining alternatives. The applied collection instead requires at least three hard-valid rows in the combined bearing, bypass, and orbit population. Forward-local rows do not contribute to this count. There is no additional total-shell minimum and no per-family minimum: three bearing rows with no feasible bypass or orbit rows can satisfy the combined rule. Family provenance defines target-relative proposal opportunity, not measured target visibility or label availability.

Each profile applies its declared check to the actual seeded step-zero shell immediately after generation and before any oracle candidate evaluation. A preliminary feasibility screen does not replace this check because it can generate a different shell. Rejection retains the actual shell’s valid count and reason, performs no oracle scoring for that root, and remains visible in the population report together with family-specific failures.

Implementation correspondence. The mixture generator also stores `sampler_probability`, currently the reciprocal of the configured total row count for every candidate. This is a uniform weight over table rows. It is not the geometric sampling density, the probability of surviving pruning, or the probability that the collection policy selects that action.

3.4 Measurement and Outcomes

The candidate shell defines possible interventions; the measurement operator assigns their target-specific consequences. The important object is therefore not an isolated candidate cloud but a **paired state update**: the current observation multiset and the same multiset with one candidate observation appended both pass through one declared preparation operator. Only then are their target errors compared. This shared path makes the hypothetical label the error change that the selected observation will realize when it enters the factual history. Rendering consequently affects both the reward measurement and the selected observation from which a successor is constructed. An altered depth map can change the evaluated reconstruction and, in an observation-updated model, the input used for the next prediction. The pose-only control excludes that new geometry from the scorer even though the evaluator uses it.

The two depth sources enter this path through different, explicit camera models. Root evidence uses all observed-prefix ASE [GT](#) ray-distance frames up to the fixed rollout root. Each valid finite positive depth sample within the far plane is unprojected with its recorded camera calibration into a camera-frame point image, spatially strided, transformed through the camera-rig trajectory into world metres, and canonically voxel-fused. The result is the fixed root cloud $\mathcal{P}_0^{\text{root}}$, already a prepared evidence cloud rather than the raw collection of all depth pixels. MPS semi-dense points remain an actor-side representation and legacy diagnostic they do not define the default oracle root error.

For candidate i , the oracle renders metric camera- z depth at the declared virtual pinhole camera against the **whole scene mesh**. The face-hit, near-plane, far-plane, and finite-depth masks remove unusable pixels. The remaining image

lattice is sampled at the declared spatial stride, and selected pixel centres are backprojected by the renderer camera directly into world metres to form the full acquired cloud $\mathcal{P}_{t,i}^{\text{acq}}$. Rendering before the target crop preserves occlusion by the surrounding scene. These ray-distance and camera- z conventions are part of the declared implementation contract; this thesis does not treat their present cross-source calibration as independently validated merely because both produce world-frame points.

Pixel stride, voxel fusion, and point capping are three different reductions. Stride selects a regular subset of the **image lattice before** 3D fusion. Voxel fusion groups finite world points by $\lfloor \frac{\mathbf{p}}{h} \rfloor$ and replaces every occupied cell by the arithmetic mean

$$n_v(\mathcal{P}) = \sum_{\mathbf{p} \in \mathcal{P}} \mathbb{1} \left[\left\lfloor \frac{\mathbf{p}}{h} \right\rfloor = \mathbf{v} \right], \quad \mu_v(\mathcal{P}) = \frac{1}{n_v(\mathcal{P})} \sum_{\mathbf{p} \in \mathcal{P}} \mathbb{1} \left[\left\lfloor \frac{\mathbf{p}}{h} \right\rfloor = \mathbf{v} \right] \mathbf{p} \quad (24)$$

Here n_v is the number of retained samples in occupied voxel $\mathbf{v}$ and $h > 0$ is its edge length in metres; setting $h = 0$ disables voxel aggregation. Each retained sample has equal weight, so repeated observations influence a voxel centroid in proportion to the number of samples they contribute. If an old cloud and a new cloud were averaged separately and their two representatives were then averaged again without their counts, the result would generally differ from the mean of all underlying samples. This non-associativity of unweighted representative fusion is why the fixed root, every full selected observation, and the hypothetical candidate are concatenated **before** voxel statistics are recomputed.

After fusion, an optional cap deterministically thins the ordered point representatives. It neither samples randomly nor optimizes geometric coverage, and it reserves no capacity for the newest observation. It can therefore discard all points contributed by that observation. The exact ordering and retained-index rule are given in [Section C.10.1](#). Crop, fusion and cap settings remain part of the measured outcome, rather than interchangeable forms of downsampling.



Figure 4: Result-affecting preparation order for both the current and candidate-appended target states, read $1 \rightarrow 2 \rightarrow 3$, then $4 \rightarrow 5 \rightarrow 6$. Configurations may make two voxel stages numerically redundant, but their distinct caps and declared grids remain separate protocol stages.

Let $\mathcal{F}_{h,N}$ denote finite filtering, voxel-mean fusion at edge length h , and the ordered cap $\mathcal{K}_N$; let C_e crop points to the target OBB. The target route uses no cap before cropping, then applies its target-local cap, and finally performs the metric preparation:

$$\Phi_e(\mathcal{P}) = \mathcal{F}_{h_m, N_m} \left(\mathcal{F}_{h_e, N_e} \left(C_e \left(\mathcal{F}_{h_e, \infty}(\mathcal{P}) \right) \right) \right) \quad (25)$$

Here h_e is the evaluation-grid edge length in metres, N_e is the target-local cap count, h_m is the final metric-grid edge length in metres, and N_m is the final metric cap count. The symbol ∞ means that the pre-crop target route is uncapped. A zero edge length disables the corresponding voxel aggregation while leaving that stage’s finite filtering and cap semantics declared.

For the default full-scene rendering route, the current and hypothetical states differ only in whether the full candidate cloud is present in the multiplicity-preserving concatenation:

Here $\mathcal{P}_j^{\text{sel}}$ denotes raw samples from the selected acquisition at step j ; $\mathcal{P}_{t,i}^{\text{acq}}$ denotes the current candidate’s raw samples. These observation clouds are distinct from the candidate-pose table $\mathcal{Q}_t$.

$$\begin{aligned}\mathcal{P}_t^e &= \Phi_e \left(\text{Concat} \left(\mathcal{P}_0^{\text{root}}, (\mathcal{P}_j^{\text{sel}})_{j < t} \right) \right), \\ \mathcal{P}_{t|i}^e &= \Phi_e \left(\text{Concat} \left(\mathcal{P}_0^{\text{root}}, (\mathcal{P}_j^{\text{sel}})_{j < t}, \mathcal{P}_{t,i}^{\text{acq}} \right) \right)\end{aligned}\tag{26}$$

The paired label/commit procedure is summarized in [Figure 44](#) in the Reproducibility Record. The measurement diagram gives the operation order; the displayed operator fixes its mathematical meaning. Label evaluation and committed updates use the same preparation, rather than separate source-wise averaging.

An optional conservative viewing crop reduces scene geometry for hypothetical candidate rendering. At each step, a world-axis-aligned region encloses the candidate camera centres and expanded target-box corners. Whole faces whose bounding boxes intersect this region are retained, including potential foreground occluders between camera and target. A wall outside the target box can therefore continue to hide the cabinet surface. The expansion preserves rendering and fusion support; the metric still uses the unexpanded target box. This is conservative face selection rather than exact scoring-surface clipping; background pixels outside the required support may change.

After selection, the procedure reacquires the chosen observation against the original full scene before constructing its successor. The cropped candidate cloud cannot substitute for this full-scene causal acquisition. The tested profile uses 48×48 candidate acquisition; separately retained 240×240 selected-view depth serves later inspection and is distinct from the observation committed to causal history. Candidate cropping is optional and disabled by default.

Validation TODO: Validate repeatability, numerical-failure handling, construct interpretation, root normalization, and matched endpoint re-evaluation before treating target gain as a reliable study outcome.
Source: docs/literature/tex-src/arXiv-VIN-NBV; metric-validation artifacts
Gate: frozen metric protocol, repeatability table, and claim-level source locators

3.4.1 Operational Reconstruction Error

The error is evaluated on the prepared clouds $\mathcal{P}_t^e$ and $\mathcal{P}_{t|i}^e$ defined above. Both combine the ASE-depth root with selected history through the same operator Φ_e ; the latter also includes the hypothetical candidate observation. The distinction

between this evaluation geometry and actor-side semidense evidence remains unchanged.

The reference-mesh crop is part of the measurement identity. The historical policy retains a whole face whenever any vertex lies inside the target OBB and its declared margin; it can omit a crossing face whose vertices all lie outside, and include portions outside the box. The exact-box policy instead intersects every triangle with the six closed half spaces of the GT OBB and triangulates the resulting in-box polygon without inventing box surfaces. Crossing triangles therefore contribute their interior portions without a scoring margin. Metric point centres use the same unexpanded OBB, while the recorded fusion order is retained. These policies define different reference geometries and their labels must not be pooled as equivalent measurements. The mesh crop limits error reduction rather than visibility: foreground occluders outside the target still need to participate in candidate rendering. A rendering-only visibility or fusion halo does not enlarge the metric target.

The evaluation cloud measures what has been reconstructed. The actor’s representation determines what the policy can use to choose its next view. In the static-root baseline, the former accumulates selected-view points while the scene features remain fixed. A measured reconstruction improvement therefore does not by itself mean that the actor has incorporated the new surface information. The Method chapter defines its retained state and update rule.

For target mesh $M_e = \mathcal{M}_e^{\text{GT}}$, the local distance $d(\mathbf{p}, \mathbf{f})$ follows the operational backend convention. Let T_f be the filled triangle, ∂T_f its three boundary segments, and a_f its area. Then

$$d(\mathbf{p}, \mathbf{f}) = \begin{cases} \min_{\mathbf{x} \in T_f} \|\mathbf{p} - \mathbf{x}\|_2 & \text{if } a_f \geq a_{\min} \text{ and normal valid} \\ \min_{\mathbf{x} \in \partial T_f} \|\mathbf{p} - \mathbf{x}\|_2 & \text{otherwise} \end{cases} \quad (27)$$

Here $a_{\min} = 0.005 \text{ m}^2$, or 50 cm^2 . The first branch applies when the area is at least that threshold and the backend accepts the triangle normal as nondegenerate; otherwise the distance is to the boundary segments. The strict area comparison

and the backend’s separate numerical normal test are part of the pinned measurement protocol. They do not define an ideal surface integral.

A right triangle with two 2 cm legs has area 2 cm². A point at (5, 5, 0) mm lies inside its plane and filled surface, so its ideal distance is zero. Under the operational small-triangle branch it is 5 mm from the nearest edge, giving squared distance 2.5×10^{-5} m². Thus even a point on a small face can have nonzero operational error.

The point-to-mesh term averages this squared distance from each reconstructed point to its closest reference triangle:

$$D_{P \rightarrow M}(\mathcal{P}, \mathcal{M}^{\text{GT}}) = \frac{1}{\|\mathcal{P}\|} \sum_{\mathbf{p} \in \mathcal{P}} \min_{\mathbf{f} \in \mathcal{F}^{\text{GT}}} d(\mathbf{p}, \mathbf{f})^2 \quad (28)$$

It measures how closely the reconstructed evidence lies on the operational target surface and is sensitive to extra, noisy, or misregistered points. The reverse mesh-to-point term is

$$D_{M \rightarrow P}(\mathcal{P}, \mathcal{M}^{\text{GT}}) = \frac{1}{\|\mathcal{F}^{\text{GT}}\|} \sum_{\mathbf{f} \in \mathcal{F}^{\text{GT}}} \min_{\mathbf{p} \in \mathcal{P}} d(\mathbf{p}, \mathbf{f})^2 \quad (29)$$

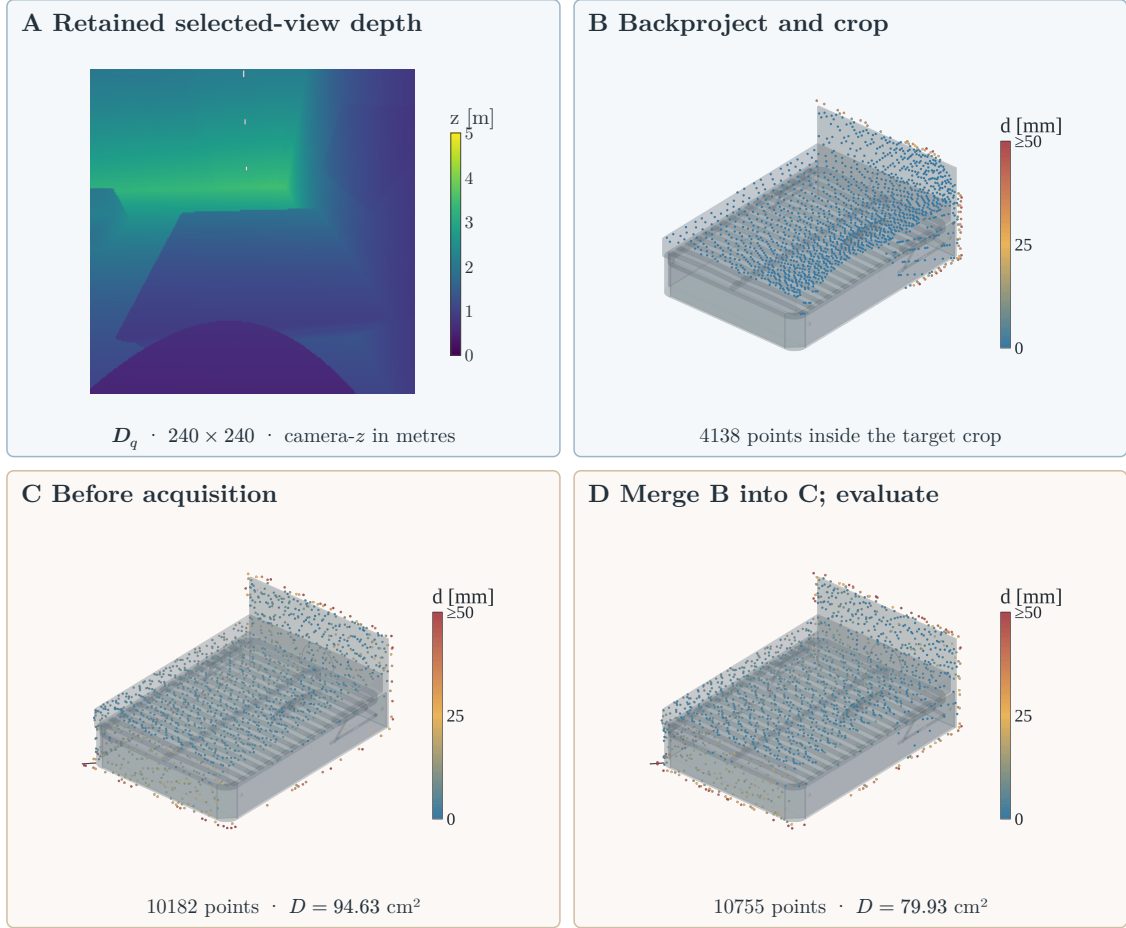
and measures which target faces lack nearby reconstructed evidence. Their sum is the target error

$$\Delta_t^e = d(\mathcal{P}_t^e, \mathcal{M}_e^{\text{GT}}) = D_{P \rightarrow M, t}^e + D_{M \rightarrow P, t}^e \quad (30)$$

with units of square metres. Current and candidate-augmented clouds pass through the same deterministic fusion and point-cap operator, so the difference isolates the candidate evidence under a fixed update rule. The resulting estimand is defined on the capped point set, fixed target mesh, and declared fusion operator. The pinned PyTorch3D implementation supplies the local distance above, with the unchanged 0.005 m² threshold in both directional kernels. Its version and threshold are part of the frozen metric protocol.

Figure 5 illustrates the two directional errors on a measured ASE target under an earlier whole-face crop and retained-depth protocol. Its values demonstrate that historical transition, rather than the current exact-box and native-render

measurement. The point colours show the point-to-triangle distances the scalar error also includes the reverse, equal-face completeness term.



$$D_{P \rightarrow M} : 1.43 \rightarrow 1.32 \text{ cm}^2 \quad D_{M \rightarrow P} : 93.2 \rightarrow 78.61 \text{ cm}^2$$

Transparent mesh: $\mathcal{M}_e^{\text{GT}}$ $\cdot$ points coloured by nearest-triangle distance.

Same target frame, 3D camera and distance scale; display subsets, full-cloud metrics.

Figure 5: Retained-depth reconstruction for ASE 82254/000008, selected candidate 55 at step 0. The privileged root cloud is reconstructed from the 20-frame ASE depth prefix. Coloured points are display subsets; reported errors use the complete cropped clouds and all 39,058 target-mesh faces. This historical illustration retains whole triangles with any vertex inside the target box expanded by 5 cm; it does not depict the applied exact-box, zero-margin measurement. The recomputed improvement is 0.15538. This uses separately retained 240×240 selected-view depth and does not reproduce the logged native-render RRI of 0.02639; it illustrates the measurement operations.

Most of the error reduction comes from target faces gaining nearby reconstructed points; the point colours show only point-to-surface distances and therefore do not display this completeness improvement directly.

The completeness reduction gives every mesh face equal weight. It is deliberately neither an area-weighted surface integral nor invariant to retessellation: subdividing part of an unchanged surface changes how much that region contributes. A distinct tessellation effect can change the accuracy term: subdividing a large filled triangle into faces below the area threshold changes which local distance branch applies, even when the planar surface is unchanged. Exact box clipping and retriangulation can create such small faces. The size of this effect on the recorded campaign’s labels or rankings has not been measured.

Validation TODO: Quantify small-face incidence and label/ranking sensitivity on retained campaign geometry before asserting practical invariance to retriangulation.

Source: Pinned PyTorch3D point-face and face-point kernels; operational metric regression fixtures

Gate: This is a sensitivity question, not an instruction to relabel historical evidence.

Open decision: If a different small-triangle threshold is considered for future collection, define and validate a prospective metric identity before adopting it. The existing threshold and frozen labels remain unchanged.

Source: Operational point-mesh metric protocol

Gate: An alternative threshold is unselected; no historical rerun is required.

The measured transition above shows the operational update on ASE data; the controlled fixture in [Figure 41](#) isolates the two asymmetric reductions and demonstrates the tessellation sensitivity. Together, they bound the interpretation of the operational estimand but do not invalidate comparisons that share one fixed mesh and scoring protocol.

3.4.2 Reliability under the Frozen Protocol

For fixed source state, target, candidate, and manifest, repeated oracle evaluation should reproduce both $\Delta_{t|i}^e$ and the induced candidate ordering. This requirement covers the complete measurement operator—rendering, backprojection, target cropping, fusion, capping, and numerical failure handling—because instability in any component changes the label. Deterministic code and a fixed mesh make exact replay possible, but RQ1 still requires repeatability measurements over the intended scene and target population.

The rollout store may additionally retain a selected-depth audit image at its declared storage resolution and precision. That payload documents and inspects

the chosen acquisition; it is not a claim that reloading the audit image reconstructs the exact in-memory scoring cloud, whose render resolution, spatial stride, masks, and floating-point path are governed by the scoring manifest. Replay claims must therefore distinguish exact stored label scalars from approximate geometry reconstructed from an audit payload.

Point-sampled Chamfer distance depends on how the reference surface is sampled; the selected equal-face point–mesh error instead depends on mesh tessellation. Thresholded precision/recall and volumetric measures answer additional geometric questions. Their definitions and comparison are retained in [Section C.11](#); they do not replace the selected outcome.

3.4.3 Construct Validity

The operational construct is **target-surface evidence after a declared acquisition and update**. ASE depth, the selected target crop, equal-face mesh reduction, and the fusion operator jointly define it. Perceived object quality and whole-scene completeness require different constructs. Two systems that share the final point–mesh formula but apply different state updates estimate different intervention effects.

The historical whole-face crop includes a mesh face when any vertex lies inside the target [Oriented Bounding Box \(OBB\)](#). It retains the whole triangle rather than clipping it at the box boundary: a face with one vertex inside can extend beyond the crop, while a crossing face with all vertices outside can be omitted. A geometry-invalid candidate is excluded by $m_{t,i}^{\text{act}}$. Measurement support is checked after preparing the candidate successors because joint voxel fusion can move a previously supported representative outside the target region. If any successor has no target-evaluation points, the current procedure rejects the target’s entire aligned candidate evaluation rather than assigning a numerical reward or retaining only its other measurable rows. Insufficient root support or an unusable render likewise makes the oracle outcome undefined and clears $m_{t,i}^{Q,h}$. Passing geometric feasibility and admitting a measurable reconstruction outcome are therefore distinct conditions.

Comparative policy evidence therefore requires the same target mesh and crop, ASE-depth source, renderer, backprojection, fusion, point cap, and metric configuration. Replacing ASE depth with MPS semi-dense evidence changes both the observation process and the error population and therefore defines a separate experimental outcome. Exact clipping at the box boundary also defines a different scoring surface from retaining whole triangles. Agreement under one crop rule cannot establish equivalence under another. Candidate-render optimizations must preserve the selected observation and its effect on later state preparation; agreement for isolated candidates alone does not prove agreement along an acquired trajectory.

The initial fixed-state renderer test established agreement at that state only. The first long-chain checkpoint failed and remains negative evidence. After the camera repair, the matched 16-acquisition test completed 20 attempts across five physical scenes, two nested seeds, and two rendering modes: ten qualified pairs and 160 matched decisions. Within this tested recipe, candidate outcomes, selected identities, full-scene selected observations, prepared histories, and terminal measurements satisfied the frozen renderer-equivalence contract. Chapter 6 reports the numerical checks and measured local costs.

Both rendering arms used the same exact-box, zero-margin scoring policy. This validates the optional rendering contrast on that finite roster; historical whole-face labels and exact-box labels remain different measurements. Its scope fixes float32 arithmetic, zero blur radius, 48×48 candidate rendering, the applied four-family mixture, disabled scene-wide RRI, and full-scene selected acquisition. Other meshes, cameras, policies, and scoring cohorts require their own source-bound validation.

3.4.4 RRI, Training Reward, and Endpoint Gain

The learning and evaluation quantities are derived from the same error sequence $\Delta_0^e, \Delta_1^e, \dots, \Delta_H^e$, but answer different questions. State-relative RRI asks how much one action improves the **current** state; target-root reward assigns additive credit along a chain; endpoint gain asks how much the target improved after a fixed

budget. Their denominators determine whether they can be accumulated or compared across states.

The VIN-compatible marginal diagnostic is

$$\text{RRI}_{t,i}^e = \frac{\Delta_t^e - \Delta_{t|i}^e}{\Delta_t^e + \varepsilon_{\text{err}}} \quad (31)$$

Its selected-chain running sum is

$$C_t^{\text{RRI},e} = \sum_{k=0}^{t-1} \text{RRI}_{k,a_k}^e \quad (32)$$

Each term divides the error decrease by the error just before that action, $\Delta_t^e + \varepsilon_{\text{err}}$. The running sum therefore describes accumulated local relative improvements. At a fixed state, all candidates share this denominator; their immediate ordering is the same as with root-normalized reward below. Along a trajectory the denominator changes, so the sum is a different objective from fractional initial-to-final error reduction.

VIN-NBV motivates evaluating views through relative reconstruction improvement [1]. To accumulate improvements along a rollout, this thesis uses the original root error Δ_0^e in every step’s denominator. For candidate i , the numerator is $\Delta_t^e - \Delta_{t|i}^e$, where $\Delta_{t|i}^e$ is the prepared target error after adding that candidate observation. When $i = a_t$ is selected, that after-error becomes Δ_{t+1}^e , giving the stored reward:

$$r_t^e = \frac{\Delta_t^e - \Delta_{t+1}^e}{\Delta_0^e + \varepsilon_{\text{err}}} \quad (33)$$

With unit discount, summing these rewards over a factual chain telescopes:

$$J_t^e = \sum_{k=0}^{t-1} r_k^e = \frac{\Delta_0^e - \Delta_t^e}{\Delta_0^e + \varepsilon_{\text{err}}} \quad (34)$$

Positive rewards indicate reduced error; negative rewards retain an observed worsening. No clipping removes them. With unit discount, intermediate errors cancel because every term uses the same root denominator. The identity requires a complete chain, each after-error to match the next before-error, and the same

measurement and preparation throughout. Discounting or clipping changes the trajectory objective and this equality.

Root normalization makes the outcome a fraction of the original error removed. Averaging it across tasks gives a unit of absolute error reduction more weight when the initial error is smaller. Absolute error increments would answer a different question about geometric improvement across targets. The selected study utility is fractional initial-error reduction; the population and aggregation weights still determine whose improvements this average represents. Comparing a policy with a named reference is a separate operation, defined below.

Open decision: RRC01 — Fractional initial-error reduction is selected. Specify the target/root population and aggregation weights, including the treatment of recurring objects and overlapping root contexts.

Source: Current fixed-root reward and endpoint definitions; reward/continuation synthesis, 2 October 2026.

Gate: Declare the study utility and population weighting; align Method and Design without relabelling historical outcomes.

The stabilizer ε_{err} keeps the arithmetic defined near zero. It does not establish that a very small root error is measured reliably. Errors before and after acquisition, and the distribution of root errors, remain necessary for interpreting the normalized rewards.

Validation TODO: RRC02 — Assess small-root normalization against measurement variability. Decide explicitly whether any reliability floor or exclusion is needed; either changes the utility or evaluated population.

Source: Fixed-root normalization and scoring preparation; the arithmetic stabilizer is not a measured noise floor.

Gate: Source-bound error/reliability audit and a stated population decision, without an invented numerical threshold.

For a realized chain, the finite discounted return defined in Foundations sums these fixed-root rewards over the requested acquisitions.

A complete stored h -step sequence samples the return of the collector that produced it; it is one outcome, rather than an expected or optimal value. A shorter recorded segment still describes its observed length. Missing or nonfinite increments cannot be read as observed zero improvement or used to certify a complete longer return. Foundations separates this cutoff from the finite query boundary and a genuine task end [19].

Open question: RRC04 — Under the declared action protocol, does an empty candidate set end acquisition, trigger a retry or prompt new proposals? Classify recorded end reasons as task stops or evidence cutoffs.

Source: Replay end reasons; finite-query/task/recording distinction in Foundations.

Gate: Declare stop/retry semantics and reason-specific eligibility for the requested horizon.

Validation TODO: RRC06 — Certify observed length, finite increments, transition continuity and scoring identity before reporting a complete return. A finite partial sum is not a completeness receipt.

Source: `aria_nbv/rri_metrics/returns.py` skips nonfinite terms; historical endpoint reconciliation does not certify every scoring cohort.

Gate: Cohort-specific completeness evidence and independent endpoint reconciliation, preserving valid negative rewards and partial outcomes.

Fitted-Q training uses a recorded transition differently. At one step, oracle gains provide direct targets. For a supported nonterminal transition at a longer horizon, the target combines the selected action’s measured reward with an estimate of greedy continuation from its recorded successor. The collector need not have chosen that continuation. This estimates actor-admissible value under the declared process; a bounded privileged search instead provides an oracle reference under its own action and information limits.

The fitted target in [Section 4.9.5](#) specifies how this continuation is selected and evaluated, including termination and missing successor evidence. Replay supplies the first transition; the backup estimates what to do next.

The [root-normalized endpoint gain](#) summarizes the fixed-budget reconstruction outcome:

$$J_e^{(H)} = \frac{\Delta_0^e - \Delta_H^e}{\Delta_0^e + \varepsilon_{\text{err}}} \quad (35)$$

It compares target error after the same acquisition budget H , relative to the initial error. To compare two policies directly, the [relative endpoint-error reduction](#) is

$$R_e^{(H)}(\pi; \pi_{\text{ref}}) = \frac{\Delta_H^e(\pi_{\text{ref}}) - \Delta_H^e(\pi)}{\Delta_H^e(\pi_{\text{ref}})}, \quad \Delta_H^e(\pi_{\text{ref}}) > 0 \quad (36)$$

Here π is the evaluated policy and π_{ref} is a named comparison policy. Both act on the same target task and acquisition budget, and their errors use the same reference geometry and reconstruction preparation. The ratio tells us what fraction

of the comparison policy’s remaining error π removes. Positive values mean lower endpoint error; negative values mean worse reconstruction. Report the endpoint errors alongside this optional ratio: it is undefined when the comparison error is zero and unstable when that error is very small. These comparisons do not require an oracle advantage or a separately trained myopic scorer.

Local RRI measures improvement relative to current error; fixed-root reward accumulates relative to initial error. For a complete, consistently measured undiscounted chain, its sum equals endpoint gain. Discount, clipping, crop, point preparation and evaluation geometry remain fixed in each comparison. The secondary log transform and its near-zero qualification are retained in [Section C.12](#).

3.5 Collection and Proposal Diagnostics

3.5.1 Candidate-Support and Geometric-Coverage Diagnostics

Candidate generation limits the alternatives a learner can rank. We distinguish attempted proposals, feasible actions, and feasible actions with finite oracle labels. Both retained counts and retained fractions matter: the same fraction can leave very different numbers of choices. Family-wise diagnostics reveal whether a large forward family hides the loss of target-conditioned alternatives. Failed roots remain zero-support observations rather than disappearing from the population. State statistics are averaged within physical scenes before scene means are combined; undefined directional measurements are reported separately.

The maximum measured immediate gain describes the best labelled feasible view. It is exact on a completely labelled candidate set, a lower bound when labels are missing, and undefined without labels. Low gain can reflect either limited proposals or little remaining improvement; it can also precede a useful later view. Nested proposal comparisons in [Section 5.4](#) separate some of these explanations. A no-render support audit cannot measure reconstruction gain.

Spatial diagnostics describe where proposals and selected cameras lie. They do not establish which target surfaces are visible: target-centre projection only checks

image framing. [Section C.13](#) gives the coordinate frames, formulas, denominators and summary table; [Section C.14](#) defines the separate angular-perturbation check.

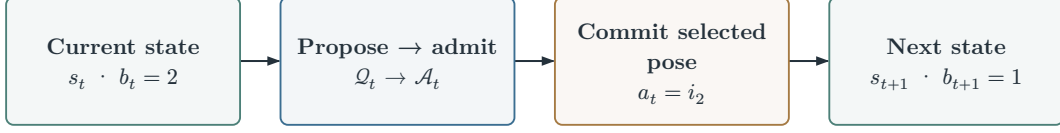
3.5.2 Rollout Branch Sampling and Dataset Impact

A **selection policy** maps the information available at a decision to a choice among feasible candidate views. We use the policy symbol π introduced in Foundations. When this rule selects the acquisitions recorded for training, we call it the **collection policy**. A scorer predicts values; the policy turns those values into choices, for example by taking the largest score or sampling from score-dependent probabilities. A collection policy may also use privileged oracle scores, which are unavailable to a learned actor at decision time.

Rollout recipes specify these collection choices. Uniform sampling explores support without a utility preference; one-step oracle greedy selects the largest immediate target gain; bounded oracle lookahead ranks finite action sequences; and temperature-softmax samples according to oracle score while retaining stochastic diversity. These recipes define the behavior distribution from which replay is collected. They serve as data-generation policies and reference strategies. Chapter 5 evaluates the learned policy on held-out tasks.

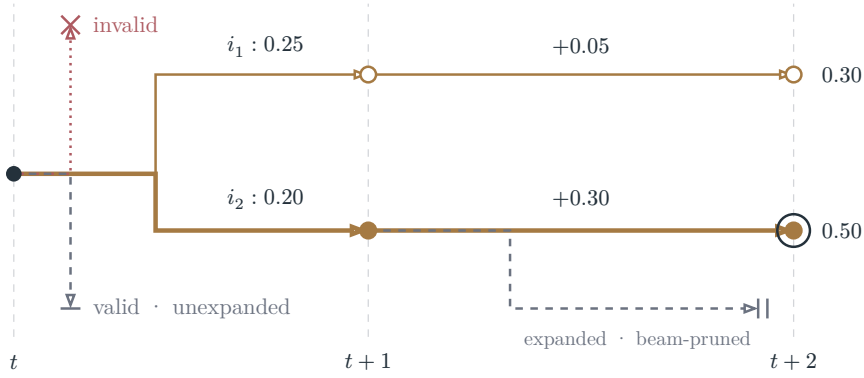
The first action of the highest-ranked bounded-lookahead chain can differ from the one-step greedy choice because lookahead ranks retained sequences rather than immediate gain alone ([Figure 6](#)). The persisted artifact contains beam-retained chains and their full per-step candidate shells; it is not an exhaustive counterfactual action tree.

A. Commit one selected transition



Fixed e , R_0 and root normalization · regenerate Q_{t+1}
 S0: selected poses · S1: selected rendered depth → surface points
 Factual budget: $b_t \rightarrow b_t - 1$ · backup query: $h \rightarrow h - 1$

B. Bounded search can reverse the immediate ranking



Constructed rewards, $\gamma = 1$. Thin and thick paths retained; ring marks the larger two-step return.

Figure 6: Selected transition and bounded oracle search. The chosen action extends factual pose history; target identity and root normalization remain fixed. S0 uses that pose history. When enabled, selected-only CF-GT depth is persisted and backprojected into the privileged S1 surface carrier. Acquisition budget and requested backup horizon have distinct roles. The constructed depth-two search preserves invalid, valid-but-unexpanded, and expanded-then-pruned branches. Its return ordering illustrates possible lookahead value, not measured planning headroom.

For stochastic collection, the temperature-softmax recipe converts each finite oracle score s_i into the robust logit

$$\ell_i = \frac{s_i - \text{median}(s)}{d_s \tau}, \quad (37)$$

$$p_{k,i} = \frac{m_{k,i}^{\text{elig}} \exp(\ell_i)}{\sum_{j=1}^{|s|} m_{k,j}^{\text{elig}} \exp(\ell_j)}, \quad m_{k,i}^{\text{elig}} \in \{0, 1\},$$

Here s contains the original finite candidate scores. The scale d_s is the interquartile range for at least four scores and the population standard deviation for smaller sets, with a machine-epsilon floor. Lower temperature τ concentrates probability on the highest eligible scores; higher temperature makes the eligible rows more

nearly equiprobable. At draw k , the binary mask $m_{k,i}^{\text{elig}}$ records whether row $i \in \{1, \dots, |s|\}$ remains eligible. Each draw excludes previously selected rows and applies the distance and optional diversity guards. The optional sibling-heading guard can reduce the number of retained branches; its vertical-view and fallback rules are specified in [Section C.8.2](#). The normalization statistics still use the original score vector, while the masked probabilities are recomputed ([Figure 34](#)).

The applied development collection retains temperature 1, a configured robust score-scale floor of 0, and heading-continuity weight 0. Machine-epsilon stabilization still applies. The scale floor bounds amplification of small score differences; it does not impose a minimum action probability. This selector adds no uniform probability mixture. The retained settings do not establish individually optimal parameters.

The resulting replay distribution is determined jointly by target sampling, candidate support, pruning, and rollout recipe. Reporting these factors isolates the population and behavior distribution on which learned ranking is assessed. Train-only pilots provide generation-cost and support-diversity evidence; held-out evaluation supplies policy evidence.

3.5.3 Proposal support and behavior distribution

A useful multi-step sequence requires both available actions and experience of their consequences. The successor generator must offer its later poses, and training must contain evidence about the relevant transitions. The generator determines the available actions; the collection policy determines which transitions are actually recorded.

For a fixed physical state, target, action, and measurement operator, changing other rows in the current candidate table does not change that action’s immediate reward. Changing future candidate support can change its longer-horizon value. Planning with stochastic action sets makes this dependence explicit by maximizing over available successor actions [\[36\]](#). The formulation in [Section 4.9.1](#) retains that distinction; tabular convergence results do not certify the present partially observed neural model.

A collector that always chooses the largest immediate gain may rarely execute such preparatory motions. Labelling every current candidate still leaves the unselected candidates without recorded successors. Exploratory collection can provide these missing transitions, including low-gain or redundant views that help distinguish useful sequences. It cannot compensate for a generator that never offers the revealing second view. This is why proposal opportunity and continuation coverage are measured separately. A recorded suffix measures the collector’s choices. A fitted target can instead estimate a better successor choice, using the off-policy mechanism introduced in Foundations and specified in [Section 4.9.5](#). It still relies on evidence of the first action’s consequence and on coverage of useful successor actions; imperfect behavior alone does not invalidate learning, and densely scoring current poses does not supply their unobserved future states.

3.6 Dataset and Storage Semantics

[Section 3.2](#) defines which quantities are causal actor information, privileged causal evidence, and oracle-only outcomes. The replay dataset preserves those roles by recording where each observation came from, which action produced a successor, and which labels are unavailable. The replay dataset is organized around decision-state relations. Each state binds an information state s_t and target e to a candidate table $\mathcal{Q}_t$, row-specific feasibility and label masks, oracle outcomes where defined, and the selected successor. References between these records connect each source to its target tasks, retained trajectories, decision states, and candidate rows.

One logged source can support several target tasks. Each task can be explored by several rollout recipes. A retained chain contains ordered selected steps. Each step owns the full finite candidate shell and designates exactly one selected row. This layout preserves two forms of supervision. Candidate rows record the counterfactual outcomes available at one state; only the selected row a_t links that state to the factual successor s_{t+1} . The first supports one-step ranking on labelled rows, becoming dense only when the declared action shell is completely labelled. The second supports finite-horizon returns along a causal trajectory.

Here, **factual** means the selected, recorded branch of the simulated process; it does not mean a physically deployed acquisition. A [rollout transition](#) links a selected action to this branch’s successor.

A single-branch recipe realizes one trajectory under its implemented policy π^{oracle} . This does not limit the oracle to scoring one immediate candidate: a state can retain labels for many hypothetical actions while only its chosen action supplies the next acquired observation. The general collector also supports several selected children per state, depth-dependent branch counts, stochastic branch-count choices, and beam retention. Each retained chain then has its own selected path and causal observation prefix; a sibling’s observation does not enter that chain. Such a tree supplies more recorded continuations, but it does not become an exhaustive tree merely because its immediate labels are dense. Specialized collection studies can restrict this general capability to one retained branch. Chapter 5 specifies what sampled returns and retained search outcomes can establish.

The stored relations and their materialized training view are shown in [Figure 42](#) and [Table 25](#). They preserve the scientific distinction between labels at a state and observed transitions from that state.

3.6.1 Immutable Sources and Causal Replay

An immutable source store owns the logged root substrate: geometry, reference pose, trajectory, target assets, and the configured visual evidence. Optional scene-level oracle products form a separate capability. A root can therefore support target-specific rollout generation without already containing scene-RRI labels or candidate renders. Replay references the immutable root identity and adds target tasks, per-state candidate shells, selected-action successors, and recipe provenance. Changing the rollout recipe creates new replay over the same initial evidence.

Root preparation can retain visual evidence, geometry, boxes and trajectory without scene-level oracle products. Missing scene labels remain absent; consumers requiring them must reject the input rather than substitute zero. Target-specific labels and selected observations may be generated later. Reference boxes remain

privileged regardless of the preparation profile. Storage capability therefore does not establish actor-visible learning admission.

Finalized fragments retain immutable content identities, so extending a collection preserves existing replay references. Fragment recovery and schema compatibility are documented with the storage implementations; the measured input compatibility is recorded in [Section C.2](#).

Every selected transition records the chosen action and its successor linkage. When selected-depth retention is enabled, replay also stores a calibrated depth raster for each selected action. The ordered rasters record the retained selected-depth prefix for inspection or an admitted depth-input profile; their resolution and precision need not reproduce the oracle scoring cloud exactly. Source role stays attached to each raster: mesh-rendered depth remains privileged even after selection. Storing a raster does not require the scorer to consume it. The pose-only control tracks the selected camera poses while keeping its root visual evidence fixed. Unselected candidate renders supply oracle labels but never become part of that branch’s acquired history.

Within a candidate shell, i denotes an arbitrary row. After selection, $a_t \in \mathcal{A}_t$ denotes the one full-shell row retained by a particular factual branch, so its endpoint is q_{t,a_t} rather than the generic $q_{t,i}$. For that branch, the stored data-generation update is

$$s_{t+1}^{\text{S0-pose}} \leftarrow \mathcal{T}_{\Xi}(s_t^{\text{S0-pose}}, a_t), \quad a_t \in \mathcal{A}_t \quad (38)$$

The pose-only update preserves root evidence and target identity, appends the selected pose, consumes one acquisition and generates the next candidate set. The expanded update and index-derived seed lineage are retained in [Section C.6](#). Seed identity is tied to the selected row path, so reordering physical candidates can change later proposals even though a scorer should remain equivariant on a fixed candidate set.

Open decision: H434-SEED01 — Decide whether replay identity should remain tied to stable full-shell indices or be invariant to reindexing the same physical poses. The latter requires a different seed identity and separate protocol validation.

Source: Replay state child-path identity and proposal seed derivation

Gate: A declared lineage identity with permutation tests; no dataset or seed change is implied by this annotation.

A visual successor adds a separately specified observation operator and state update for RGB, depth, EFM3D features, or surface memory. The rollout recipe retains the chains it selects or keeps in its beam. Matched oracle endpoint re-evaluation then supplies a common fixed-budget outcome for policy comparison.

3.6.2 Missingness and Derived Views

Every candidate row carries separate predicates for materialization, actor-feasible action $m_{t,i}^{\text{act}}$, finite value label $m_{t,i}^{Q,h}$, and factual successor m_t^{succ} . These predicates distinguish a materialized row, action admissibility, [label availability](#), and [successor availability](#). A padded row represents no proposed action; an unlabelled feasible row represents an action whose gain is unknown. A training objective may select the intersection of the required predicates; the stored fields preserve the reason each row enters or leaves that population.

Model-ready arrays are deterministic projections of the normalized store. They pad candidate shells, expand a chain into state queries, and cache features while preserving source identity, target identity, state order, masks, requested horizon, and label provenance. Several states from one chain may share a batch, but each query receives only its own causal prefix.

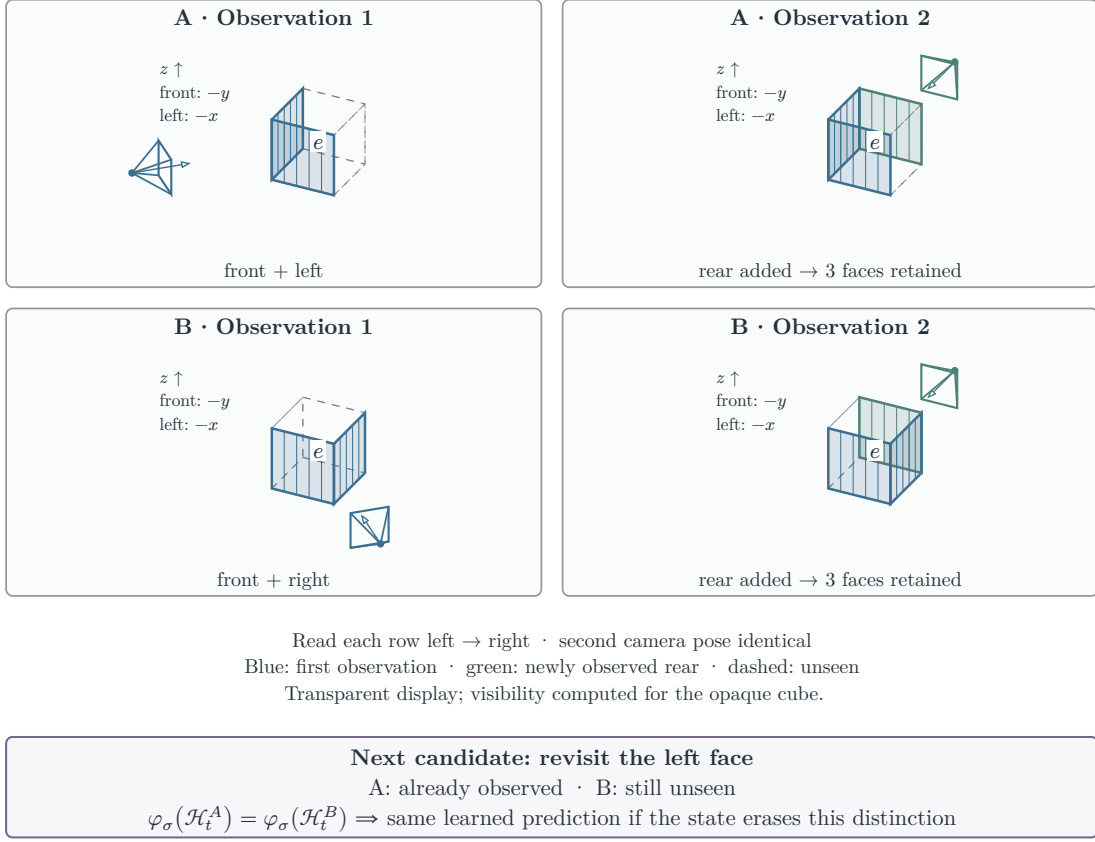
Reproducibility requires content identity and transformation identity. The source manifest fixes scene, snippet, target, geometry, and oracle products; the rollout manifest fixes proposal profile, pruning, recipe, seed, and retention; the training view fixes tensorization and supported horizons. Coverage, failed roots, family survival, undefined metrics, source-role counts, and exclusions define the population to which a learned value claim applies. Compression, chunking, hashes, and execution commands complete the reproducibility record. Development bandwidth pilots contribute cost estimates; held-out policy evidence comes from the evaluation protocol in Chapter 5.

4 Method

Chapter 3 defined the controlled decision process; this chapter specifies the finite-candidate value model used within it. At each realized state, the model receives a causal scene representation, a target description, a finite set of candidate views, the selected-view history, the remaining acquisition budget, and a requested prediction horizon. It assigns one conditional value to every materialized candidate. A separate hard feasibility mask determines which of those candidates the policy may select.

The compact baseline follows one forward path. It pools the fixed root scene, encodes the target pose and size, and summarizes only earlier selected camera poses. Root- and current-relative candidate poses form a physical embedding; target-relative geometry turns it into a value query. That query attends to five shared tokens for scene, target, history, budget, and horizon. A decoder combines the query, attended context, and their elementwise product into a scalar return prediction. The external action mask then restricts selection. Selected rendered depth enters a separate privileged state control, described below; it is not part of this pose-only baseline.

The ideal value conditions on the complete actor-visible observation and action history, whereas the learned scorer receives a finite representation of that history. This compression must retain the distinctions needed by the decision: which target surfaces were observed, which space is known free, and which continuations remain possible. For example, two histories may end at the same camera pose after viewing the rear surface while only one has previously observed the left surface. A representation that maps these histories to the same state cannot assign them different revisit values. This loss of a decision-relevant distinction is the aliasing risk illustrated in [Figure 7](#).



Fixed target, candidate, budget and horizon. Constructed counterexample, not a measured model failure.

Figure 7: Schematic counterexample to state sufficiency. For a fixed target, candidate, remaining budget, and horizon, two observation histories may have different ideal revisit values yet share a lossy representation. The learned scorer then emits the same prediction for both. This is a conceptual possibility, not a measured failure of a particular trained model. The sufficient condition for Bellman recursion is stated in [Section 4.9.1](#).

Bellman recursion adds the expected reward of the next acquisition to the expected value of subsequent choices [15, Sec. 3.5]. For a fixed target and protocol, a compressed state is sufficient for this calculation if histories encoded alike imply the same expected immediate reward and the same distribution of later decision opportunities under each admissible candidate pose. Here, those opportunities include the next representation, remaining budget, valid candidates and stopping. Matching the full reward-and-successor distribution is a stronger condition than preserving a particular expected return [24, Sec. 2, Theorem 1; App. A.2.2]. [Section 4.9.1](#) formalizes this stronger condition and its limits. The following sections examine which acquired distinctions each implemented representation retains.

Geometry enters through relations rather than absolute world coordinates. Candidate poses are expressed from the rollout root and current camera, the target is expressed relative to each candidate, and candidate rows remain exchangeable. The proposed research architecture maintains one fixed-chart geometric volume: root evidence initializes it, and each selected observation updates only the causal successor state. A small shared spatial encoder runs on that state before independent target and candidate reads produce one feature per candidate. The current compact scorer and the current experimental volume are executable controls for this proposal; neither is renamed as the proposed architecture.

Realization	Scientific role and evidence boundary
Compact baseline	Implemented pooled root scene and selected-pose history; contract tests do not establish held-out value accuracy or policy benefit.
Experimental volume	Implemented spatial control with a shared small encoder and independent candidate reads; historical and corrected root-support profiles are distinct. Results identifies the corrected privileged Q1 fit.
Corrected and calibrated volume	Calibrated target/ray reads are proposed work on the corrected support profile. Selected mesh depth makes the first study a privileged representation diagnostic.
Static root state	RQ2 control with static root scene evidence; an actor-visible policy claim also requires actor-visible target and action support.
Actor-observed update	RQ2 comparison in which selected actor observations update state under matched information; privileged depth cannot establish this claim.

Table 2: Architecture roles are separate from implementation maturity and scientific promotion. The diagnostic volume study does not redefine the core research questions.

Open decision: Choose the central Method realization for the completed thesis: the compact baseline, the corrected privileged volume, or an evaluated actor-observed update. Keep the implemented paths and their measured evidence distinct from the proposed calibrated readout; the current results do not select an actor-visible winner.

Source: RQ2; [Table 2](#); [Chapter 6](#)

Gate: A named realization with matched information access and corresponding evaluation determines the final chapter emphasis.

Immediate labels supervise every evaluable feasible candidate; multi-step training also needs the selected action’s factual successor. The learning section explains that difference after the complete forward path and its representation controls.

The proposed comparisons test information retention and value prediction; none currently establishes held-out actor-visible policy benefit.

Status axis	Question answered
Scientific role	Is this the selected method, a matched control, or a proposed extension?
Implementation	Does the mechanism have an executable interface, or is it still planned?
Evidence	Has a contract been tested, or has a held-out scientific result been established?

Table 3: Independent status distinctions used in the development draft. An executable interface and a passing contract test do not establish held-out policy benefit.

Remove or rewrite before submission: Complete held-out evaluation of the selected realization before presenting it as an evaluated method. Preserve the distinction between tested interfaces and demonstrated value accuracy or policy benefit.

Source: this chapter and its source-owner decision gates

Gate: frozen objective, source-matched controls, and end-to-end acceptance evidence

The detailed path now follows scene evidence, geometric and history encoding, shared contexts, interaction, and scalar prediction. Representation controls follow this complete computation; replay then supplies the learning targets.

4.1 Current and Intended Actor State

Implementation: partial **Evidence:** pending

The selected $s_t^{S0\text{-pose}}$ baseline and the independent actor-observed input descriptor path are implemented and tested. The current selected experiment still uses privileged ground-truth target geometry and privileged oracle action support; no frozen actor-visible target, action-support, or causal observation-updated state evidence yet supports the scientific target.

Literature: [6], [16]

Promotion gate: freeze and evaluate an actor-visible target corpus and action-support protocol; implement the causal state-update contract; pass source-dropout, no-future-observation, leakage, and held-out task-sufficiency tests

Development source: `aria_nbv/aria_nbv/targets/protocol.py`; `aria_nbv/aria_nbv/targets/selection.py`; `aria_nbv/aria_nbv/oracle/pipelines/campaign.py`; `aria_nbv/aria_nbv/oracle/pipelines/rollout_dataset.py`; `aria_nbv/aria_nbv/data_handling/qh_data/views.py`; `aria_nbv/aria_nbv/vin/modules/qh_scene_encoders.py`; `aria_nbv/aria_nbv/rollouts/storage/qh_reader.py`; `aria_nbv/aria_nbv/vin/models/target_finite_horizon.py`

The `rollout state` contains the information available before action a is selected; it does not include every quantity stored in the same replay row. Oracle target gains, mesh crops, ground-truth associations, current all-candidate renders, and labels are excluded from the scorer graph. Three independent protocol axes determine this information boundary. The target-source axis distinguishes privileged ground-truth geometry from a target descriptor constructed from actor-visible observations with explicit source and construction provenance. The dynamic-state axis distinguishes the pose-only $s_t^{\text{S0-pose}}$ carrier from a state updated with selected observations. The action-support axis distinguishes the current mesh-derived oracle support from actor-observed support or a separately calibrated learned-feasibility route. Sharing tensor shapes or field names across these axes does not make their scientific claims interchangeable.

The current baseline and the intended actor representation retain the same decision interface,

$$\mathbf{h}_{t,e,i} = \text{Read}(\Phi_t^{\text{scene}}, \mathbf{h}_e^{\text{tgt}}, q_{t,i}, \mathbf{H}_t^{\text{pose}}, b_t, h) \quad (39)$$

where b_t is the factual remaining acquisition budget and h is the requested prediction horizon. They differ in what the state makes available to that read. The current root component contains semidense evidence and lossy global moments of locally supported EFM3D features. Its dynamic component is only the strictly causal selected-pose prefix. Candidate regeneration updates the reference pose, prefix, remaining budget, and action table; it does not imply a new RGB observation, a refreshed EFM3D field, or fused selected depth.

Component	Current realization	Intended actor representation
Root scene	immutable semidense points and global moments of locally supported EFM3D features	immutable actor-visible evidence with explicit finite support and missingness; absent evidence must not mean observed free space
Target	selected experiment: privileged ground-truth target geometry; an independent actor-observed input admission and I/O path is implemented but not frozen or evaluated	actor-observed input: actor-visible identity and geometry with support, source, construction provenance, and explicit matching failures
Dynamic state	$\mathbf{H}_t^{\text{pose}}$ only	RQ2 comparison: one fixed root-chart volume updated strictly causally from selected actor observations, with observed surface, observed free, unknown, support, source, and recency kept distinct
Action support	privileged oracle action support: mesh-derived physical-validity support	actor-observed action support, or a learned-feasibility route with frozen calibration, threshold, abstention, and false-valid evidence
Decision context	$\mathcal{Q}_t$, b_t , and h under the named oracle-support protocol	the same finite-candidate interface under actor-visible target, state, and action-support protocols

Table 4: Current baseline and scientific targets. The right column describes actor-visible target, support, and observation-updated state requirements for RQ2; it is a promotion contract, not an architecture already implemented or a result obtained.

Depth and camera information have four distinct roles. Root evidence initializes the shared pre-acquisition state under its declared source protocol. A previously selected depth observation may update only the next causal state; selected mesh depth remains privileged. Current candidates supply actor-known poses and calibration, without their unselected renders entering actor features. Counterfactual renders of those candidates supply offline labels only. Scoring a current shell therefore uses one shared causal state, rather than one GT-rendered successor volume per candidate.

The two evidence sources also differ in what an update can contain. Historical RGB-derived EFM3D features carry learned appearance and spatial context at the root. Selected rendered depth contributes calibrated surface locations and ray-based free-space evidence, but no new RGB appearance. Fusing that depth with the root carrier therefore enriches geometry while its appearance evidence remains historical. A learned encoder can transform the combined evidence; it cannot turn

a depth-only acquisition into a newly observed image feature. This asymmetry motivates explicit source and availability channels, rather than treating root features and selected depth as interchangeable modalities.

The [causal observation state](#) specifies the actor-visible information requirement; the static-root representation remains a control. For the proposed fixed-chart volume, an actor-visible update is admissible only if its deterministic fusion rule exposes source and availability, uses the selected actor-visible observation and no future or unselected render, and preserves the distinctions in [Table 4](#). Source-dropout, no-future-observation, and leakage tests establish the information boundary. Held-out task-sufficiency evidence must then show that the retained distinctions help predict target-specific value.

4.1.1 Sufficiency and the local-evidence limit

A decision state is task-sufficient here only if it preserves the distinctions needed to predict target-specific future return. Under partial observability, this is an empirical requirement on the representation, not a claim that the study solves a general belief-state POMDP [\[16\]](#). $s_t^{S0\text{-pose}}$ cannot satisfy that condition by construction when two histories share poses but reveal different surfaces, occlusions, or free-space evidence. It is therefore a deliberately compact baseline and an interface test, not a claim that pose history is a complete reconstruction state.

[EFM3D](#) makes the representation problem concrete. Its [EVL](#) field has a known pose and finite extent [\[6\]](#). Evidence inside that chart can retain local spatial structure, but a target surface or candidate frustum outside the chart has no corresponding lifted feature. Missing lifted features do not by themselves make a physically feasible candidate invalid or imply negligible reconstruction gain. Evidence coverage describes the representation; action validity follows the declared feasibility checks. Pooling the chart into global moments removes a second distinction: two states with different spatial arrangements can acquire the same summary. A failure of the selected compact state can therefore arise from absent evidence or from evidence discarded by compression; value learning alone does not distinguish those causes.

The scene path in [Figure 8](#) connects these two losses to the actual computation. Calibrated image features are lifted into a finite voxel chart and combined with surface and free-space evidence. The selected compact carrier then pools field statistics and root-frame point statistics. Its input may contain spatial structure that the pooled vector no longer retains.

tinct. The first selected-depth diagnostic uses privileged mesh renders; admitting selected actor observations requires a separate source-bound path. A small shared 3D encoder derives features from the updated evidence. Independent candidate queries then read the target region and intervening rays without reducing the scene to one global vector. Finite chart extent and resolution remain explicit limitations; sparse or re-centred storage is deferred until measurements show that either boundary, rather than the readout, is limiting.

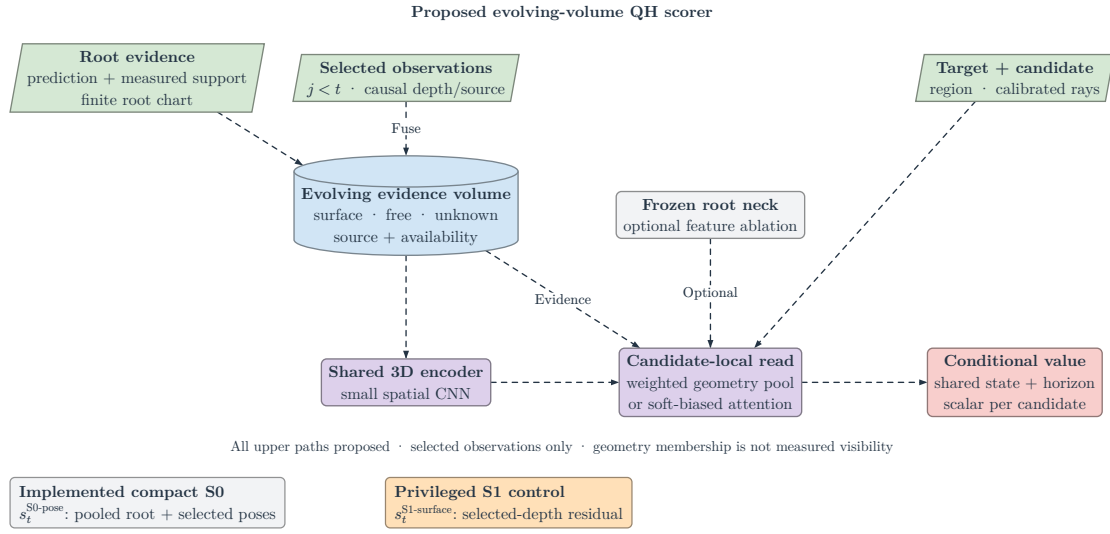


Figure 9: Proposed state path for candidate-value prediction. Root evidence initializes one fixed-chart volume; selected observations update only causal successors. A shared spatial encoder is followed by independent candidate reads and the existing scalar value interface. Geometric pooling and one spatial-attention block are matched readout hypotheses; a frozen EVL feature carrier is a separate late-fusion ablation. The lower strip shows the implemented compact S0 scorer and privileged S1 selected-depth control; the latter currently uses mesh-rendered depth, while actor-observed selected depth requires separate admission. The current experimental volume is described in Figure 15. This graph states a research architecture, not an implemented or superior method.

4.2 Target, Candidate, and History Encoding

Implementation: implemented **Evidence:** pending

The selected encoder materializes one target-conditioned query per candidate from root evidence, local relative geometry, H0 selected-pose history, remaining budget, and requested horizon. Its tensor contract is tested; scientific usefulness remains pending.

Literature: [18], [25]

Promotion gate: preserve local-frame, mask, source, and strictly causal prefix tests; evaluate held-out ranking

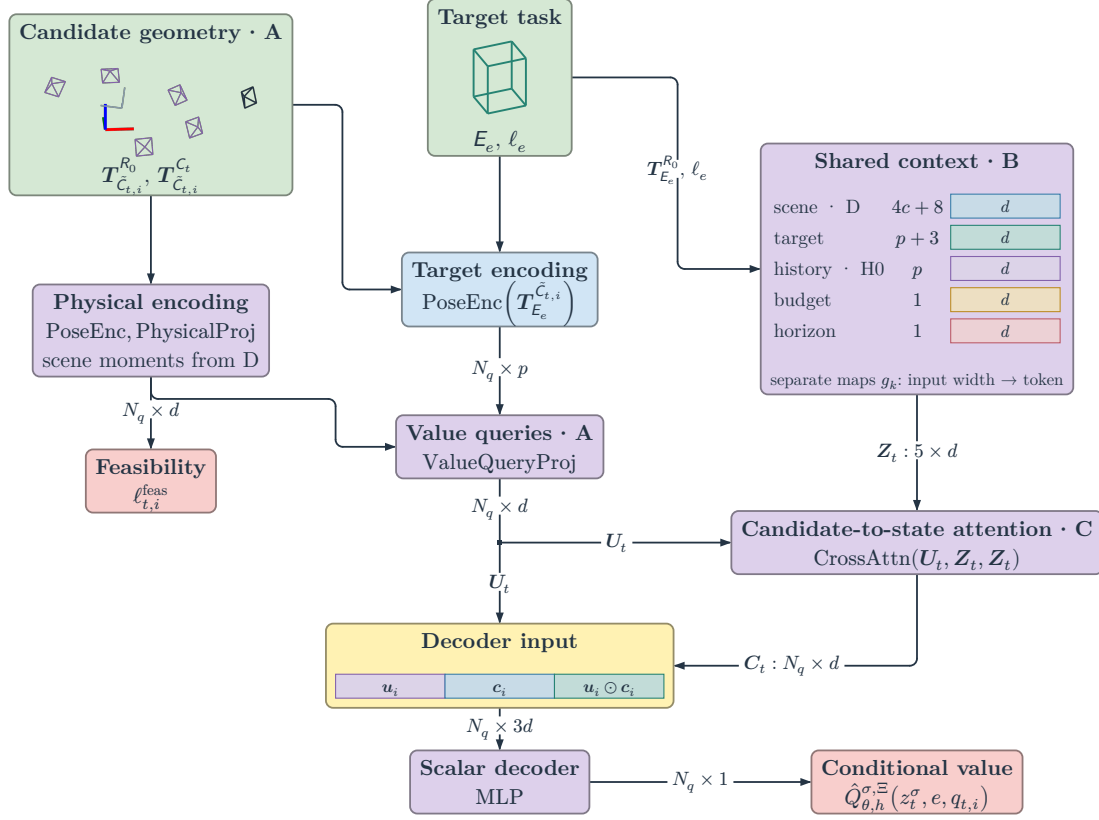
Development source: aria_nbv/aria_nbv/data_handling/qh_data/views.py; aria_nbv/aria_nbv/vin/mod-els/target_finite_horizon.py; aria_nbv/aria_nbv/vin/modules/qh_history_encoders.py; aria_nbv/tests/vin/test_target_finite_horizon.py; aria_nbv/tests/vin/test_qh_history_encoders.py

For target e and candidate $q_{t,i}$, the model predicts how much the target’s reconstruction can improve over the requested horizon, given the causal scene evidence, target state, selected history, and remaining acquisition budget. The scorer input is the structured relation

$$\mathcal{J}_{t,e} = \left(\Phi_t^{\text{scene}}, \mathbf{T}_{E_e}^{R_0}, \ell_e, \left\{ \mathbf{T}_{\tilde{C}_{t,i}}^{R_0}, \mathbf{T}_{\tilde{C}_{t,i}}^{C_t}, \mathbf{T}_{E_e}^{\tilde{C}_{t,i}}, m_{t,i}^{\text{cand}} \right\}_{i=1}^{N_q}, \mathbf{H}_t^{\text{pose}}, b_t, h \right) \quad (40)$$

Scene evidence records what is available about the environment, while the target state identifies the object whose reconstruction is evaluated. Selected history summarizes earlier actions. The remaining budget records how many acquisitions are still available; the requested horizon states how many future rewards the prediction covers. These five contexts are shared across candidate rows. Each candidate adds its own root- and current-relative pose and its relation to the target, so every row queries the same state without reading another candidate row.

The physical and value paths separate in [Figure 10](#). Physical feasibility reads only root-scene evidence and root/current-relative candidate geometry. The value query also reads the target relative to that candidate and the five shared state tokens. Both raw predictions precede authoritative action masking. The auxiliary feasibility loss is separately weighted when enabled and is disabled by default in the selected configuration; its logit does not multiply the conditional value.



Default widths: $p = 32$, $c = 5$, $d = 128$. Each candidate reads five state tokens independently.
ASE 83269/000056 geometry; shell reconstructed without validity filtering.

Figure 10: Independent candidate scoring in A1. A forms queries and predicts target-independent feasibility; B embeds each shared context into one equally wide state token, including the scalar conditioning variables budget and horizon; C attends to those tokens and decodes the query, context and their product. D supplies pooled root evidence. Insets use ASE 83269/000056 geometry. The candidate shell reconstructs unfiltered proposals around the measured root with stored calibration; five separated proposals are displayed. Dark ink identifies the recorded candidate; RGB and grey axes denote the target and root frames. These are input illustrations, not a measured admissible set or a declaration of the training target source. Tensor widths use configuration defaults. Action masking remains outside the scorer.

Panels A–D lead to the detailed candidate encoding (Figure 11), shared state tokens (Figure 12), attention and decoder (Figure 13), and root evidence (Figure 8), respectively. These panel roles differ from the A0/A1 architecture controls.

Source and availability masks distinguish absent evidence from measured zeros; padding is separate from physical invalidity; supervision and audit lineage never enter the learned query.

The architecture atlas (Chapter C) connects these local views to a full-size computation graph. Its paired view-footprint illustration (Figure 37) distinguishes

alternative candidates at one state from the ordered selected acquisitions of a stored rollout.

4.2.1 Logical data roles

A normalized batch may collate actor inputs, supervision, transition linkage, and audit lineage, but tensor colocation does not grant the model access. The Method therefore describes the logical contract rather than enumerating storage columns:

Logical role	Scientific content	Model access
Scene and target state	root actor evidence, source-bound target descriptor, $\mathbf{H}_t^{\text{pose}}$, b_t	candidate-value input
Candidate table	stable row identity, local pose relations, materialization, physical validity, reason and proposal provenance	geometry and materialization only; $m_{t,i}^{\text{act}}$ is external
Value query	scalar h with $1 \leq h \leq b_t \leq H_{\max}$	input only when bundle-supported
Candidate supervision	root-normalized gain, fitted target, $m_{t,i}^{Q,h}$	loss admission after prediction
Selected transition	factual selected row, reward, terminal flag, successor identity	Bellman linkage only
Audit lineage	scene, target, store, generator, policy, seed, source, and configuration identity	never a learned feature

Table 5: Logical data roles at the dataset–model boundary. Storage readability and batch colocation do not imply actor visibility.

This separation matters for causal interpretation. Target-source provenance is experiment identity rather than a shortcut feature; generator family is proposal-support evidence rather than a semantic action label; and oracle returns supervise $\hat{Q}_{\theta,h}^{\sigma,\Xi}(z_t^\sigma, e, q_{t,i})$ only after that value has been emitted. The same replay loader can therefore support privileged controls and actor-visible experiments without pretending that they estimate the same value function.

4.2.2 Target-conditioned query

The selected target protocol supplies a root-relative target pose and metric OBB extents. The target token is therefore

$$\mathbf{h}_e^{\text{tgt}} = \text{TargetProj}\left(\text{concat}\left(\text{PoseEnc}\left(\mathbf{T}_{E_e}^{R_0}\right), \ell_e\right)\right) \quad (41)$$

The pose locates and orients the requested object relative to the root; the extents describe its size. Together they let the same scene evidence support different requests: a view directed at a small cabinet and one directed at a larger adjacent

cabinet need not have the same value. Expressing the target from each candidate then relates that view to the object’s position and sides. Class, detector confidence, and optional observation-support summaries remain descriptor or audit information unless the selected encoder admits them.

The source protocol that produced this geometry identifies the experiment; it is not another model input. Current oracle tasks use ground-truth-derived target geometry. This privileged source limits deployability even though gains, meshes, associations, and audit lineage remain outside the prediction graph.

4.2.3 Relative candidate geometry

Canonical world poses remain reproducibility facts, but the selected scorer encodes each candidate both relative to the rollout root and relative to the factual current camera. Its physical trunk is

$$\begin{aligned} \mathbf{p}_{t,i}^{\text{root}} &= \text{PoseEnc}\left(\mathbf{T}_{\tilde{\mathcal{C}}_{t,i}}^{R_0}\right), \\ \mathbf{p}_{t,i}^{\text{current}} &= \text{PoseEnc}\left(\mathbf{T}_{\tilde{\mathcal{C}}_{t,i}}^{C_t}\right), \\ \mathbf{h}_{t,i}^{\text{phys}} &= \text{PhysicalProj}\left(\text{concat}\left(\mathbf{p}_{t,i}^{\text{root}}, \mathbf{p}_{t,i}^{\text{current}}, \Phi_t^{\text{scene}}\right)\right) \end{aligned} \quad (42)$$

The intermediate pose features separate root-relative placement from motion relative to the factual current camera before scene context is fused. The conditional-value query then adds the target pose expressed from that candidate:

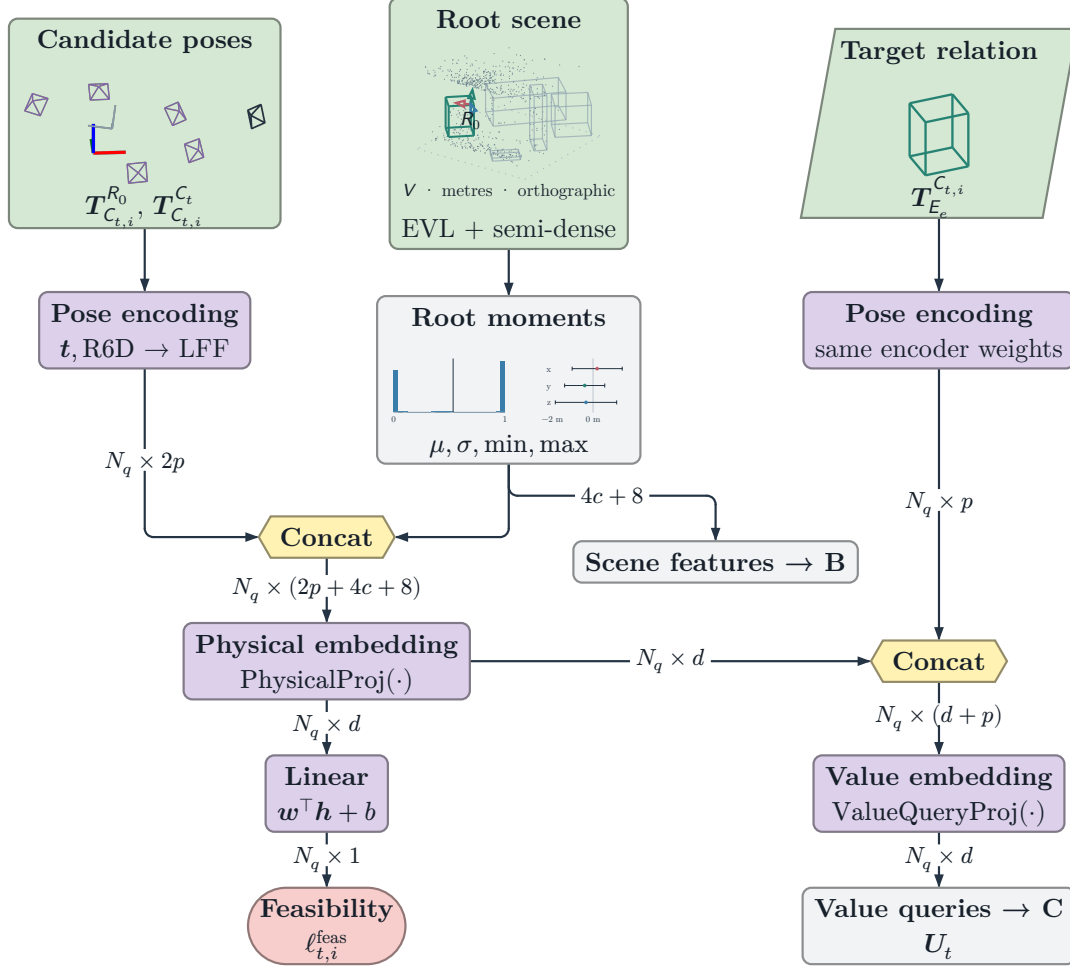
$$\mathbf{x}_{t,i} = \text{ValueQueryProj}\left(\text{concat}\left(\mathbf{h}_{t,i}^{\text{phys}}, \text{PoseEnc}\left(\mathbf{T}_{E_e}^{\tilde{\mathcal{C}}_{t,i}}\right)\right)\right) \quad (43)$$

This factorization removes arbitrary world origin while retaining the complete relative rotations and translations represented by the shared PoseTW encoder. It does not append handcrafted range, bearing, height, frustum, sampler-family, or generator-provenance features. The physical token feeds feasibility without target or horizon information; only the conditional-value query receives the candidate-from-target transform.

Figure 11 expands the two branches. The shared pose encoder maps the root- and current-relative poses before the physical projection the target-relative pose

enters only the value branch. Sharing pose-encoder weights does not merge the information available to the two heads. The rotation embedding uses six coordinates from the first two rows of the rotation matrix. Translation and rotation receive separate learned scales. Learnable Fourier features then map linear combinations of these coordinates to sine and cosine pairs before a small multilayer perceptron projects them. This periodic embedding gives the network a richer pose description; it does not itself guarantee geometric equivariance.

A · Candidate encoding and the two prediction branches



Defaults: $p = 32$, $c = 5$, $d = 128$; batch axis omitted. LFF: learnable Fourier features.
Shell reconstructed without filtering; root statistics measured.

Figure 11: Candidate encoding and the physical/value split (A). Root/current-relative poses and root moments form the physical embedding; candidate-relative target geometry supplies the additional value input. Tensor dimensions label the connections. The inset displays five separated, unfiltered proposals reconstructed around ASE 83269/000056 and its stored candidate; root moments are measured. Widths $p = 32$, $c = 5$, and $d = 128$ are configuration defaults.

4.2.4 Strictly causal pose history

For every realized state, the model input contains the complete $\mathbf{H}_t^{\text{pose}}$ for $k < t$, expressed from the factual current camera. Each selected pose is first mapped to a feature in that frame:

$$\mathbf{p}_{t,k}^{\text{hist}} = \text{PoseEnc}\left(\mathbf{T}_{C_{k+1}}^{C_t}\right), \quad k < t \quad (44)$$

The feature $\mathbf{p}_{t,k}^{\text{hist}}$ therefore retains the relative rotation and translation of view k as seen from the decision pose at t . The selected H0 encoder discards acquisition order after pose encoding. It applies a padding-aware arithmetic mean and projects the result to the shared state-token width:

$$\mathbf{h}_t^{\text{histH0}} = \text{HistProj}\left(\frac{1}{\max(1, t)} \sum_{k < t} \mathbf{p}_{t,k}^{\text{hist}}\right) \quad (45)$$

Thus H0 is invariant to permutations of the same selected-pose multiset. Both padding and the lack of acquired surface observations remain explicit: this pooled pose summary cannot replace a reconstruction state. An ordered pose-history control is described after the complete scoring path.

The current encoder implements the $s_t^{\text{S0-pose}}$ actor-state carrier described above. The actor-observed input path constructs its target token from detected OBBs, the actor-visible snippet transform, sample identity, and the world-frame reference rig pose; it does not require an oracle payload. The physical reference pose is actor input; its storage location does not make oracle labels or ground-truth boxes admissible features. Focused tests establish numerical identity with the neutral carrier and invariance to oracle RRI and GT OBB payloads. The resulting actor-only construction is implemented, but no frozen observed-target-protocol corpus or held-out policy evaluation exists. Target provenance and dynamic state remain separate axes. The proposed observation-updated state is described in [Section 4.1](#); the compact five-token scorer and current four-token experimental volume remain implemented controls.

The maximum horizon $H_{\max}$ binds data, model, and checkpoint. Remaining budget b_t records the factual state and h selects one return from $1 \leq h \leq b_t \leq$

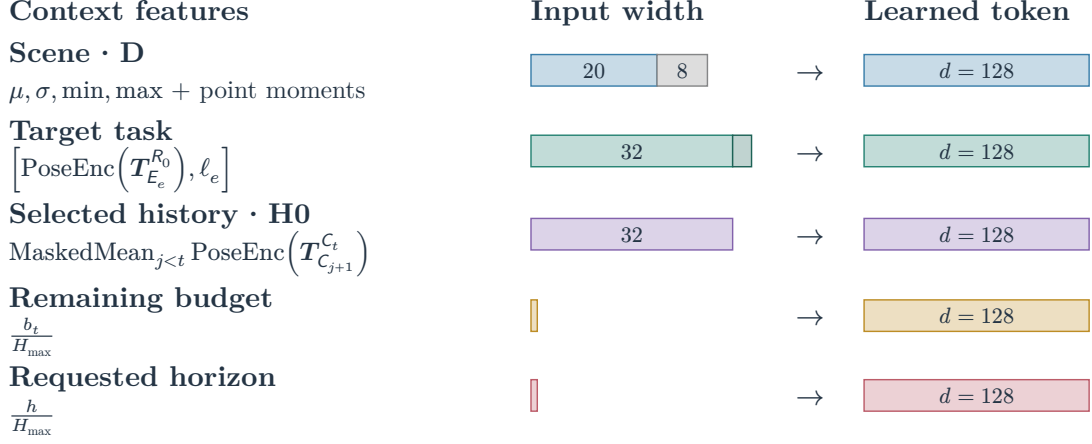
$H_{\max}$. Current bundles record trained horizons rather than budget–horizon pairs, so deployed inference requests the factual diagonal $h = b_t$. A manifest-bound pair gate is required before any off-diagonal $h > 1$ query is promoted beyond the syntactic scorer interface. The public interface scores one scalar horizon at a time and preserves candidate order in its $[B, S, N_q]$ output.

4.2.5 Shared state tokens

Figure 12 makes the distinction between input dimension and token capacity explicit. Each of the five contexts has its own learned projection into the same width. Budget and horizon therefore each occupy one complete token despite being scalar inputs. Equal token width does not imply equal importance. It only gives each context the same representational width; attention determines how a candidate combines their learned values, but does not by itself establish empirical feature importance.

This candidate-to-state interaction is distinct from the proposed spatial readout. In the compact A1 control, attention combines five already pooled state tokens. In the volume study, geometric pooling or candidate-local attention combines spatial samples from the target and intervening rays; the resulting candidate feature then joins target, history, budget, and horizon conditioning. Attention weights in either location are learned routing coefficients, not direct measurements of modality importance or physical visibility.

B · Five contexts, five independent maps, five tokens



Embedding family $g_k(x) = \text{LN}(\text{GELU}(W_k x + b_k))$

$$\text{Stack}_{k=1}^5 g_k(x_k) \rightarrow \mathbf{Z}_t \in \mathbb{R}^{5 \times d} \rightarrow \mathbf{C}$$

Inputs use one width scale; learned tokens use another. Scalar inputs each produce a full token.

Distinct weights per row; equal width describes capacity, not attention importance.

Default feature dimensions; the bars are not measured activations.

Figure 12: Five shared contexts become five equally wide tokens (B). Raw widths differ: field/point moments, target pose/extents, and selected-pose history are vectors, whereas normalized budget and requested horizon are scalars. Independent maps produce $d = 128$ features per token, stacked into $\mathbf{Z}_t$. The two bar scales encode dimensions rather than measured feature magnitudes or attention weights.

4.3 Candidate-to-State Interaction

To score a view, the model combines the candidate’s geometric features with five shared summaries: scene, target, action history, remaining budget, and requested horizon. Cross-attention allows each candidate to weight these summaries differently. The candidate embedding supplies the query, the five state tokens supply keys and values, and the resulting context enters the value decoder (Figure 10). Candidates do not attend to one another.

The implementation names this interaction A1. Its A0 control concatenates the same candidate embedding and ordered state tokens, then applies an MLP shared across rows. Both produce context vectors of the same width for the same

decoder. The comparison therefore changes how the same evidence is combined. It does not by itself match parameter count or optimization behavior. The choice of scene representation remains a separate factor (Table 4).

Implementation: implemented **Evidence:** pending

A1 candidate-to-state cross-attention is the selected interaction; A0 is its feature-matched independent-row control. Both pass the same executable acceptance tests. Their scientific comparison remains pending.

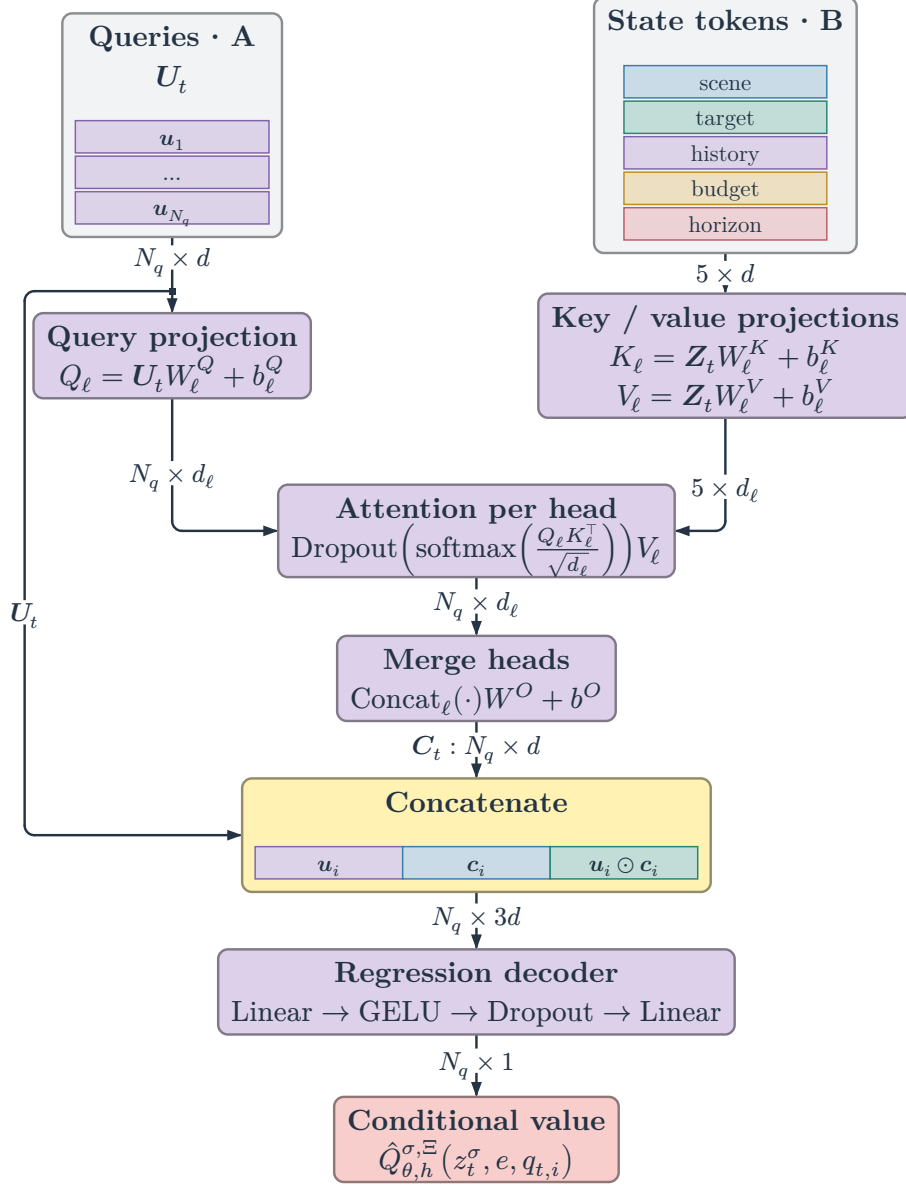
Literature: [25], [27]

Promotion gate: retain row-equivariance, invalid-row isolation, frame, source, mask-independence, and horizon tests; measure the A0/A1 control

Development source: `aria_nbv/aria_nbv/vin/models/target_finite_horizon.py`; `aria_nbv/aria_nbv/vin/modules/`
`qh_state_fusion.py`; `aria_nbv/tests/vin/test_qh_state_fusion.py`; `aria_nbv/tests/vin/test_target_finite_horizon.py`

Figure 13 exposes the candidate-to-context computation. Each head projects the candidate into a query and the five shared tokens into keys and values. The candidate retains a direct path to the decoder alongside the attended context and their elementwise product; attention can therefore modulate the context without replacing the original candidate representation.

C · Independent candidate queries read shared state



Softmax is over five state tokens; candidate rows never serve as keys or values.
 ℓ indexes attention heads; tensor drawings are schematic.

Figure 13: Independent candidate-to-state attention and decoding (C). Softmax normalizes over the five state tokens, separately for each candidate and attention head. Merged head outputs form the context. The regression decoder receives the candidate query, context, and their elementwise product. The head index ℓ is distinct from the requested horizon h ; token and tensor glyphs depict structure, not learned values.

4.3.1 Candidate independence and consistency

At a fixed state, inserting a duplicate candidate should not change the selected model’s raw prediction for an existing view. This independence is built into both A0 and A1. Reordering candidates should likewise reorder their predictions without changing the value attached to each pose. For a joint row permutation Π ,

$$f_{\theta}(\Pi \mathbf{X}_t, \Pi \mathbf{m}_t^{\text{cand}}) = \Pi f_{\theta}(\mathbf{X}_t, \mathbf{m}_t^{\text{cand}}) \quad (46)$$

Padding has no physical meaning and must not affect predictions for real candidates:

$$\begin{aligned} \text{Mask}(\mathbf{X}_t, \mathbf{m}_t^{\text{cand}}) &= \text{Mask}(\mathbf{X}'_t, \mathbf{m}_t^{\text{cand}}), \\ &\Rightarrow \\ \text{Mask}(f_{\theta}(\mathbf{X}_t, \mathbf{m}_t^{\text{cand}}), \mathbf{m}_t^{\text{cand}}) &= \text{Mask}(f_{\theta}(\mathbf{X}'_t, \mathbf{m}_t^{\text{cand}}), \mathbf{m}_t^{\text{cand}}) \end{aligned} \quad (47)$$

Here $m_{t,i}^{\text{cand}}$ is one for a materialized candidate and zero for padding. It is distinct from $m_{t,i}^{\text{act}}$, which determines whether a proposed action may be selected. The scorer uses only the padding mask; action feasibility is applied to selection and training outside the scorer. Changing the hard action mask therefore leaves raw conditional values and feasibility logits unchanged in this architecture.

The row-permutation example in [Figure 14](#) summarizes these checks: the identity of a pose follows its row through reordering, while padding and external action support remain different operations.

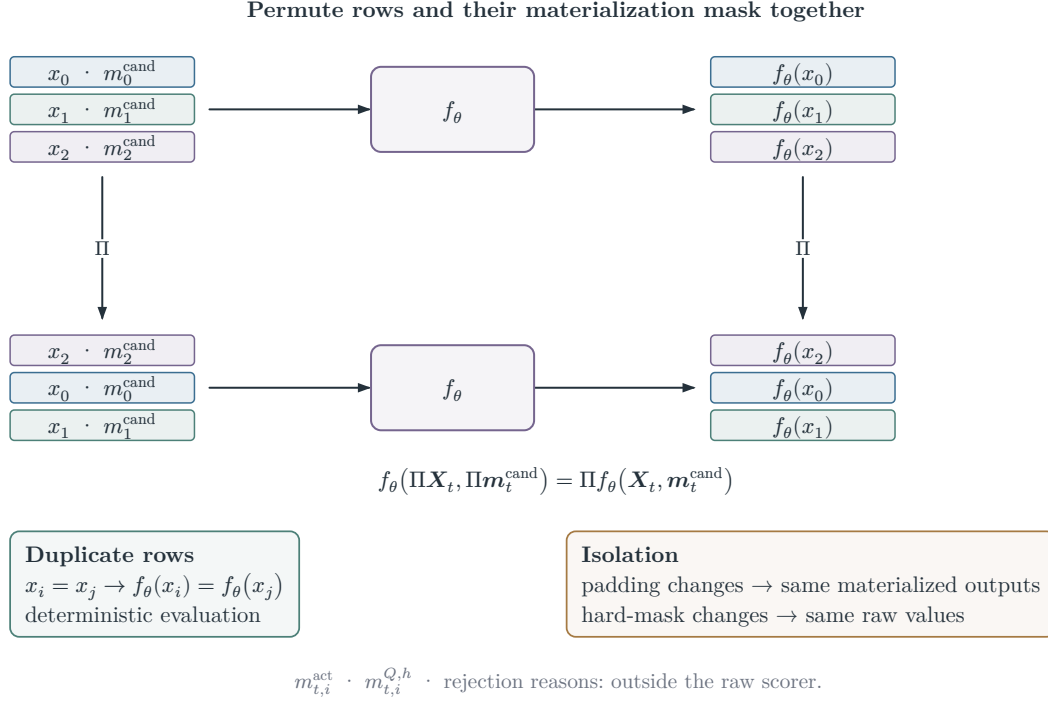


Figure 14: Row consistency of the independent scorer. Permuting candidates and their materialization mask permutes the outputs; identical inputs produce identical values in deterministic evaluation. Padding and authoritative action support remain distinct from the raw value computation.

Independent scoring is a modelling choice, not a requirement of every value function. A model with candidate-set summaries or self-attention may condition on which other actions are offered, provided that the offered set is part of its state and learning target. The distinction concerns the information in the value definition, not whether the network uses attention.

4.3.2 Geometry, information, and horizon

Local-frame encoding expresses candidate poses relative to the root and current frames, and the target pose relative to each candidate. These relations avoid using an arbitrary world origin as a feature without establishing exact SE(3) equivariance [25]. The shared PoseTW encoder retains both translation and rotation.

Tests separately check the actor’s information boundary: target gains, meshes, associations, and current candidate renders are excluded from its inputs. Previously selected privileged depth is available only in an explicitly non-deployable state protocol.

At $h = 1$, the target has no bootstrap term. A request with $h > b_t$ is rejected rather than clamped. When a longer-horizon target has a supported bootstrap, it queries the recorded successor at $h - 1$. Because an admissible action can have negative immediate gain, values are not constrained to increase with horizon.

4.4 Target-Conditioned Finite-Horizon Value

Implementation: partial **Evidence:** pending

The selected executable baseline is A1-H0- $s_t^{\text{S0-pose}}$ with a scalar requested horizon, direct continuous Huber regression, fitted-Q training, and runtime selection under the configured hard action-support mask. Current experiments use oracle-derived support. The implementation path is tested, but learned exact-Q2 accuracy and policy evidence remain absent.

Literature: [15], [17], [18], [23], [37], [38]

Promotion gate: populate a frozen held-out exact-Q2 receipt; evaluate matched endpoint reconstruction errors on untouched scenes

Development source: aria_nbv/aria_nbv/vin/models/target_finite_horizon.py; aria_nbv/aria_nbv/lightning/finite_horizon/module.py; aria_nbv/aria_nbv/lightning/finite_horizon/objective.py; aria_nbv/aria_nbv/lightning/finite_horizon/certification.py; aria_nbv/aria_nbv/oracle/pipelines/online_qh.py; aria_nbv/tests/lightning/test_qh_q2_certification.py

The evolving-volume architecture proposed in this chapter retains this output and learning interface. It changes how the causal actor state is formed and how one candidate reads it; it does not change the meaning of the scalar return, move hard feasibility into the value, or grant the scorer access to labels. The compact A1-H0-S0 path remains the executable baseline, and the current experimental volume is the spatial implementation control. Proposed geometric pooling, spatial attention, and frozen-feature inputs require matched evidence before any is selected.

4.4.1 Conditional scorer

The encoder defined above maps root evidence, target and candidate geometry, selected-pose history, remaining budget, and requested horizon to one query per materialized candidate. Oracle gains, target crops, meshes, associations, and audit lineage remain excluded. Because its dynamic component contains H0 pose history

rather than selected observations, this executable **S0-pose**, denoted $s_t^{\text{S0-pose}}$, is not assumed to be a task-sufficient reconstruction state.

Before the hard action mask is applied, the scorer emits

$$\begin{aligned} (\hat{Q}_{\theta,h}^{\sigma,\Xi}(z_t^\sigma, e, q_{t,i}), \ell_{t,i}^{\text{feas}}) &= f_\theta(z_t^\sigma, e, q_{t,i}, h), \\ 1 \leq h \leq b_t &\leq H_{\max}, \\ h = b_t &\text{ when omitted} \end{aligned} \tag{48}$$

for every materialized row. The conditional value alone carries return units and enters regression, Bellman backup, and online ranking. Feasibility logits are a separate auxiliary prediction; multiplying them by value would mix meanings. The training adapter owns hard-valid label and bootstrap admission, while online inference owns the final masked selection.

The selected A1 interaction lets each $q_{t,i}$ query shared scene, target, history, budget, and horizon tokens. A0 supplies the matched independent-row control. Both use the same geometric query and value decoder. Their comparison therefore asks whether query-dependent state reading helps under a fixed state and objective; it does not isolate attention independently of parameter count or runtime.

The policy then chooses among admissible candidates:

$$a_t^\theta = \operatorname{argmax}_{i:m_{t,i}^{\text{act}}=1} \hat{Q}_{\theta,h}^{\sigma,\Xi}(z_t^\sigma, e, q_{t,i}) \tag{49}$$

The hard action mask restricts selection after prediction. This completes the forward path; the following controls change representations or interaction, while the later learning section defines the return being fitted.

4.5 Representation Comparisons

A bounded volumetric alternative restores spatial correspondence before the candidate-value query (Figure 15). Frozen root fields and semidense points initialize a fixed, root-chart-aligned grid with a one-metre margin beyond the original EVL extent. For each decision state, the selected-depth variant reconstructs exactly its admitted observation prefix. Projecting voxel centres into each calibrated depth

image supplies separate surface and free-space votes; centres beyond the surface-tolerance band behind a return gain neither vote. The band is a centre-sampling heuristic, not a conservative bound on the projected extent of a rotated voxel. Observation identities are integrated once; recency records the normalized last-observed index. Learned features are recomputed for each forward pass.

The common grid has tensor layout $(C, D_{\text{vol}}, H_{\text{vol}}, W_{\text{vol}})$ per state; its width, height and depth axes correspond to metric x, y, z , respectively. Each channel of $\mathbf{M}_t$ assigns one scalar to every cell: frozen root fields and voxelized semidense points, selected-depth surface and free-space counts, recency, source availability, and original-chart extent. In the historical projection-count profile, the Boolean [composite support mask](#) A_t combines positive root projection counts, root occupancy input, semidense occupancy, and selected-depth surface or free-space votes. These sources have different meanings: projection coverage alone does not establish measured geometry. Availability and chart extent separate a missing source or a cell outside the original chart from a numerical zero within an available field.

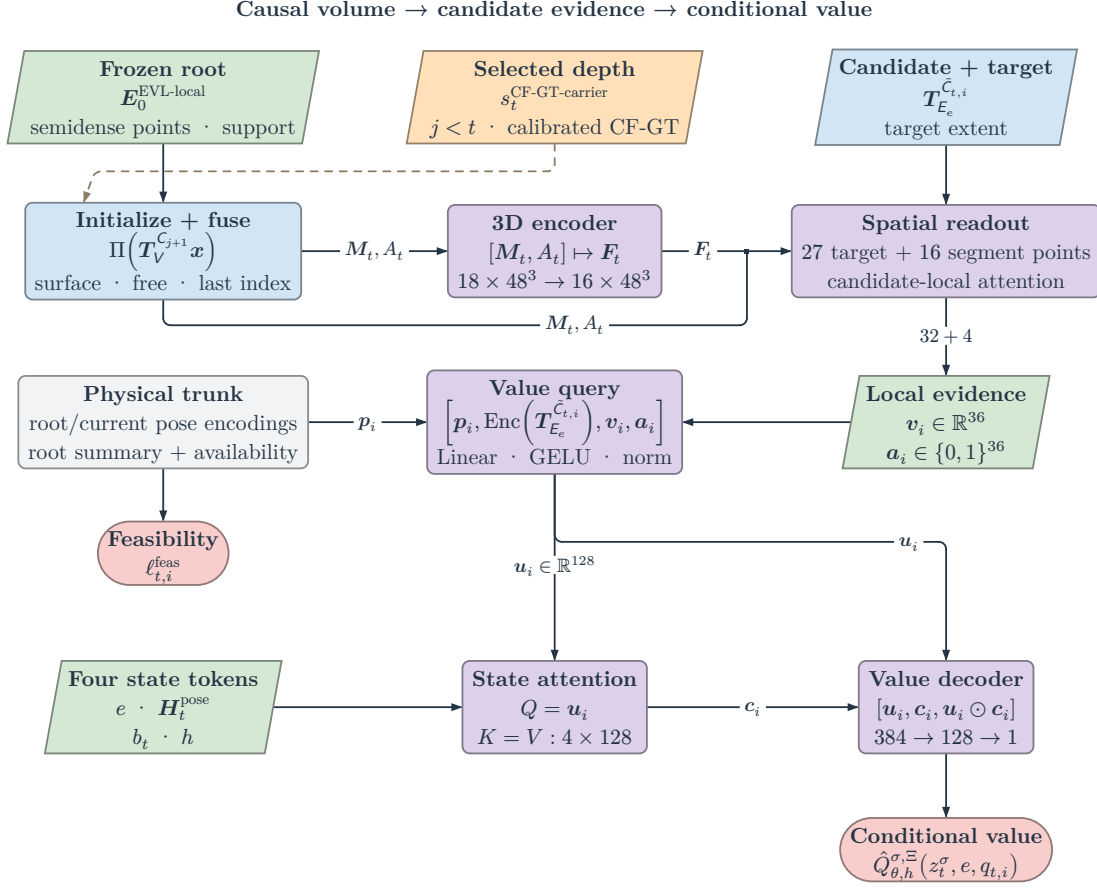


Figure 15: Implemented volumetric alternative and its QH interface. Green denotes evidence, blue deterministic geometry, purple learned operations, and pink predictions. The dashed orange input is the privileged selected-depth path; the root-only variant omits it. $\mathbf{M}_t$ contains seventeen evidence/source channels. The current support mask A_t in the historical profile conflates positive root feature-projection counts with measured surface/free support. The learned volume is $\mathbf{F}_t$. Each read returns 32 attended features and four support statistics, paired with availability bits $\mathbf{a}_i$. Four shared tokens condition value; the feasibility branch remains target-independent. Dimensions describe the configured diagnostic model, not measured activations or comparative performance.

A two-scale 3D encoder produces learned spatial features while preserving the composite support mask. Candidate reads sample both these features and the deterministic channels at a $3 \times 3 \times 3$ target lattice and sixteen points along the segment from the candidate optical centre to the target centre. These points probe spatial context between the proposed view and target; they are not calibrated camera rays and do not establish visibility. Each spatial token adds candidate-frame coordinates and a target-versus-segment indicator. A relative-target-pose

query attends to these tokens independently for each candidate. In-map unknown cells remain readable; out-of-map samples are excluded. For the target lattice and segment separately, the readout reports the fraction of samples with composite support and whether any sample has such support. These statistics describe the supplied composite mask, without identifying which source supplied it. In the historical profile, positive root feature-projection counts enter that mask, so those statistics do not establish measured support. Selected-depth votes retain their own measured-source provenance. The corrected support profile instead uses occupied root input and admitted semidense points as measured root evidence, excluding image-feature projection counts; selected-depth surface/free votes are unchanged. A root free-input field without native-ray provenance is not promoted to measured support. The larger Q1 fit in Results uses this corrected profile, while its centre-segment attention remains distinct from the proposed calibrated multi-ray readout. [Figure 39](#) expands this encoder/readout interface.

The downstream spatial scorer combines this evidence with pose encodings and the physical token, then conditions each value query on four state tokens: target, causal pose history, remaining budget, and requested horizon. Spatial evidence enters through the candidate read instead of a fifth pooled scene token. The existing value decoder, fitted-Q objective, and Lightning experiment path remain in use. The selected-depth variant consumes privileged mesh depth and is therefore a diagnostic control. Root-only inference has a supported bundle path; that software boundary does not establish actor-visible target admission or calibrated action support. Neither variant establishes a full-frustum readout, calibrated uncertainty, or a held-out value advantage.

The state carrier should become more elaborate only when a matched comparison identifies a distinction that the current carrier loses. The intervention must restore that distinction with the smallest change that leaves the task, candidate support, replay population, and evaluation protocol fixed.

The full conditional intervention ladder is retained in [Section 9.6](#). The present comparison keeps the admitted observations and value target fixed; new carriers are motivated by a measured loss of spatial support or information.

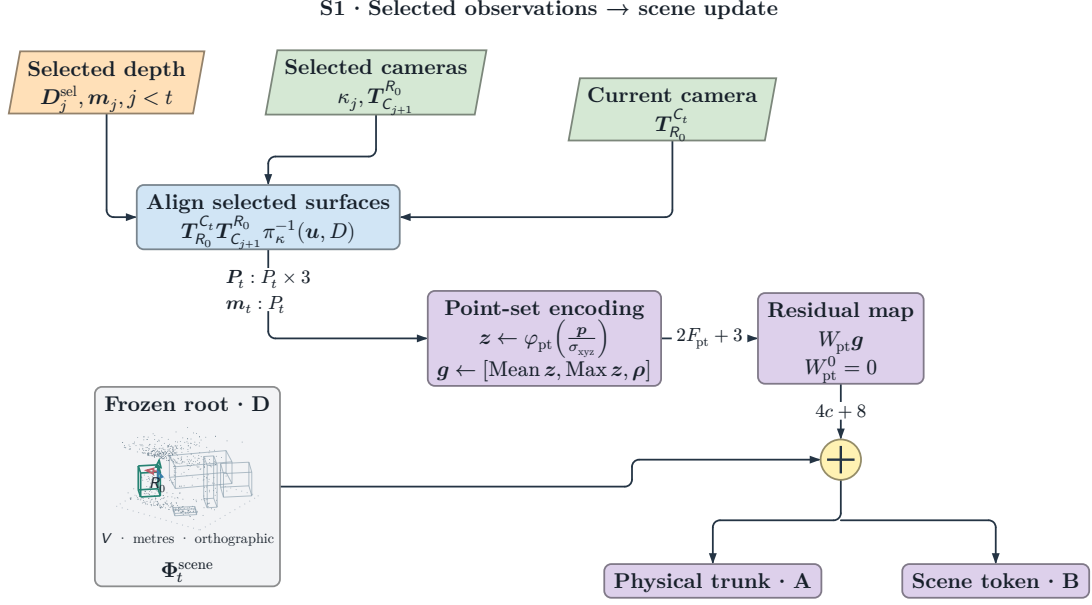
Open decision: Choose commissioning order explicitly. CF+ H0 versus S1 tests selected-surface consumption; geometric pooling versus spatial attention tests readout weighting; frozen features test representation reuse.

Source: Method commissioning proposals; Experimental Design task and policy comparisons

Gate: Choose one bounded commissioning comparison with frozen purpose, controls, population, and information contract.

The source-matched CF+ H0 versus CF+ S1 diagnostic uses identity-start initialization. CF+ H0 validates the selected-depth structure while remaining invariant to its numeric payload; S1 consumes the selected surface summary. Their difference therefore asks whether any selected-surface signal reaches the scorer through the current compression seam. Because S1 also changes geometry consumption, point encoding, initialization, and model capacity, a positive result would motivate an actor-visible realization but would not validate one. A null result would remain ambiguous between absent signal and a lossy surface summary.

[Figure 16](#) shows where the privileged control introduces selected observations. Each selected depth map is unprojected with its calibration, aligned to the current frame, and summarized by a masked point encoder. A zero-initialized residual map adds that summary to the frozen root carrier, which feeds both the physical trunk and the shared scene token. Before point encoding, current-camera coordinates are divided by a metric scale to make them dimensionless; this does not crop or clip the geometry. The encoder pools a masked mean and maximum, accompanied by three summaries: whether any valid point exists, the fraction of sampled pixels yielding valid points, and the fraction of selected views contributing at least one point. Thus many points from one view and sparse evidence from many views can carry different support summaries. Empty evidence yields a zero residual.



Root inset: measured ASE 83269/000056. Selected-depth prefix shown symbolically.
CF-GT depth is privileged. m_t masks points; ρ summarizes support. Defaults: $F_{pt} = 64$, $c = 5$.

Figure 16: Selected-surface residual in the privileged S1 control. Only the selected prefix $j < t$ is aligned and pooled; point and support masks distinguish missing evidence from measured zeros. The residual map starts at zero, so the initial carrier equals the frozen root. The root inset is measured; the selected-depth prefix is symbolic. Ground-truth mesh depth makes this an explicitly non-deployable control.

A spatial comparison can reuse persisted local evidence before acquiring or re-lifting evidence. The corrected experimental volume is the implementation starting point: positive EVL image-feature projection counts are excluded from measured support. With the declared evidence semantics, the deterministic volume is encoded once per state and read independently for each candidate. Geometric pooling and candidate-local spatial attention test whether learned weighting improves those reads. Their matched observations, samples, encoder, mass features and decoder, including the context path without geometric bias, are specified in [Section 4.7](#).

The first optional pretrained-feature arm samples a frozen EVL neck field at the same calibrated locations and fuses it after the geometric read, using a trained projection into the existing downstream decoder. It does not replace surface/free/unknown evidence, and its benefit cannot be attributed to the readout unless the readout is fixed first. Feature extraction remains frozen, with chart, source,

availability, and pretraining provenance declared. This arm can be tested during Q1 commissioning; Q2 is not its prerequisite. Only a measured chart-extent or resolution failure motivates sparse or re-centred storage.

Renderable or symbolic scene models remain exploratory responses to failures that have not yet been diagnosed. Because they change optimization cost, supervision, or the meaning of state, they remain in the development alternatives register until a matched result shows that causal local evidence and candidate-relative reads are insufficient.

4.6 Pose-history order and observed content

The selected H0 encoder discards pose order. The implemented, unselected H1 control attaches relative age and applies causal attention to the same pose-only prefix. Neither supplies observed surfaces or free-space evidence. Its full construction is retained in the Method Alternatives register ([Section 9.3](#)); the core observation-updated comparison asks a different question about the information supplied to the scorer.

4.7 Spatial Readout and Interaction Comparisons

The evolving-volume study preserves the same candidate independence and external feasibility boundary. Deterministic geometric pooling provides the control over target samples and candidate-to-target ray samples. Integration weights define the normalized summaries, while separate mass features record how much target, ray, observed, unknown, and in-map support contributed. This prevents a normalized mean over one supported sample from appearing equivalent to the same mean over a broadly supported region.

The matched treatment replaces only that aggregation with one candidate-local cross-attention block. It receives the same encoded volume, deterministic channels, physical samples, integration weights, mass features, and downstream decoder. Calibrated image inclusion, observed occlusion evidence, and uncertainty remain different quantities. A bounded soft geometric or visibility term may bias attention, but does not set an attention weight or remove the context path without

that geometric bias. Consequently, an unknown or currently occluded region can remain relevant to a setup action. Whether this learned spatial weighting improves immediate candidate ranking is a hypothesis for the matched Q1 experiment, not a property inferred from the mechanism.

The A0/A1 comparison above asks how a candidate combines already formed state summaries. This proposed comparison asks how spatial samples become one candidate feature. The two attention sites answer different questions and are not exchanged under one label.

4.7.1 Planned interaction comparisons

Candidate-relative relation embeddings remain a planned test when held-out errors retain target or history structure after controlling for state, action support, labels, optimization, seeds, and capacity. Candidate-set interaction tests whether the offered alternatives provide useful additional context. Recurrent reading tests repeated access to the same state, rather than adding new observations. The F0–F5 development register retains these alternatives and their one-factor comparisons; their value must be established empirically.

4.8 Finite Action and Causal Replay

Implementation: partial **Evidence:** pending

Finite candidate tables, mesh-derived privileged oracle action support, and the $s_t^{\text{SO-pose}}$ replay transition are implemented and tested. This is a privileged oracle-support control. The scientific target additionally requires actor-visible action support and a causal observation update, which remain pending.

Promotion gate: retain deterministic row identity, source roles, generator and action-support identities, hard-mask isolation, and selected-transition validation; add actor-visible action support, selected-observation fusion, and no-future-observation tests

Development source: `aria_nbv/aria_nbv/targets/protocol.py`; `aria_nbv/aria_nbv/rollouts/replay/engine.py`; `aria_nbv/aria_nbv/rollouts/storage/reader.py`; `aria_nbv/aria_nbv/rollouts/storage/writer.py`; `aria_nbv/aria_nbv/rollouts/storage/validation.py`; `aria_nbv/aria_nbv/rollouts/storage/qh_reader.py`; `aria_nbv/tests/rollouts/test_qh_reader.py`

At step t , the named generator protocol returns the full candidate table $\mathcal{Q}_t$, the hard **validity mask** $m_{t,i}^{\text{act}}$, and versioned failure reasons. The admissible action set is

$$\begin{aligned}
q_{t,i} &:= \mathbf{T}_{\tilde{\mathcal{C}}_{t,i}}^{R_0} \in \text{SE}(3), \quad \mathcal{Q}_t = \{q_{t,i}\}_{i=1}^{N_q}, \\
\mathcal{A}_t &= \{i \in \{1, \dots, N_q\} : m_{t,i}^{\text{act}} = 1\}, \quad a_t \in \mathcal{A}_t
\end{aligned} \tag{50}$$

Invalid rows remain auditable but cannot enter selection, Q supervision, or a bootstrap maximum. A feasible row may have negative utility; infeasibility is therefore never encoded as a small value. Padding, materialization, action validity, Q-training eligibility, and any predicted feasibility remain distinct. Under the dense-valid supervision profile, the writer proves and the reader revalidates equality of Q-label support and hard action support on every realized state. Unknown or legacy label-density profiles fail closed for this claim.

Candidate orientations factor a component-specific base gaze from bounded yaw–pitch perturbations. Paired proposal components may reuse a camera centre across two base-gaze families, but both remain separate actions. This makes translation and viewing direction distinguishable without collapsing their values. Generator family, resolved configuration, randomness derivation, sampled support, hard rejection, mask semantics, and stable shell identity remain attached to every row so proposal coverage can be audited before policy quality is interpreted.

Each $q_{t,i}$ is an endpoint pose in $\text{SE}(3)$, expressed by the prospective camera frame $\tilde{\mathcal{C}}_{t,i}$. The common row index i aligns that pose with provenance and admission metadata, but those fields are not part of the pose. The prospective frame does not denote an acquired observation. Selecting its row index realizes the endpoint within the current rollout:

$$\mathbf{T}_{\mathcal{C}_{t+1}}^{R_0} := q_{t,a_t}, \quad a_t \in \mathcal{A}_t \tag{51}$$

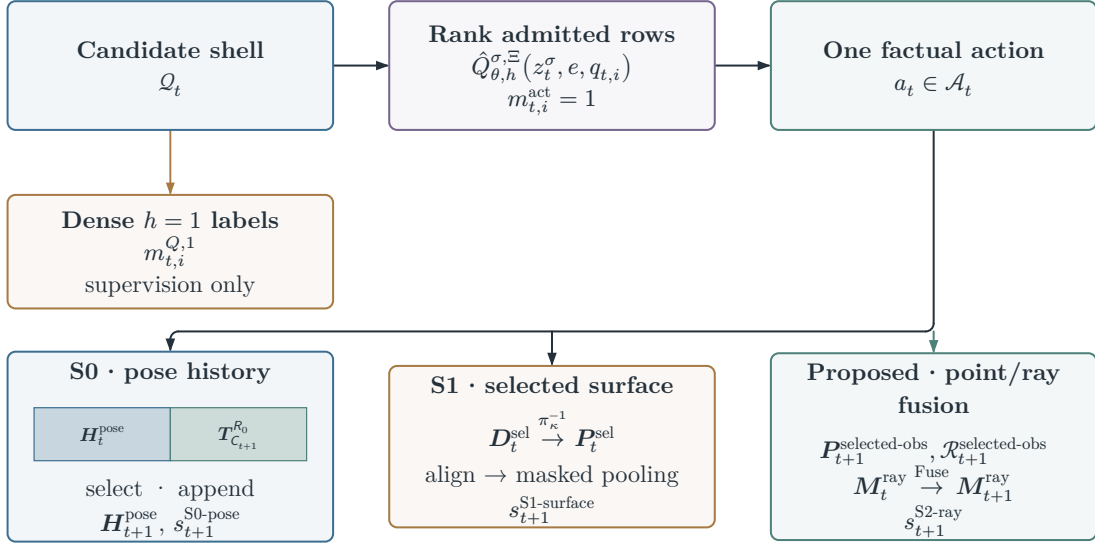
The realized path can be synthetic and need not coincide with the recorded Aria trajectory. New actor-visible evidence becomes available only after the selected acquisition under the declared observation protocol.

4.8.1 Immediate labels and one factual successor

Chapter 3 defines the normalized replay relations and their provenance; [Figure 42](#) shows how a state owns a full candidate shell while one selected row links to the

next state. The Method uses that relation through one causal rule. An eligible candidate row may receive an immediate target-utility label from privileged offline evaluation, but a label is supervision for a prediction at the current state, not an observation available to the actor. Chapter 3 owns the general support and lineage semantics; the receipt-bound Phase-A floors and coordinate convention are recorded in [Section A.1](#); and [Chapter 6](#) owns the authenticated counts and their interpretation.

Dense immediate supervision and factual replay therefore branch at the same candidate shell. Oracle evaluation can label every eligible row for the one-step target, so those rows can train or compare immediate-value predictions. The rollout policy nevertheless chooses exactly one admissible action, and only that row supplies the factual successor used by a longer-horizon target. When the protocol materializes an observation for the selected row, that observation may update the causal prefix. The current S0 replay updates pose, history, budget, and support without such an observation update. Unselected candidate renders neither become actor observations nor define alternative successor states.



Each protocol: $b_t \rightarrow b_t - 1$ · regenerate Q_{t+1} · query horizon $h - 1$

Solid: implemented controls. Dashed: proposed observation path. Unselected views never update history.

Figure 17: Counterfactual supervision and factual replay serve different roles. Immediate oracle labels may supervise every eligible candidate at the current state. The authoritative action mask is applied outside the scorer, and only the selected row supplies the causal successor for the next query at horizon $h - 1$. If its protocol supplies an observation, only that selected observation may update state. Current S0 replay updates pose and support without an observation update; persisted selected CF-GT depth is backprojected into the privileged surface control, while actor-visible fusion is a separate proposal. Unselected renders never update actor state.

4.8.2 Current replay transition

After selecting an admitted row as shown in Figure 17, the current replay engine applies

$$s_{t+1}^{\text{S0-pose}} \leftarrow \mathcal{T}_\Xi(s_t^{\text{S0-pose}}, a_t), \quad a_t \in \mathcal{A}_t \quad (52)$$

Here a_t is the selected row index, whereas i continues to index arbitrary rows in the shell. Thus only q_{t,a_t} realizes a factual successor. The relation states the transition between implemented $s_t^{\text{S0-pose}}$ states. Its component assignments are defined once in Equation 74: they realize the selected pose, append it to H_t^{pose} , decrement b_t , then regenerate the next candidate table and hard action mask with the frozen seeded generator g_Ξ . Proposal and action-selection randomness use separate streams keyed by the factual selected-action history, so reordering retained trajectories cannot change a previously defined state.

Together, the compact relation and its component definition in Equation 74 specify the implemented $s_t^{\text{S0-pose}}$ transition under privileged oracle action support. It changes the reference pose, selected-pose history, remaining budget, and finite action support. It does not assert that the actor received RGB, fused depth, or an updated spatial reconstruction, and the mesh-derived action mask is not presented as actor-visible. A selected mesh-depth raster may support the distinct privileged $s_t^{\text{S1-surface}}$ control; it is not evidence that an actor-visible observation update has been implemented.

4.8.3 Scientific target transition

The [causal observation state](#) retains the same factual action, budget, and finite-candidate interface, but replaces privileged action support with an actor-observed mask or a separately calibrated learned-feasibility route. After the policy chooses a_t , the dynamic state must update from the observation acquired at that pose. The update may use the previous actor-visible state, the selected action, and the selected observation; it may use neither future observations nor renders of unselected candidates. The resulting carrier must preserve observed surface, observed free space, unknown space, finite support, uncertainty, source, and recency, so two pose-identical histories with different observations need not collapse to the same state.

This actor-visible update is a scientific requirement rather than behavior of the current replay engine. Promotion requires deterministic fusion, source-dropout and no-future-observation tests, target-source leakage checks, and held-out evidence that the added state improves target-specific value prediction. Persisted selected mesh depth remains a privileged control until an actor-visible observation path satisfies those gates.

The normalized replay view expands one retained chain into its realized decision states. Each state sees only the prefix available before its action. Several states from one chain may share a batch, but that tensor colocation does not create temporal communication. The factual selected action supplies the only successor

link used by finite-horizon learning; no counterfactual transition is fabricated for unselected rows.

Training, validation, and test populations used for one claim must share the target, actor-state, generator-family and resolved-configuration, randomness-derivation, action-mask, reward, horizon, and label-support protocols. Different random seeds may realize new candidate tables under that frozen stochastic generator; changing the generator protocol or admissibility semantics instead changes the supported decision problem. Such a comparison is reported as an explicit support-transfer experiment, not absorbed into an ordinary split. Behavior-policy changes likewise remain identified because they alter replay coverage even when the greedy target continuation is unchanged. Physical-scene splits and evaluation-population selection must be independent of oracle outcomes and model error. Within a training split, acquisition may depend on oracle scores through the declared collector; that dependence remains part of the replay provenance.

4.9 Value Definition and Learning

4.9.1 Ideal value and learned representation

The value estimand is meaningful only relative to one frozen decision protocol,

$$\Xi = (g, \tau, \sigma, \nu_{\text{mask}}, \rho, \gamma, H_{\text{max}}) \quad (53)$$

where g fixes the candidate generator and its resolved configuration; τ fixes target-source admission; σ fixes state construction; ν_{mask} fixes the provenance and meaning of action support; ρ fixes reward, execution, and terminal conventions; and γ and H_{max} fix discounting and the largest residual horizon. Changing the admitted observations, reward, action support, or transition and observation laws can change the decision problem. Encoding the same admitted history differently instead changes the approximation used to predict its value; a lossy encoding may also restrict the policies that can be represented. Matched representation comparisons therefore retain the same history-conditioned value target. Changes to the decision problem are identified separately as protocol-transfer studies. Future candidate-set randomness is part of this protocol, as in planning with stochastic

action sets [36]. The collection policy instead determines the replay distribution used to estimate the value. Under a fixed, sufficiently observed decision process, a different behavior policy need not change the optimal value; it changes which consequences are available to learn from. Finite data, restricted function approximation, and aliased histories can make the fitted predictor strongly dependent on that collection distribution.

Let $\mathcal{H}_t$ denote the actor-visible observation and action history available before decision step t . The finite return defined in Foundations sums the task rewards from Chapter 3 over the requested h acquisitions. Its ideal actor-admissible control value is

$$\begin{aligned} Q_{h,e}^{\star,\Xi}(\mathcal{H}_t, q_{t,i}) &= \sup_{\pi \in \Pi^{\text{act}}} \mathbb{E}_{\pi, a_t := i} \left[G_{t,e}^{(h)} \mid \mathcal{H}_t \right], \\ i &\in \mathcal{A}_t, \quad 1 \leq h \leq b_t \leq H_{\max}, \\ Q_{0,e}^{\star,\Xi}(\mathcal{H}_t, q_{t,i}) &= 0 \end{aligned} \tag{54}$$

Here $q_{t,i}$ is the candidate action identified by its physical endpoint; i is only its table-local handle in the protocol-defined admissible set $\mathcal{A}_t$. The first action is prescribed: the question is what would follow if the actor executes that pose, rather than what happens among records in which a collector happened to choose it. The policy π ranges over continuation policies using only actor-visible information. The expectation includes the transition, observation, and future candidate-set randomness fixed by Ξ . This is a history-conditioned object: it does not assume that the selected finite-dimensional model input is Markov-sufficient.

The executable scorer instead receives the representation

$$z_t^\sigma = \varphi_\sigma(\mathcal{H}_t) \tag{55}$$

where φ_σ denotes the state-construction protocol and z_t^σ is its output. Under the sufficient condition stated below, the pointwise training target can be written as

$$\hat{Q}_{\theta,h}^{\sigma,\Xi}(z_t^\sigma, e, q_{t,i}) \approx Q_{h,e}^{\star,\Xi}(\mathcal{H}_t, q_{t,i}), \quad \text{if } z_t^\sigma \text{ is decision-context sufficient} \tag{56}$$

from the frozen replay distribution. The hat distinguishes this learned predictor from the ideal value; the superscripts retain the representation and decision-

protocol identities that determine what was fitted. The following condition is sufficient for exact Bellman recursion on the compressed representation:

$$Q_{h,e}^{\star,\Xi}(\mathcal{H}_t, q_{t,i}) = Q_{h,e}^{\star,\sigma,\Xi}(z_t^\sigma, q_{t,i}) \quad \text{if } z_t^\sigma \text{ is decision-context sufficient} \quad (57)$$

Decision-context sufficiency means that histories mapped to the same z_t^σ induce the same joint conditional law of reward, successor representation, residual budget, regenerated candidate table, hard action-support mask, and termination for every physical candidate pose. This condition guarantees the displayed factorization and Bellman recursion. For expected returns, matching the mean immediate reward and the distribution of subsequent decision opportunities already suffices; matching the full reward distribution is stronger than necessary. A representation can also preserve a particular value function without preserving that complete model [24, Sec. 2, Theorem 1; App. A.2.2]. Finite held-out tests can assess prediction adequacy on the evaluated histories and action support; they cannot prove this equality for every relevant history. Equal successor representations alone are insufficient because the available continuation support can still differ. Without demonstrated sufficiency, fitted-Q training yields a representation- and replay-conditioned projection that may still be useful for ranking candidates, but it is not silently reinterpreted as the Bellman-optimal value of a Markov state. This is the admitted status of $s_t^{\text{S0-pose}}$. Causal observation-updated state ablations test the sufficiency hypothesis rather than merely enlarging a token.

A further limitation concerns how the recorded action was chosen. If a collector uses hidden geometry that affects both selection and future gain, then records of its selected actions can differ from the outcomes an actor would encounter when prescribing those actions from the same visible history. This is a possible hidden-selection problem, not evidence that oracle use alone biases this dataset. Kallus and Zhou study how unobserved selection information can obstruct off-policy identification [39]; their stationary infinite-horizon estimator is not applied here. Whether the relevant information is already captured in the actor history remains an empirical question. Generators and validity masks must be included in that information audit, since they can also depend on privileged geometry.

Research TODO: RRC08 — Audit whether outcome-relevant information outside the actor history influences collection, candidate generation, or action support. Retain available selection probabilities and their conditioning information.

Source: Protocol-conditioned value; collector and actor information boundaries; Kallus and Zhou 2020, Sections 2–3

Gate: Information and probability audit; diagnose selection bias only with evidence beyond privileged access itself.

4.9.2 Scalar requested horizon

One model call follows the frozen interface

$$f_{\theta}\left(s_t^{\text{S0-pose}}, \mathbf{T}_{E_e}^{R_0}, \ell_e, \{q_{t,i}\}_{i=1}^{N_q}, h\right) \rightarrow \left(\left\{\hat{Q}_{\theta,h}^{\sigma,\Xi}(z_t^{\sigma}, e, q_{t,i})\right\}_{i=1}^{N_q}, \left\{\ell_{t,i}^{\text{feas}}\right\}_{i=1}^{N_q}\right) \quad (58)$$

The notation Q_H names the bounded predictor family rather than a collection of separately configured models. Remaining budget b_t describes the factual state; h selects one return from the triangular domain $\{(b_t, h) : 1 \leq h \leq b_t \leq H_{\max}\}$. Omitting h requests the full factual residual budget. The mathematical boundary $Q_0 = 0$ closes recursion, whereas zero in executable tensors denotes padding and is never a learned query.

Dense labels train $h = 1$ for every hard-valid candidate. For $h > 1$, only the factual selected action has a stored successor state and replay link, so multi-step transitions are sparse even when immediate labels are dense. The current recursive construction and exact- Q_2 surface populate the factual diagonal $h = b_t$, with exact Q_2 executable at $(b_t, h) = (2, 2)$ but its held-out receipt still pending. Dense Q_1 additionally supports $(b_t, 1)$ across realized budgets. Current bundles record trained horizons rather than realized pairs, while deployed online inference requests the diagonal and applies that horizon gate. Before any off-diagonal $h > 1$ query is exposed, the bundle and runtime must add pair-bound training and evaluation evidence and reject unsupported pairs. A wide syntactic interface is not evidence of wide learned capability. The default profile bounds both remaining budget and requested horizon by five; this is a configuration limit, not a theoretical limit on finite prediction. In particular, requesting two steps at a remaining budget of sixteen requires a compatible larger-budget profile and separate pair-specific support. It is not enabled by choosing a smaller horizon alone.

Open decision: RRC07 — Extend the D05 input-profile and M30 coverage decisions to specify the actor-visible information, action-support privileges, and admitted budget-horizon pairs for each comparison.

Source: D05 exact input profiles; M30 query support; `qh_actor_preparation.py` budget and horizon checks

Gate: Frozen per-arm input and support profile, with evidence for each exposed pair.

A recording boundary also affects these queries. In the legacy countdown profile, record absence can produce a terminal flag and an empty successor mask, leaving a selected row with an immediate-reward-only target. A missing record leaves future rewards unknown; it does not establish that the requested query or real task ended. The implemented `continuing_selected_rows_v1` profile instead requires a recorded successor with finite, hard-valid reward-label support for recursive loss. Saved terminal flags and collection-cutoff discounts do not define its task boundary. This admission rule avoids treating absent continuation as zero, but does not recover missing observations or settle the intended stopping rule. Boundary handling must be checked across reader, batch and loss, rather than inferred from a terminal flag alone.

Implementation TODO: RRC05 — Separate true query/task boundaries, missing successor records, and stored successors without actor-valid actions in reader, batch, and loss admission. Add pair-specific support before exposing earlier-state queries with shorter multi-step horizons.

Source: Replay successor construction and fitted-Q admission; M30 query-domain limitation

Gate: Boundary-aware admission tests and pair-bound training/evaluation evidence; a terminal-flag edit alone is insufficient.

4.9.3 Direct continuous objective

The decoder maps each candidate representation directly to the continuous, root-normalized finite-horizon return. Larger target values represent greater cumulative reconstruction gain over the requested horizon. The implemented Huber penalty is quadratic for small prediction errors and linear beyond its chosen threshold. This limits the influence of large errors during training. The intended prediction is the average return for the given model inputs. Huber fitting does not always recover that average, so this agreement must be checked on held-out data.

Validation TODO: RRC09 — Extend D06 and M31 evaluation to distinguish fitted backup-target agreement, learned Q2 prediction accuracy, agreement with a named actor continuation, and actual endpoint reconstruction improvement.

Source: D06 independent targets; M31 mean-value agreement; exact-Q2 certification and matched policy evaluation

Gate: Independent supported references and named-policy outcomes under matching horizon clock, measurement, and actor information.

Losses are first averaged within each realized state and then within each horizon. Consequently, a state with more valid candidates does not receive more weight merely because it contributes more rows, and abundant one-step states do not silently dominate a sparser recursive horizon. Reported calibration and support remain stratified by horizon.

At $h = 1$, evaluation can use dense candidate labels to report continuous error, within-state pairwise ranking, and greedy regret. Factual $h > 1$ transitions do not supply counterfactual values for unselected rows, so longer-horizon ranking and regret remain undefined until exact or independently oracle-rescored candidate tables exist. Recording zero support is more faithful than inventing a metric from the selected transition alone.

4.9.4 Selected decoder and matched control

Direct regression is selected because the Bellman target, endpoint gain, and reported calibration share continuous return units. The implemented CORAL path is retained as the one matched decoder-objective control under the same input and underlying continuous fitted-Q target. It represents that target with ordinal threshold labels, optimizes the CORAL loss, and decodes the ordinal output through frozen representative values. The comparison therefore changes the output parameterization, loss, discretization/support artifact, and representative decoding as one controlled package; it does not define another action value.

The complete decoder comparison and its frozen-support requirements remain in [Section 9.7](#). This optional package effect does not change the underlying continuous value target.

Immediate-value commissioning uses $h = 1$ at every realized causal state. This preserves each state’s actual observation prefix and remaining budget while

removing learned continuation from the comparison. Dense immediate labels then support calibration, within-state ranking, and greedy regret for compact, current centre-segment-attention volume, proposed calibrated multi-ray geometric-pooling, and proposed soft-biased multi-ray attention arms under a common objective. The Q1 study cannot establish non-myopic value. Supported recursive Q2 is a separate continuation experiment requiring factual successor evidence. A positive historical usefulness threshold, prefix-reuse optimization, or reuse across horizon queries is not a prerequisite for first Q1 fitting. The optional frozen-feature trial can precede Q2; neither trial establishes actor-visible endpoint recovery.

4.9.5 Fitted-Q recursion

Batch fitted Q iteration turns the fixed replay collection into successive supervised problems [23]. The immediate reward is

$$r_t^e = \frac{\Delta_t^e - \Delta_{t+1}^e}{\Delta_0^e + \varepsilon_{\text{err}}} \quad (59)$$

and for $h > 1$ the selected method uses the supported lower-horizon successor set

$$\mathcal{A}_{t+1}^{(Q,1)} = \{i : m_{t+1,i}^{\text{act}} = 1 \wedge m_{t+1,i}^{Q,1} = 1\} \quad (60)$$

The set $\mathcal{A}_{t+1}^{(Q,1)}$ contains successor rows that are both hard-valid and supported by a persisted finite one-step reward label. This admission mask is shared across requested horizons; the network is queried at $h - 1$, but replay does not certify a separate $(h - 1)$ label for every successor row. The online parameters θ select one row from that set, while the delayed parameters θ^- evaluate the selected row. The following display gives the legacy countdown target and continuation rule. The continuing profile uses the same online selector and delayed evaluator only on recorded, supported successors; its exclusions are specified below:

$$y_t^{(h,e)} = r_t^e + \gamma B_t^{(h,e)} \quad (61)$$

$$\begin{aligned}
i_{t+1}^* &= \operatorname{argmax}_{i \in \mathcal{A}_{t+1}^{(Q,1)}} \hat{Q}_{\theta, h-1}^{\sigma, \Xi}(z_{t+1}^\sigma, e, q_{t+1, i}), \quad \text{when the support is nonempty} \\
B_t^{(h, e)} &= \begin{cases} \hat{Q}_{\theta^-, h-1}^{\sigma, \Xi}(z_{t+1}^\sigma, e, q_{t+1, i_{t+1}^*}) & \text{if } h > 1 \wedge d_t = 0 \wedge \mathcal{A}_{t+1}^{(Q,1)} \neq \emptyset \\ 0 & \text{if } h = 1 \vee d_t = 1 \vee \mathcal{A}_{t+1} = \emptyset \\ \perp & \text{if } h > 1 \wedge d_t = 0 \wedge \mathcal{A}_{t+1} \neq \emptyset \wedge \mathcal{A}_{t+1}^{(Q,1)} = \emptyset \end{cases} \quad (62)
\end{aligned}$$

For a two-row successor, suppose the online values are $(0.30, 0.45)$ while the delayed values are $(0.36, 0.31)$. The online network selects the second row, so the continuation is 0.31; it is not the delayed network’s own maximum 0.36. In the legacy countdown profile, a terminal flag, $h = 1$, or an empty actor-valid successor mask reduces the target to the immediate reward. Its narrower support failure occurs when actor-valid successor rows exist but none has one-step label support; that target is excluded from loss. The continuing profile admits recursive rows only when the next state is recorded and has finite, hard-valid reward-label support, querying that state at $h - 1$ regardless of its scheduled query. Missing records and empty or label-unsupported successor sets therefore supply no recursive loss in that profile. Neither profile reconstructs unrecorded continuation. The structural rule is $Q_h \leftarrow Q_{h-1}$, so a horizon never bootstraps from itself. This Double-Q selector/evaluator split follows DDQN-style online selection and delayed evaluation, rather than two independently trained alternating estimators. It may limit maximization bias [37], but it cannot repair unsupported actions, aliased state, privileged selection, or a missing selected successor. These are separate from the bias addressed by the estimator.

Figure 18 connects the replay asymmetry to this target: dense immediate labels supervise candidate rows, whereas recursive targets follow the selected factual successor. The support check precedes online selection and delayed evaluation.

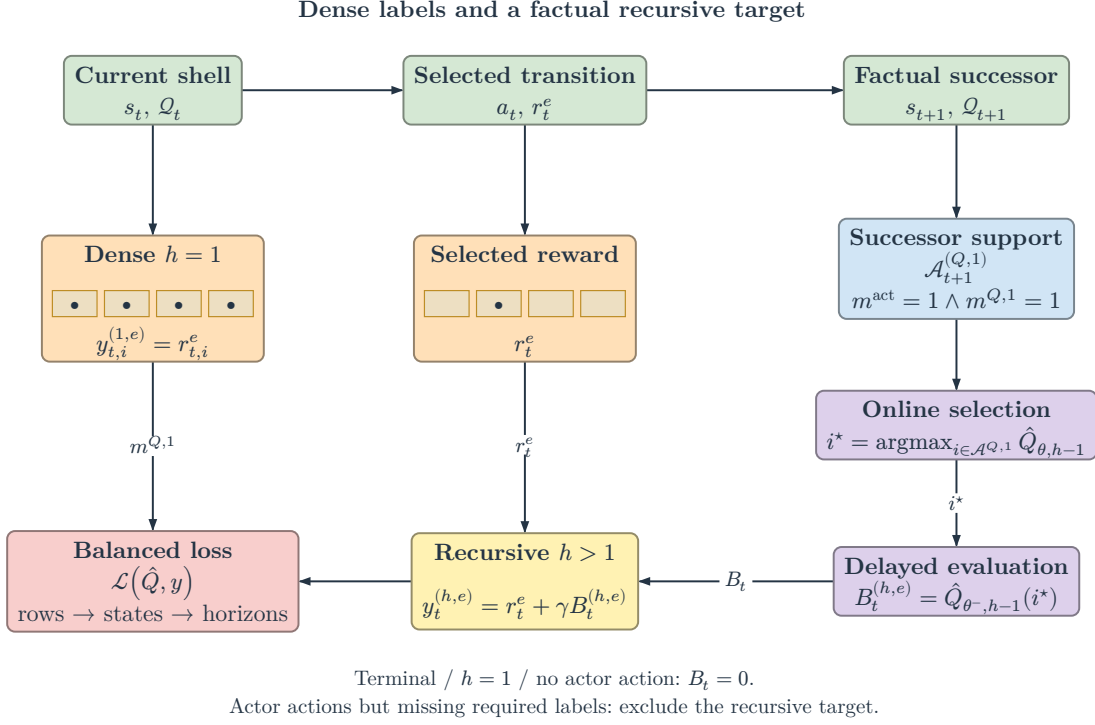


Figure 18: Evidence used by finite-horizon fitted-Q learning. Dense $h = 1$ labels and selected-transition $h > 1$ targets have different row support. Successor admission intersects hard action validity with finite one-step reward-label support. Online selection and delayed evaluation use that same set. Missing required labels exclude a recursive target; they do not imply zero continuation.

The recursion estimates greedy continuation over the generated, hard-valid finite support under the named target-source, actor-state, generator, action-mask, reward, discount, and horizon protocols. Unlike an immediate per-row target, the $h > 1$ continuation value changes when the successor-support protocol changes because that protocol changes the bootstrap maximum. This is off-policy continuation: the recorded first transition supplies the next state, but the learned selector can choose a different next action from the collector. DDQN changes how that learned choice is evaluated; it is not what makes continuation off-policy. The resulting target is therefore distinct from a realized collector return, and it is not an optimum over ungenerated continuous camera poses. This distinction fixes the meaning of a successful fit before any policy comparison is attempted.

4.9.6 Checking the two-step training target

When the selected action has a successor table with dense one-step labels, its finite-support two-step target is computable exactly:

$$y_t^{(2,\text{exact})} = r_t^e + \gamma_t \max_{i \in \mathcal{A}_{t+1}^{(Q,1)}} r_{t+1,i}^e \quad (63)$$

Here γ_t is the persisted transition discount that scales the continuation reward. Terminal status and successor-label support are checked separately; a missing label does not become a zero discount.

Horizon two is the first recursive backup that can be checked against this finite-support reference. A unit test can inject exact Q_1 values and prove that the tensor path implements the equation. A frozen population test instead measures learned recursion error,

$$\varepsilon_t^{(2)} = \left| y_t^{(2,\text{recursive})} - y_t^{(2,\text{exact})} \right|, \quad \varepsilon_t^{(2)} \leq \tau_{\text{abs}} + \tau_{\text{rel}} \left| y_t^{(2,\text{exact})} \right| \quad (64)$$

on held-out supported rows. This certification compares the fitted backup target with the exact reference; it does not compare the scorer’s predicted Q_2 with that reference. It tests the learned Q_1 approximation, state encoding, masks, successor linkage, and recursive target construction together. The target is exact for the frozen factual successor table and reward contract; the learned $\hat{Q}_{\theta,h}^{\sigma,\Xi}$ remains a representation-conditioned predictor. Agreement supplies evidence about the backup construction. Learned Q_2 prediction accuracy, actor-continuation value, and endpoint policy benefit need separate tests; none follows from this certification or proves Markov sufficiency.

The held-out evaluation begins with the complete eligible chain population rather than only rows on which an exact target is available. Chains that terminate or lose support before $h = 2$ remain in the denominator with zero exact rows. Scene- and target-stratified coverage, error, uncertainty, numeric tolerance, candidate width, generator, recipe, and behavior policy are frozen before inspection. Low error on a small supported subset cannot justify a longer horizon when minimum support or coverage fails.

Exact targets support prediction checks; they do not establish that the resulting policy improves reconstruction. Policy evaluation compares endpoint errors on the same target tasks, reference geometry, scoring protocol, and acquisition budget. These matched comparisons do not require positive oracle headroom or training a learned-myopic control. The endpoint quantities and optional relative comparison to a named reference policy are defined in [Section 3.4.4](#). Endpoint errors remain necessary for interpreting that ratio, especially when the reference error is small or zero.

4.9.6.1 Oracle continuation and actor-achievable value

Exactness for a stored successor shell does not establish that the actor can identify its maximizing action. Consider two equally likely hidden scene configurations with identical actor-visible history and the same two candidate actions. Suppose opposite actions yield unit gain in the two configurations and the other action yields zero. An informed oracle obtains one by selecting separately in each scene; any actor restricted to the shared history has expected gain one half. This constructed example isolates an information advantage, not a measured defect of the present model.

The oracle maximum in an exact two-step target must therefore be distinguished from continuation selected using actor-visible information and subsequently evaluated by the oracle. Maximizing after observing hidden scene information can exceed maximizing expected gain before that information is available. Baisero and Amato analyze a related distinction between state-based and history-conditioned critics under partial observability [\[40\]](#); their results do not directly quantify this thesis’s oracle-reference gap. Here the consequence is interpretive: an unclosed gap to oracle continuation cannot by itself be attributed to failed learning.

The proposed continuation diagnostic in [Section 5.4](#) holds the successor shell fixed and separates actor selection from oracle evaluation. A comparison to one learned actor measures its gap to the oracle; it does not identify the optimal actor

policy or an irreducible information gap. The ideal history-conditioned value and the oracle-supervised diagnostic retain their respective information assumptions.

The resulting evidence boundary is explicit. The current implementation supports the $s_t^{\text{SO-POSE}}$ scorer, direct objective, and recursion under privileged oracle action support, together with the exact- Q_2 evaluation path. The scientific target additionally requires an evaluated actor-observed corpus, a causal observation-updated state, actor-visible or calibrated action support, a frozen held-out exact- Q_2 receipt, and matched endpoint comparisons. Oracle headroom and normalized recovery remain optional diagnostics. The actor-observed input admission, actor-only construction, and writer–reader path are implemented, but that route has no frozen corpus or evaluation and has not passed these evidence gates. Until they pass, this chapter establishes an executable method and its falsification tests, not a successful non-myopic actor. The development register retains the alternatives as one-factor responses to diagnosed failures. Their activation conditions are hypotheses; none is evidence that the corresponding failure has been observed.

5 Experimental Design

The experiments first test whether the target-specific measurement is repeatable and whether the scene, target, candidate and replay populations provide representative support for learning and evaluation. They then compare descriptors, volumetric readouts and selected-observation fusion under matched information, separating representation effects from additional observation access. Finally, they evaluate finite-horizon value predictions and learned acquisition policies against declared references. Independent prediction accuracy and fixed-budget endpoint performance answer distinct questions. Bounded oracle lookahead supplies an optional diagnostic of delayed opportunity; relative endpoint-error reduction and legacy recovery retain their distinct reference policies and denominators.

The scene is the independent experimental unit. Confirmatory policy comparisons therefore use scene-disjoint data, matched target and candidate conditions, equal acquisition budgets, and independent endpoint oracle evaluation. Generation throughput, invalidity, support, and resource use remain necessary feasibility evidence, but they cannot substitute for paired policy outcomes. This design keeps architectural capacity and systems readiness distinct from evidence that non-myopic target reconstruction improves the endpoint outcome.

5.1 Study Population and Evidence Gates

A valid comparison must distinguish better decisions from easier tasks. The study therefore fixes the scene and target population before comparing candidate profiles or learned policies. Additional views from one scene increase the observations available for learning, but not the number of independent environments on which generalization is assessed.

Validation TODO: Preregister the eligible population, exclusions, primary estimands, aggregation unit, number of independent runs, uncertainty interval, minimum meaningful effect, and multiplicity policy before inspecting confirmatory results.

Source: experiment manifest and analysis specification

Gate: immutable analysis plan plus matched held-out policy table

5.1.1 Candidate support and paired profile comparison

The RQ1 support analysis treats the scene as the independent unit. Candidate-support observations are computed per state, averaged within scene, and then macro-averaged across the cohort. These within-profile diagnostics are descriptive. A comparative claim instead uses paired per-scene differences on eligible scene/target/root tasks observed under both profiles, with equal candidate budgets and each profile’s frozen generation settings. Its uncertainty is assessed across the paired scene differences, not across candidate rows.

Missing or failed states remain in the audit. Support counts and fractions use the preregistered failure values in [Table 27](#). An undefined direction is not replaced by an invented angle; eligibility for a geometric contrast and its excluded-scene reasons are reported separately. The descriptive support summary and paired profile comparison may therefore refer to different populations and answer different questions.

Minimum valid support reports the worst observed case, while its fifth percentile describes the lower tail. Actor-valid fraction, configured-family zero rate, side balance, and circular span characterize complementary aspects of the proposals. Projection, observed oracle opportunity, and jitter compliance are secondary diagnostics. Projection measures framing rather than visibility; opportunity describes the best labelled immediate gain rather than a policy effect; jitter compliance checks the declared production perturbations. None alone establishes wearer-motion realism or identifies the cause of a poor endpoint result.

The source population comprises ASE/ATEK snippet windows from scenes with the configured [GT](#) mesh and object-box table. Entire scenes are assigned to training, validation, or test before model selection; another snippet cannot carry a scene across these boundaries. Each run records counts of scenes, snippets, admitted tasks, chains, transitions, and candidate rows. A capped training-source pilot can characterize support and cost, but cannot estimate held-out policy performance.

5.1.2 Task population and information available to each policy

The legacy ground-truth task control samples geometry-valid reference OBBs and reports the GT pool, selected tasks, classes, scenes, and later measurement failures. The implemented observed-target route instead begins with detector descriptors and records their matching and admission outcomes. Its actor instruction remains separate from the GT geometry used for crops and labels. Implementing this route does not establish either a frozen observed-target corpus or an end-to-end actor-visible policy result.

These are different populations and must remain separate in both the manifest and interpretation. Ground-truth geometry may define evaluation and privileged reference policies, but it cannot silently define the learned actor’s target, action support, or scene evidence. The boundary in [Section 3.2](#) also applies to rendered depth: hypothetical renders supply labels, and selected privileged renders belong only to named non-deployable controls. Teacher policies, distillation, and current-belief rendering remain planned alternatives until their distinct information roles are implemented and evaluated.

5.1.3 Representation commissioning before horizon claims

Engineering commissioning and scientific promotion answer different questions. Commissioning first corrects measured-support semantics, binds the admitted population and source protocol, and verifies dense immediate-value fitting at every realized causal prefix with its actual remaining budget. A small overfit and resource smoke check establish a working fit path. Historical usefulness screens are not prerequisites to starting that fit, and prefix reuse or reuse across horizon queries are optional optimizations justified by measured cost. Actor-visible policy promotion still requires the research-question evidence and matched endpoint evaluation defined below; oracle opportunity remains a separate diagnostic.

The architecture study first asks whether the representation predicts immediate candidate value under matched information. Its initial selected-depth diagnostic uses privileged mesh-rendered observations; the actor-visible state comparison belongs to RQ2 and requires a separately matched observation protocol.

The primary Q1 comparisons use the same scene split, target protocol, candidate generator, admitted observations, labels, optimizer protocol, and evaluator. Root-only compact S0 is compared with a root-only volume to estimate the matched-input architecture package effect. Their encoders, readouts, state-token topology, and capacity differ, so this contrast alone cannot isolate spatial retention. For states with admitted selected observations, a pooled selected-depth control is compared with the updated volume under the same prefix. Root-only versus selected-depth is reported separately: an updated volume with more observations cannot identify a spatial-readout benefit over a root-only compact model.

Model comparisons also declare training steps, tuning-trial budgets and random seeds, alongside parameter count, memory and inference cost. Match optimization effort and model-selection opportunities where the intended contrast requires them; report residual differences as part of the model package. A resolution study deliberately changes grid resolution while fixing the other admitted evidence and evaluation conditions.

Report prediction error and action quality separately at root states and at states containing selected observations, and by supported horizon. Otherwise a change in source availability or rollout stage can masquerade as better value prediction. If this confound appears, a control using only source availability and stage can test the shortcut before adding another representation arm.

Within the volume family, deterministic geometric pooling and one candidate-local spatial-attention block use the same encoder, physical samples, mass features, and downstream decoder. An optional pretrained-feature arm requires the head-versus-neck interface decision recorded in [Section 4.1](#). It holds the geometric evidence and readout fixed, reports feature storage and inference cost, and can be tested at Q1 without waiting for Q2. This separates feature reuse from readout and horizon changes. Q1 calibration, within-state ranking, greedy regret, latency, and peak memory are the commissioning outcomes. These comparisons test representation and readout hypotheses; they do not establish delayed planning benefit or endpoint policy recovery.

Scene diversity, rather than the product of snippets, candidates, and rollout futures, determines the independent evidence available for model selection. Overlapping windows and multiple candidate labels improve coverage within a scene but remain correlated observations. All validation summaries therefore aggregate within scene before uncertainty across scenes, and the untouched test scenes are reserved for the final selected configuration.

5.1.4 Oracle opportunity and the learned comparison

Two policy comparisons answer different questions. Oracle lookahead versus one-step oracle-greedy tests whether the generated finite support exposes useful delayed gain:

$$\Delta_{\text{look}} = J_e^{(H)}(\pi_{\text{oracle-look}}) - J_e^{(H)}(\pi_{\text{oracle-1}}) \quad (65)$$

The learned comparison measures fixed-budget endpoint gain against declared reference policies. It may additionally express relative improvement through $R_e^{(H)}$, the fraction of a reference policy’s remaining endpoint error removed by the evaluated policy. The reference error is reported alongside the ratio; zero makes the ratio undefined and near-zero values make it unstable. Neither the endpoint comparison nor this optional normalization requires positive oracle headroom.

A target-conditioned learned-myopic scorer is an optional matched-information control. If included, it is evaluated through calibration, candidate ranking, oracle-rescored selections, and fixed-budget endpoints. The historical scene-level VIN scorer is not itself this target-conditioned control. The legacy ratio η_Q can then describe improvement over that control relative to its gap to bounded oracle lookahead. It remains a separate diagnostic, with no interpretation as recovered opportunity when its reference gap is small or nonpositive.

Evaluation readiness and policy outcomes remain separate. Oracle comparisons do not require a neural myopic control, and learned-policy comparisons do not depend on a positive oracle result.

Policy success is assessed with matched endpoint evaluation, not predicted values or training loss. A completed null lookahead comparison is bounded by

its scenes, target protocol, generator, horizon, search width, and validity regime. A gap between a learned policy and an oracle reference can reflect target observability, available actions, continuation coverage, representation, or approximation. The distribution and support diagnostics separate these explanations before a larger model or a new collection strategy is prescribed.

Question	Comparison	Outcome	Population and limit
RQ1: measurement and experience	Matched acquisition/scoring; proposal profiles and collection rules	Error consistency, valid actions, continuation coverage and cost	Declared target/root cohort; held-out coverage differs from train-only yield
RQ2: representation and scorer	Matched observations for compact/spatial and readout controls; added-observation contrast separately	Independent value error, ranking, selected-action regret and resources	Scene-disjoint data; root/successor and horizon strata; disclose package effects
RQ3: learning and selection	Supported horizon/learning variants; learned policy against named references	Independent numerical checks and fixed-budget endpoint effects	Physical-scene pairing, supported queries and shared information rules

Table 6: Research questions, comparisons and evidence. Exact arm choices and independent references remain the decisions stated in the adjoining protocol.

5.2 Replay Eligibility and Finite-Horizon Learning Gate

The factual rollout tables determine which records may supervise each learning problem. Five predicates answer different questions. **Proposal validity** says that a materialized row passes the declared proposal and hard-geometry checks. **Action support** says that the admitted actor protocol permits selecting it. **Label support** adds an admitted target-label state and a finite root-gain label. **Selected support** marks the one factual action taken on this branch. **Bootstrap support** admits its continuation: a reproducible successor and the supported lower-horizon candidates required by the objective. Padding and modality presence remain separate structural predicates.

A feasible side view whose render fails can remain an action while lacking a label; an unselected labelled view can supervise immediate gain while lacking a factual successor. Selection uses action support, immediate loss uses action and label support, and recursive loss uses the selected transition and its continuation

evidence. The diagnostic RRI audit has its own availability predicate, so a missing auxiliary metric does not remove a finite root-gain label. Exclusion reports retain the failed predicate for each operation. Equal counts in two masks do not establish that their definitions, information contracts, or label populations are identical.

Dense one-step supervision requires finite immediate labels for every action in the declared comparison support. Legacy subset supervision instead learns only from its labelled eligible rows; completing a rollout does not make that subset dense. Exact two-step supervision is narrower: the factual first action must have a stored reward and either an explicit terminal outcome, whose continuation is exactly zero, or a valid successor step with finite one-step root-gain labels for every hard-valid successor candidate. This completeness condition makes the masked successor maximum exact over the stored action set; one finite label is insufficient because an unlabeled hard-valid action could have the true maximum. Recursive supervision has a different continuation requirement: it uses a factual selected transition and immediate reward. When continuation is required, it evaluates a learned lower-horizon value on a recorded successor. The current bootstrap set contains hard-valid successor rows with finite one-step labels; it does not certify a factual multi-step return for each of those rows. Exact two-step supervision instead needs complete immediate labels over the stored hard-valid successor set, so that no unlabelled feasible action could change its maximum. The derived finite-horizon arrays align these fields on a padded state-candidate view; they do not create labels for unobserved transitions, make selected GT depth actor-visible, or turn sparse long-horizon action support into dense support.

Learning surface	Purpose	Minimum factual evidence
dense $h = 1$	immediate candidate value	finite one-step root-gain labels on every action in the declared support
exact $h = 2$	base-case finite-support value	selected reward plus either terminal outcome or complete finite one-step labels over the hard-valid successor action set
recursive $h > 1$	variable-horizon fitted value	selected reward and boundary convention; continuation uses a stored successor, hard-valid rows with finite one-step labels, and a learned lower-horizon estimate
behavior return	policy-conditioned Monte-Carlo control	complete rewards for the declared query, or a verified task boundary, plus behavior-policy identity

Table 7: Required replay evidence by learning target. Dense immediate labels, exact two-step targets, recursive optimal-continuation targets, and behavior-policy returns are distinct supervision surfaces.

The measured root/replay compatibility and batch probes are recorded in [Section C.2](#). Readable storage and successful joins do not establish admission to a particular learning objective.

Activity or claim	Required evidence	Limit
Executable scoring	compatible checkpoint, declared input/state protocol, hard support and supported query	A runnable scorer does not establish predictive accuracy or policy benefit.
One-step prediction	finite held-out immediate labels; source-matched inputs, scene split and prediction/ranking measures	Does not require a two-step or policy-endpoint result.
Multi-step prediction	supported targets for each declared horizon and budget; independent reference matching the stated continuation	An exact-Q2 backup check certifies the target construction, not the learned prediction.
Observation-updated state	typed causal reader, deterministic fusion, source masks and no-future-observation tests	Construction tests do not establish state sufficiency or policy benefit.
Policy improvement	held-out paired endpoints under equal budget and a matched evaluation contract	Policy outcomes do not establish unbiased numerical values. Optional myopic controls require their own frozen scorer.

Table 8: Claim-specific evidence requirements. Executable scoring, held-out prediction and acquisition outcomes answer different questions.

Implementation: partial **Evidence:** pending

The scalar requested-horizon A0/A1 controls, dense-Q1 plus selected-recursion Double-Q learner, feasibility auxiliary, manifest-bound trained-horizon support, and bounded stratified exact-Q2 certification surface are implemented. A1 remains the default; a qualifying population receipt, task-sufficient dynamic state, and comparative policy evidence remain unavailable.

Literature: [17], [21], [22], [23], [37]

Promotion gate: populated held-out exact-Q2 receipt, supported $H > 2$ targets, compatible checkpoint, frozen state protocol, and independent held-out oracle re-evaluation

Development source: `aria_nbv/aria_nbv/data_handling/qh_data/batching.py`; `aria_nbv/aria_nbv/lightning/finite_horizon/datamodule.py`; `aria_nbv/aria_nbv/lightning/finite_horizon/module.py`; `aria_nbv/aria_nbv/lightning/finite_horizon/objective.py`; `aria_nbv/aria_nbv/lightning/finite_horizon/certification.py`; `aria_nbv/aria_nbv/rollouts/storage/qh_reader.py`

Method defines the scorer and externally masked selection rule. The masked argmax is already the discrete decision rule. A separate actor network and online data collection are not required to train or execute this finite-candidate policy. Batch fitted Q iteration explicitly learns a greedy Q function from a fixed collection of transitions by repeatedly solving supervised regression problems [23].

5.2.1 Prediction support and independent backup checks

The comparisons use Method’s immediate and recursive targets (Section 4.9.5). Immediate supervision includes finite hard-valid labels at each realized state; recursive supervision needs its selected transition and the successor support required by the chosen profile. Padding and parallel batching add no observations. The pure one-step control preserves factual budgets and acquired prefixes while omitting recursive targets and feasibility supervision, so it does not also test joint validity learning.

Two-step verification separates arithmetic from empirical agreement. Exact-table injection tests the backup calculation. A frozen-bundle comparison measures its error against the factual exact two-step reference on supported rows, using Section 4.9.6. Agreement checks the backup target, not the learned two-step prediction or policy outcome.

Physical scenes define the statistical clusters even when several stores or chains reuse them. Cross-store deduplication remains unverified; the minimum of five provenance groups in the existing check is neither proof of five independent scenes nor a power calculation. Missing support is unavailable evidence, not zero error. The sampling strata, profile-specific boundaries, support gates and batching procedure are retained in [Section C.5](#). Longer-horizon claims require independent outcomes; the terminal-step proxy alone cannot supply them.

5.2.2 DDQN-style and behavior-return controls

The Double-Q protocol uses Method’s online selector and delayed evaluator on the same supported successor rows. Horizon one has no continuation. Terminal and empty-action masks can also yield zero continuation in the legacy countdown profile; record absence can therefore leave an immediate-only target. The continuing profile instead excludes recursive rows without a recorded, finite-label-supported successor. Comparisons must name the profile and verify its admission, rather than equating a recording cutoff with a task endpoint. A simpler single-estimator frozen lower-horizon maximum remains a planned matched control until it has an executable configuration and receipt. Its comparison must hold state, action support and scalar horizon targets fixed.

As distinguished from the greedy value objective in [Section 4.9.5](#), a retained chain also yields a complete observed return for a declared query length, using the reward sum defined in Foundations and specialized in Chapter 3.

A shorter retained segment can describe the gain actually observed along that collector, but it is incomplete evidence for a longer query. Its unobserved suffix is not filled with zeros. A complete behavior-return comparison therefore requires continuity and the same measurement convention throughout the query, or an independently verified task endpoint.

These observed returns describe continuation under the named collector μ , rather than greedy finite-support continuation. A predictor fitted to them is a behavior-return control. Interpreting it as the expected Q^μ additionally requires appropriate coverage and information assumptions, and the mean-value check for

Huber fitting described in Method. Behavior returns from random-valid, greedy, softmax, and oracle-lookahead chains must therefore remain identified rather than pooled as if they represented one optimal value function.

For a fixed implemented oracle policy π^{oracle} , each complete chain supplies $G_{t,e}^{(h,\text{oracle})}$. Estimating its expected return requires repeated outcomes from the declared conditional population, with policy, initial information and decision protocol held fixed. Branches sharing a prefix and trajectories sharing a scene can carry common sources of dependence; their count alone does not establish independent support. Sampling weights and uncertainty estimates must respect the sampling design, such dependence, and any branch-selection probabilities. A deterministic replay of the same scene and seed can reproduce a return without adding a new outcome.

Selecting the best explored return answers a search question. Beam pruning retains the highest accumulated selection scores among explored partial paths; when those scores are reward sums, this selects strong realized returns within the explored tree. Its retained maximum is neither an unbiased estimate of a fixed policy’s expected return nor a certificate of $Q_{h,e}^{\star,\text{oracle},\Xi}$. Reward-scored search and expected-return estimation must therefore report their sampling and retention rules separately.

Open decision: ORACLE-BRANCH01 — Choose branch probabilities, branching depths, action-sampling policy and computation budget for a matched collection comparison. Compare a schedule favouring early branches with a flat depth schedule under the same rendering budget. Early branching explores different prefixes; late branching explores alternatives after a shared prefix. Which allocation supplies more useful continuation evidence depends on the task and measured acquisition cost.

Source: Implemented branching and beam collector; ideal-oracle and sampled-return distinction.

Gate: Freeze the estimand, depth schedule, probabilities, selection and retention rules, and equal-cost budget before comparing collection policies; preserve shared-prefix dependence in analysis.

DDQN-style estimation addresses one source of maximization bias. It does not create missing successor transitions, make unsupported actions reliable, or repair state aliasing. CQL and BCQ motivate explicit offline-support diagnostics because a greedy learned policy can select actions whose multi-step consequences are weakly represented in the behavior data [21], [22]. Conservative regularization is

introduced only if those diagnostics reveal systematic unsupported-action overestimation.

Research TODO: RRC10 — Decide whether to run or defer a bounded same-model one-step warm start, single-estimator continuation control, or Huber versus squared-loss comparison. Each asks a different optimization or mean-prediction question.

Source: Selected fitted-Q/Huber objective; reward and continuation research, 2 October 2026

Gate: Record a chosen comparison and its result, or explicit deferral; do not make these optional studies an architecture-wide obligation.

5.2.3 Horizon-balanced replay and evaluation

In a densely labelled corpus, one-step rows can vastly outnumber selected-action targets at longer horizons. An unweighted row mean can therefore favour the myopic task while hiding longer-horizon failure. The learning contract instead averages candidates within each state, states within each horizon, and then non-empty horizons uniformly. It binds this weighting to the maximum horizon, replay and candidate configurations, behavior-policy vocabulary, reward and return semantics, and discount. The evidence bundle records the realized state and candidate counts for every trained horizon, and online inference rejects horizons without positive training support. Alternative sampling or weighting schemes require a new versioned learning identity and remain in the development register.

The primary diagnostic panel reports, separately for each horizon, admitted states and candidate labels, physical-scene and target coverage, state-averaged value loss and calibration error, and ranking and selected-action regret only where the required counterfactual comparison table exists. Counts accompany each diagnostic: no eligible comparison is unavailable evidence, not zero error. Dense one-step tables currently support ranking and regret; selected-transition recursion at longer horizons reports zero counterfactual comparison support. Candidate errors and pairwise accuracies are first reduced within factual state, and selected-action regret contributes one value per supported state.

The evidence bundle retains behavior and family composition, bootstrap and boundary fractions, online/target disagreement, and selected-chain outcome fields for diagnosis. Marginal RRI, cumulative RRI and cumulative root gain

remain distinct quantities. Held-out endpoint outcomes evaluate the resulting policy, rather than its fit to learning targets. A single validation loss can select a model only under a prespecified horizon aggregation. Cross-stage admission and checkpoint compatibility retain the contracts in [Table 8](#); diagnostic fields and full identity inventories remain in their reproducibility owners.

The optimality claim remains bounded under either interface. A learned finite-horizon value can approximate the best continuation only within the sampled finite candidate generator, hard-validity regime, represented actor state, and offline transition support. The same checkpoint cannot silently mix a pose-only state with privileged selected-depth, sensor-like, or actor-visible dynamic states.

Longer horizons can expose information lost by a pose-only approximation: two histories ending at the same pose may leave different surfaces observed and therefore different useful continuations. Whether that distinction matters is an empirical property of the task and represented state. Selected-observation geometry can preserve it, but geometry updates alone do not establish Markov sufficiency, and longer horizons do not universally require that representation. The pose-only control remains a declared approximation whose limits must be checked against the same causal information boundary.

Research TODO: Specify the selected-observation fusion comparison in RQ2: which evidence is updated, which representation receives it, and which factor changes between matched controls. Do not infer this protocol from the longer-horizon information-loss risk or from an implemented privileged state control.

Source: RQ2 representation study; Method actor-state contract

Gate: Declare the compared information states and align their manifests and chapter descriptions.

The readiness gates in [Table 8](#) separate executable learning infrastructure from supported inference and policy evidence. A dynamic state claim additionally requires deterministic selected-observation fusion and no-future-observation tests. Those tests establish causality; held-out matched comparison is still needed to establish a useful state representation.

5.3 Policy Comparison and Statistical Protocol

Implementation: partial **Evidence:** pending

The aggregation and artifact-driven reporting seam exists, but confirmatory paired policy evidence does not.

Promotion gate: frozen held-out manifest, completed paired endpoints, and validated report bundle

Development source: `aria_nbv/aria_nbv/rollouts/presentation/reports.py`; `docs/typst/thesis/experiment_data.typ`

The scene is the experimental unit. Every comparison is paired by scene, target task, root state, candidate-generation distribution, hard-validity regime, and acquisition horizon. Planner depth may differ because it defines the decision rule, but every policy receives the same acquisition budget. Repeated snippets, targets, and seeds are aggregated within scene before the primary comparison, preventing densely sampled scenes from dominating the estimand.

The primary estimand is the paired per-scene difference in fixed-budget endpoint target-quality gain, $J_e^{(H)}$. Its denominator $\Delta_0^e + \varepsilon$ also normalizes cumulative target-root gain, which equals the endpoint gain at the shared horizon for an undiscounted return. By contrast, state-relative RRI has a denominator that changes with the state and remains a secondary diagnostic. Before final aggregation, the analysis artifact freezes the scene aggregation, uncertainty procedure, interval level, comparison set, and minimum meaningful effect. Invalid-action rate, runtime, path length, and task or candidate coverage diagnose mechanism and feasibility; they do not replace the endpoint estimand. The current replay store alone cannot estimate this endpoint comparison because it does not persist an independent post-budget reconstruction for every policy; confirmatory analysis therefore requires matched endpoint oracle re-evaluation or an equivalent frozen endpoint record.

Support and failure strata are fixed before policy inspection. The report distinguishes source rows without an admitted target task, admitted tasks whose oracle evaluation fails, roots without a valid action, trajectories that terminate before the shared budget, and completed paired trajectories. A recording cutoff, an acquisition failure, and a protocol-defined task stop are reported separately.

An incomplete record cannot establish the error that would have remained at the shared budget. A genuine task stop may permit an endpoint measurement if the declared comparison treats that stop as final; this requires a measured reconstruction and an explicit stopping convention. The current shared-complete analysis gives unfinished or unmeasured cases no numerical endpoint gain, while retaining them in admission, completion and discordance denominators. Conditional endpoint-quality effects use only the shared-complete cohort, whose size is reported explicitly; incomplete outcomes are neither zero-filled nor silently dropped. Completion over all paired cells and endpoint quality conditional on shared completion are therefore reported together. This exposes whether an apparent policy difference is instead a support or systems failure.

Policy	Decision input	Acquisition budget	Scientific role
π_{rand}	actor-visible	H	valid-action lower reference
$\pi_{\text{learned-1}}$	actor-visible	H	optional learned myopic control
$\pi_{\text{oracle-1}}$	oracle immediate reward	H	one-step oracle-greedy comparator
$\pi_{\text{oracle-look}}$	bounded oracle lookahead	H	bounded privileged lookahead comparator
π_Q	actor-visible	H	planned learned finite-horizon policy comparison

Table 9: Matched policy comparison. All rows share the same acquisition budget; oracle access and planner depth define the decision rule rather than the acquisition budget.

The oracle diagnostic compares bounded lookahead against one-step oracle greedy. Every policy uses the same target mesh and crop, ASE-depth source, renderer and backprojection, fusion and point cap, and point-mesh metric configuration. Repeated evaluation must reproduce the scores within the declared numerical tolerance. This verifies deterministic execution under the fixed contract, not invariance to another mesh, sampling process, or metric. The learned comparison evaluates Q_H against declared reference policies under the same acquisition budget and endpoint metric, whether or not the oracle comparison finds positive headroom. An actor-visible target-conditioned myopic control can isolate the horizon comparison if it is included; other named references remain valid comparisons. Relative endpoint-error reduction is an optional companion to the

primary endpoint-gain difference. The named reference error is reported with it, including zero or near-zero cases where the ratio is undefined or unstable. The legacy learned-recovery ratio uses a different reference gap and remains optional. These learned-policy comparisons remain prospective because completed matched endpoint evidence is absent.

An optional fixed-scorer fusion comparison associated with [Section 1.3.2](#) compares causal observation updating with static scene evidence. Both arms track selected poses, use the same compatible frozen scorer, and regenerate candidates under the same rule. Only the update arm incorporates the selected actor-visible observation; neither arm retrains or adds transitions to the training set. The measured endpoint difference is the total closed-loop effect of updating: different decisions can produce different successor states and candidate tables. Holding the generation rule fixed therefore does not isolate a scorer-only effect on identical future candidates. Privileged mesh depth cannot establish the actor-observed claim.

Validation TODO: Populate evaluation-contract identity and deterministic repeatability, candidate and target support, and matched endpoints for each selected policy comparison. Report oracle headroom and any learned-myopic or legacy recovery diagnostics separately. Missing evidence blocks the corresponding claim rather than becoming a zero result.

Source: thesis objective-to-evidence contract

Gate: artifact-backed Results bundle

Validation TODO: Run row-shuffle, mask-isolation, duplicate-row, valid-count, frame-transform, target-source-dropout, and horizon-boundary tests for every model admitted to the policy comparison.

Source: geometric acceptance contract

Gate: architecture validity report

Open decision: Freeze the scene aggregation, interval procedure, interval level, reference-policy comparisons, and confirmatory decision rules in the resolved analysis manifest rather than in thesis prose. Keep exploratory diagnostics distinguishable from confirmatory tests.

Source: artifact-driven reporting plan

Gate: confirmatory analysis freeze

5.4 Distribution and support diagnostics

The following comparisons are proposed tests of the mechanisms identified in [Section C.8.1](#) and [Section 3.5.3](#), not completed experiments. Their settings are selected on training or development scenes before confirmatory evaluation. They retain the same endpoint objective and separate acquisition budgets from the oracle calls, rendered pixels, elapsed time, and memory consumed in obtaining training labels.

5.4.1 Target exposure under a fixed evaluation population

An optional target-exposure comparison would train with uniform target sampling and with a small observed-support-stratified challenger. Both use the same training scenes, model, optimizer, run-seed policy, and declared label-computation budget. Their endpoint evaluation uses the same reference target population, with scene-level aggregation and support-stratified reporting. Exposure counts, repeated object identities, and admission failures remain visible. Any later adaptive sampler must record its probability semantics and retain reference support rather than quietly replacing the scientific population with easier targets.

Observed-target admission is reported both over the prespecified target-task population and, separately, conditional on successful association. The end-to-end report retains admission failures with a predeclared fallback or failure outcome. An undefined oracle measurement is reported as missing, not converted to zero utility. A conditional observed-target result is never combined with an unconditional ground-truth-task oracle denominator. All arms of a learned-policy comparison use the same target population and action-support information contract.

5.4.2 Proposal opportunity versus selection regret

At a common root, let $\mathcal{A}$ be the baseline hard-valid candidate shell and $\mathcal{B}$ a broader diagnostic shell containing every row of $\mathcal{A}$. Both shells use the same feasibility and oracle reward operators. Write y_i for the measured root-normalized gain of row i , $i^\pi \in \mathcal{A}$ for the selected row, and $i^b \in \mathcal{A}$ for an incumbent baseline’s selection. Every hard-valid row in both shells needs a finite label before either maximum is

called exact. The reference opportunity and the within-shell decision then separate algebraically:

$$y_B^* - y_{i^\pi} = (y_B^* - y_A^*) + (y_A^* - y_{i^\pi}), \quad y_A^* = \max_{i \in \mathcal{A}} y_i, \quad y_B^* = \max_{i \in \mathcal{B}} y_i \quad (66)$$

The first nonnegative term is the **proposal gap**: the best reward made available only by the broader reference shell. The second is **within-shell regret**: gain already available in the baseline shell but missed by selection. For example, if the baseline shell offers gains 0.3 and 0.5, the policy chooses 0.3, and an added reference view offers 0.8, the proposal gap is 0.3 and within-shell regret is 0.2. The reference shell remains a finite diagnostic, not a continuous-pose optimum. If a valid row lacks a label, its shell maximum and the corresponding exact regret are unavailable.

Practical improvement asks a different question: how much better the selected view is than the incumbent's selected view under the same root and reward operator. It is $y_{i^\pi} - y_{i^b}$ and can be negative even when the reference shell contains a high-gain view. For a stochastic selector, the expectation under its declared action probabilities is a separate estimand from the gain of one realized selection. Policy advantage instead compares an action with the expected value of the same continuation policy. Its definition and finite-label limitation are retained in [Section C.15](#); it is neither shell regret nor an improvement against another policy's selected action.

The tolerance-regret diagnostic first fixes a completely labelled comparison shell and computes raw regret $r = \max_{i \in \mathcal{A}} y_i - y_{i^\pi}$. For a declared tolerance $\tau \geq 0$, practical regret is $r_\tau = r \mathbb{1}[r > \tau]$: it is zero at or below the tolerance and equals the full raw regret above it. For example, with $\tau = 0.005$, raw regrets 0.004 and 0.005 both become zero, whereas 0.006 remains 0.006. This threshold introduces a discontinuity; it does not subtract τ from nonzero regret. Practical-regret reduction is baseline practical regret minus learner practical regret, so a positive value favors the learner. It can differ from the incumbent gain improvement above because each regret is thresholded separately.

The recorded one-step diagnostic compares the finite labelled action subset and reports how many actor-supported labels are missing. Its maximum is exact only over that admitted comparison shell, not over unlabelled feasible alternatives. Within this shell, a tied predicted maximum selects the lowest admitted full-shell index. A constant scorer therefore defines a deterministic selection baseline even though its rank correlation is undefined. Changing that rule changes the baseline and cannot revise a frozen effect-screen verdict.

Collector comparisons change which transitions are recorded, not the value target defined in [Section 4.9.5](#). A different temperature or continuity weight therefore tests experience coverage; it does not turn the fitted-Q objective into regression on sampled collector returns. Interpretation must keep this change separate from reward or encoder changes.

At horizon two, this comparison must additionally fix or explicitly vary the continuation protocol. Changed successor generators require new continuation evaluation, not reuse of old-generator value labels. Exhaustive enumeration can establish a finite-support optimum; branch and beam truncation provide only a bounded comparator. Search-width sensitivity is therefore needed before a null bounded-lookahead result is interpreted as absent planning opportunity. Policies share a common root shell and the same declared generator and validity rules thereafter; different successor states need not have identical rows.

5.4.3 Ranking, prediction error and selected-action regret

Candidate ordering, numerical prediction error and regret at the chosen maximum answer different questions. A high overall rank correlation can still choose the second-best view. Constant scores require an explicit tie rule; retrospective affine association does not demonstrate predictive calibration. Undefined statistics retain their row/state support counts. Physical scenes remain the independent units for paired uncertainty. Exact rank/affine definitions and the retained pair-count estimator are in [Section C.15](#).

Factual-start throughput is a separate quantity-per-cost diagnostic. For one chain with L selected acquisitions, the count of overlapping length- h windows

before additional training masks is $C_h(L) = \max(L - h + 1, 0)$. A complete 16-acquisition chain supplies five length-12 windows and one length-16 window. For attempts indexed by a , its rate is $Y_h = 3600 \frac{\sum_a C_h(L_a)}{\sum_a \tau_a}$, where τ_a is attempted worker/receipt time in seconds. Failed or rejected attempts contribute cost and no usable store windows; missing or zero total cost leaves the rate unavailable. Overlapping windows are not independent scenes, and this rate weights no window by reconstruction quality. A lower rate can therefore be an acceptable cost of better endpoint utility or diversity. Report completion, validity, quality and attempted cost separately rather than using throughput as a universal admission threshold.

5.4.4 Collection coverage and the information boundary

A learner-driven collection experiment is considered only after a usable proposal opportunity is established and continuation coverage is implicated. It compares the fixed offline reference with equally costly additional exploratory and learner-driven collection. The latter mixes learned selection with declared exploration on training scenes, retains earlier replay, and refits the existing value model. The additional exploratory arm distinguishes better allocation from merely receiving more data. Test scenes do not enter collection, sampler tuning, or early stopping.

Research TODO: Specify a matched learned-policy versus immediate-RRI collector comparison as a pipeline robustness diagnostic: freeze the scorer and information boundary, match tasks and acquisition cost, and report completion, support and measurement failures separately. A change in collector alone does not prove reduced value bias. Keep refitting and additional-data effects in the separate exploratory-collection comparison.

Source: RQ1 pipeline robustness; RQ3 learned selection

Gate: Declare collector policies, measurement conventions and matched population before interpreting their outcomes.

The report distinguishes dense one-step label coverage from factual multi-step support for policy-selected actions. Family, target-relative geometry, and horizon strata localize independently oracle-rescored regret; representation-distance summaries are overlap diagnostics, not certified support probabilities. On completely labelled successor shells, an additional comparison selects continuation using only actor-visible information and then scores it with the oracle, alongside the oracle maximum. This separates privileged selection from actor-visible selection without

attributing the entire difference to information rather than approximation error. Exact-two-step support and endpoint effects, rather than training loss alone, determine the interpretation.

A proposed two-step diagnostic can use one frozen scorer to choose both the first action at $h = 2$ and continuation at $h = 1$. Oracle evaluation then measures the return of those choices, rather than replacing continuation with an oracle maximum. If a chosen action lacks the required label or successor, that query is unsupported; choosing another action after inspecting label availability would change the policy being evaluated. The current exact-two-step certification checks a different object, the fitted backup target. An earlier-state two-step diagnostic additionally needs the pair-specific admission and evaluation described in Method; it is not presently implied by support at $(b_t, h) = (2, 2)$.

These tests are separate from the optional fixed-scorer fusion comparison associated with [Section 1.3.2](#). Protocol transfer to a new proposal profile is reported separately from unseen-scene generalization under the original profile. A stable negative comparison is informative; an incomplete run or missing upstream support is not a zero effect.

5.4.5 Conditional capacity check

Capacity or data-fraction pilots are relevant only if they prevent interpreting the selected representation comparison. Their settings and unresolved H434-CAPACITY01 decision are retained in [Section 9.8](#); completing the full sweep is not a prerequisite for reporting a Q1 result.

5.5 Interpreting the optional oracle comparison

Positive, stable lookahead headroom would establish that the frozen finite-candidate setup contains exploitable non-myopic structure; it would not establish that the learned actor can identify it. A supported null would instead be a negative result for that setup, not evidence that all target-aware planning is myopic. If the evaluation contract is mismatched or non-reproducible, or if paired support is insufficient, the effect is not estimable for that comparison. Learned-policy

endpoint effects remain a separate question and do not depend on a positive oracle-headroom result. Training-only throughput pilots and failed generation attempts can guide compute configuration, but cannot establish headroom or policy superiority.

6 Results

Validation TODO: Replace fixture- and readiness-oriented output with confirmatory results that answer each research question using matched endpoint evaluation, denominators, exclusions, aggregation units, independent-run uncertainty, and immutable artifact provenance.

Source: confirmatory report bundle and analysis manifest

Gate: submission evidence bundle passes all report assertions and every stated result resolves to raw and derived artifacts

This chapter distinguishes bounded development observations from the principal comparisons still awaiting evidence. Table 10 maps that distinction to the three research questions. Each subsection retains its own population and protocol; these cohorts are not pooled.

Question	Study/population	Observation	Claim limit
RQ1	Matched renderer: five scenes, two seeds each	Candidate decisions, selected observations and audited endpoints agree	Only the tested recipe; runtime screen does not establish universal throughput
RQ1	Historical 77-root attempts; later 97-scene collections separately	Retained trajectories and explicit support/measurement exclusions	Historical evaluator and source-specific admission; nested rows are not new scenes
RQ1	Second R1 re-analysis plus incremental CUDA chains and one diagnostic	Family-specific supervision and conditional late-acquisition support	Overlapping scenes; no collector intervention or mixed-host admission
RQ2	Proposed matched representation comparisons	Comparative benefit remains unmeasured	Implemented encoders and input joins do not establish an advantage
RQ3	One-step screen: 21 scenes	Practical-effect threshold missed; exploratory ranking signal	Negative screen retained; neither multi-step nor actor-visible policy evidence
RQ3	Larger one-step analysis: exposed validation, one completed fit	Action choice and numerical prediction error diverge	Fixed weights/checkpoints; no seed or selection uncertainty
RQ3	Held-out multi-step prediction and learned-policy comparisons	Results unavailable	Missing evidence is not a measured null

Table 10: Chapter-wide evidence by research question. Independent development artifacts have their own provenance; the confirmatory report interface is documented in the reproducibility record.

6.1 Does candidate cropping preserve the measured transition?

Reducing the rendered mesh is useful only if it preserves the candidate choices and the observations that construct later states. A matched 16-step diagnostic therefore compared full-scene and cropped candidate rendering on five fixed physical scenes, with two seeds nested within each scene. Both arms used the same 60-candidate recipe, exact zero-margin target-box scoring, float32 geometry and zero rasterization blur. Candidate images were 48×48 pixels. Selected causal observations used the original full scene at that resolution: the cropped arm reacquired its selected view, while the full-scene arm reused its candidate observation to construct the successor. Separate full-scene 240×240 depth was retained for inspection and visualization.

All 20 intended rollouts completed, yielding ten qualified pairs and 160 matched decisions (320 arm-step records). Candidate scores had maximum absolute error zero under the frozen absolute tolerance 10^{-6} and relative tolerance zero. Candidate identities, probabilities, ranks and ties agreed; selected observations, prepared histories, raw endpoint clouds and target references agreed exactly. All twenty independent CPU endpoint audits passed, and all seven endpoint metric deltas were zero. These results establish agreement for this finite recipe and roster; they do not establish equivalence for arbitrary cameras, meshes or selection policies.¹

The earlier fixed-state rendering check established agreement on its recorded states only. The first whole-rollout checkpoint completed twelve attempts; all six reached pairs failed trace equivalence, and eight intentions were withheld. That failure remains retained as failure evidence. The completed contrast used the camera repair published in the camera-stability repair, with acquisition source `87bdc7b8be7a2c55918807dc43213e17567eba8`. Because both completed arms use

¹Frozen rendering contrast and primary readback: camera-stable renderer evidence, especially the original manifest, analysis and summary.

exact zero-margin scoring, this rendering result does not compare that metric with historical margin or any-vertex scoring.²

For each scene and seed, the runtime ratio divides the full-scene parent-worker duration by the cropped duration. Averaging the two ratios within each scene, then averaging the five scene means, gives approximately $2.21\times$; individual pair ratios range from $1.1023\times$ to $3.8161\times$. The same scene-equal reduction of cropped-minus-full durations gives -315.0522 seconds. The parent-worker clock includes diagnostic capture, selected-view retention, endpoint measurement and store writing. It excludes frozen preflight, the supervisor checkpoint and final readbacks. GPU acquisition was sequential and exclusive, while CPU readback and light publication work overlapped part of acquisition without a separately measured per-worker correction. These are descriptive local costs with no population confidence interval: seeds and decisions are nested within five scenes.

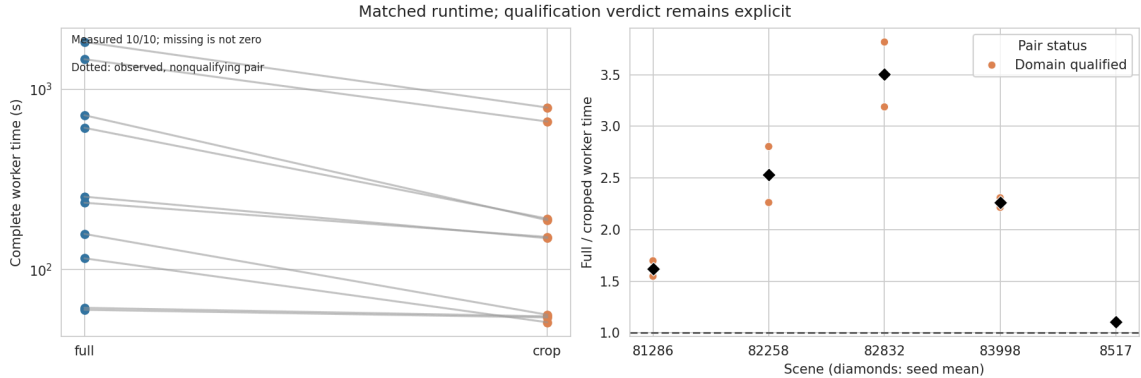


Figure 19: Paired parent-worker lifecycle costs for five scenes, with two seeds per scene. The left panel connects each matched full/crop duration on a logarithmic seconds axis; the right panel shows seed ratios as orange points and within-scene means as black diamonds. Each pair uses the same scoring and selected-observation contract. The recorded seconds include capture, retained selected depth, endpoint measurement and writing; the pairs describe this finite roster, with no population interval. The sealed plot is reproduced unchanged.

6.1.1 Bounded generation-runtime attribution

Six eight-step rollouts on three fixed scenes compared ordinary execution with capture/profiling. All independent endpoint audits passed and each scene’s pair retained identical decisions, stored observations and endpoint geometry. Mesh-preparation probes were too small to justify a larger cache. Their host clocks

²Preserved earlier evidence: fixed-state screen and failed whole-rollout checkpoint.

overlap other operations, so they identify a bounded preparation cost, rather than a universal bottleneck or speedup. The full clocks, source and execution context are retained in [Section C.17](#). No cache patch or proposed optimization contrast was selected.

6.2 Historical collection in context

The earlier September collection retained 186 validated stores from 75 roots, with 2,749 selected acquisitions. Twenty-one recorded states had fewer than two valid actions: completing trajectories did not guarantee useful alternatives at every decision. These training-split, ground-truth-target records use a historical evaluator and are not pooled with the later exact-box cohorts or learning screen. [Section C.20](#) retains the full attempt counts, continuation support, runtime denominators and family-validity figure.

6.3 What development data completed and qualified?

Two separate 100-intention collections produced 97 complete 16-step chains each, with three typed support exclusions in each roster. Their successful chains cover the same 97 development scenes. Together they provide 194 chains, 3,104 states and 107,577 label-supported candidate rows; these nested rows do not increase the independent-scene denominator. The second draw adds target and state coverage within existing scenes: among 97 successful exact-root matches, 93 selected target identities changed and four were retained.

Development cohort	Intentions	Chains	States	Supported labels
Original collection, terminal 2 October; audited 4 October	100	97	1,552	54,816
Second collection, terminal 4 October	100	97	1,552	52,761
Combined, same 97 scenes	200	194	3,104	107,577
Separate actor-root increment, 4 October	5	5	80	separate audit

Table 11: Completed development cohorts. Each 100-intention roster has three support exclusions; every successful chain has 16 states. “Supported labels” counts stored candidate labels on their declared support, rather than every proposed candidate. The five actor-only roots and their five rollouts qualify separately and are excluded from the combined totals.

Canonical readback passed for all 194 successful stores and their actor, label, selected-depth and successor joins. It materialized all eight root feature blocks and causal selected-depth prefixes, verified mask nesting and finite supported rewards, and found the sole terminal at index 15 of every chain. This qualifies the recorded privileged ground-truth-target input; it does not admit dense-valid supervision or an observed-target deployment input. The separate five actor-only roots retain their own source manifest and five qualified rollouts. Passing their reader checks does not merge their population with the two earlier-source cohorts.³

Successful-worker receipt times sum to 8.92146 hours for the original 97 chains and 3.82197 hours for the second 97 chains. These clocks exclude support exclusions and are sums of worker durations, rather than collection wall time. The acquisitions differ in source revision, seed and operational settings; their totals are not a matched acceleration estimate. The bounded renderer contrast above supplies the matched cost observation. Collection completion and input qualification establish neither corpus admission nor a learned endpoint improvement. Stored endpoint bindings and seals were checked in these collection audits; endpoint statistics were not independently recomputed from raw clouds.

The second cohort also exposes how the available choices change during a successful trajectory. Median hard-valid candidate count falls from 50 at the first acquisition to 22 at the sixteenth, while accumulated selected-query gain rises (Figure 20). Hard validity describes feasible proposals; it does not establish the training-label support examined next. The stored gain is a sum of selected query reductions, not a new independent endpoint audit. Completion therefore coexists with a contracting choice set.

³Primary audits and seals: original collection, second collection, and actor-root increment. The explicit schema correspondence is V10 for the two 97-chain cohorts and V11 for the five actor-only roots.

Candidate feasibility and stored target gain through H16

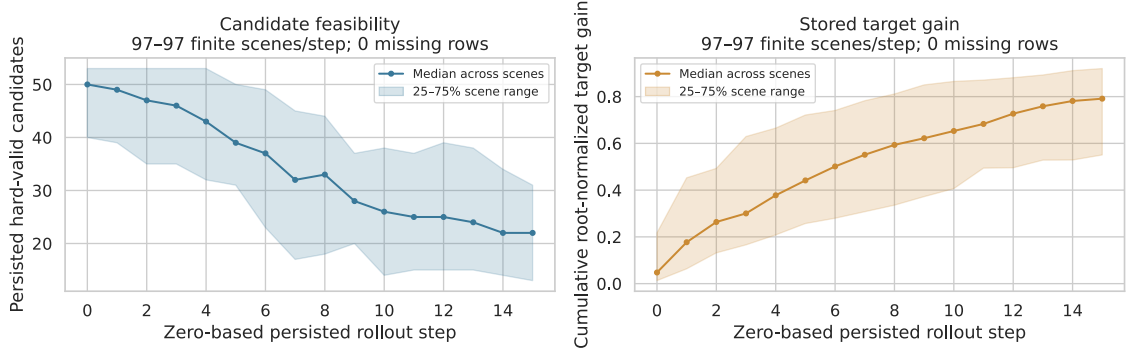


Figure 20: Candidate feasibility and accumulated selected-query gain in the second 97-scene cohort. Lines show scene medians and shading the 25th–75th percentile scene range at each acquisition; the retained plot indexes acquisitions from zero. Hard-valid counts and training-supported counts have different meanings. Gain uses the stored root-error normalization and remains distinct from an independently measured endpoint. Scene ranges describe variation, not confidence intervals.

Selection changes too: orbit proposals account for 21 of the 97 first choices and 55 of the final choices, while bearing choices fall from 20 to 3 (Figure 21). These counts describe the combination of the offered pool, its validity filters and the implemented selection policy. They do not establish that orbit proposals cause greater improvement or that the cameras move consistently clockwise or counter-clockwise around the target. Family identity and actual camera-position coverage require separate checks.

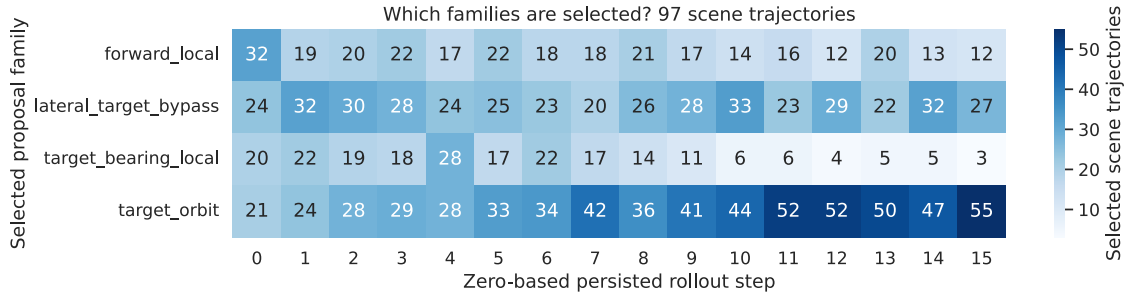


Figure 21: Selected proposal families across the second cohort’s 97 trajectories and 16 acquisitions. Each heatmap column totals 97; the retained plot uses zero-based acquisition indices. The four groups are forward, bearing, bypass and orbit proposals. Increasing orbit frequency is post-selection evidence under unequal proposal quotas, not a causal family comparison or a measurement of orbital coverage.

6.4 What supervision survives across families and acquisitions?

Collection completion supplies recorded trajectories; learning also needs supported alternatives at their decision states. A descriptive re-analysis examines 215 completed chains with 3436 recorded states and 206160 proposed candidate rows. It reuses the **second** 97-chain R1 collection summarized above and adds 117 incremental CUDA chains and one separately labelled Mac diagnostic. These totals overlap [Table 11](#) and cannot be added to its combined count. The original R0 collection is outside this re-analysis.

The primary views separate 77 historical TRAIN chains, 117 incremental TRAIN chains and 20 exposed validation chains. The 194 TRAIN chains cover 77 distinct physical scenes: the increment covers 40 scenes already represented in historical TRAIN. More chains expand recorded target and state experience within existing scenes without a corresponding increase in independent scene diversity. The diagnostic remains in the availability record and is excluded from the primary quantitative comparison.

Scope	Availability	Interpretation
Re-analysis and increment	215 chains; 3436 states; 206160 proposals	Includes the second R1 collection already reported
TRAIN	194 chains; 77 physical scenes	77 historical chains plus 117 incremental chains on 40 overlapping scenes
Exposed validation	20 chains; 20 scenes	Opened after TRAIN nominations were frozen
Mac diagnostic	1 chain	Excluded from primary quantitative figures
Root context	42 attached; 38 held stores	Held context was unlaunched and remains unavailable
Source-bound timing	213 of 215 chains	Different clock/device scopes; no backend speedup estimate

Table 12: Re-analysis of the second R1 cohort plus incremental CUDA chains and one diagnostic. The preceding two-cohort table and this census overlap: the historical chains here are the second R1 collection. Root-context availability has its own denominator and adds no rollout chains.

A candidate can be proposed yet lack a label admitted by the finite-horizon training contract. [Figure 22](#) separates the fraction of allocated candidates with training-label support from the fraction of states whose uniquely best finite actor-action label belongs to each family. The latter asks whether a family offers the sole

best labelled immediate gain. It uses finite actor-action/oracle labels rather than the training-support mask; ties do not identify a unique best. Values are reduced within each physical scene before equal-weight scene summaries are formed, with cohorts kept separate.

Bearing proposals have training-support yields of 31.4%, 37.1% and 37.4% in historical TRAIN, incremental TRAIN and exposed validation, respectively. Corresponding orbit yields are 89.5%, 89.3% and 89.5%. Orbit supplies the uniquely best finite actor-action label at 38%, 33.4% and 36.9% of states, respectively. These observations describe the incumbent quotas and scene exposures. Unequal allocations affect opportunities to attain a maximum, so exclusive-best fractions do not identify the effect of reallocating proposals.

Bearing’s zero-support frequency complements its allocation-normalized yield. A recorded state with no supported bearing candidate contributes a measured zero for that family, while an unavailable state remains outside that calculation. [Table 13](#) retains both counts and scene-equal fractions.

Cohort	Bearing: zero support	Orbit: zero support
Historical TRAIN	584 states; 47.4%	0 states; 0%
Incremental TRAIN	765 states; 40.8%	0 states; 0%
Exposed validation	119 states; 37.2%	0 states; 0%

Table 13: States with no training-supported candidate in the indicated family. Counts sum recorded states, while percentages first reduce within each physical scene and then weight scenes equally. Fractions are descriptive; unavailable support is separate from measured zero.

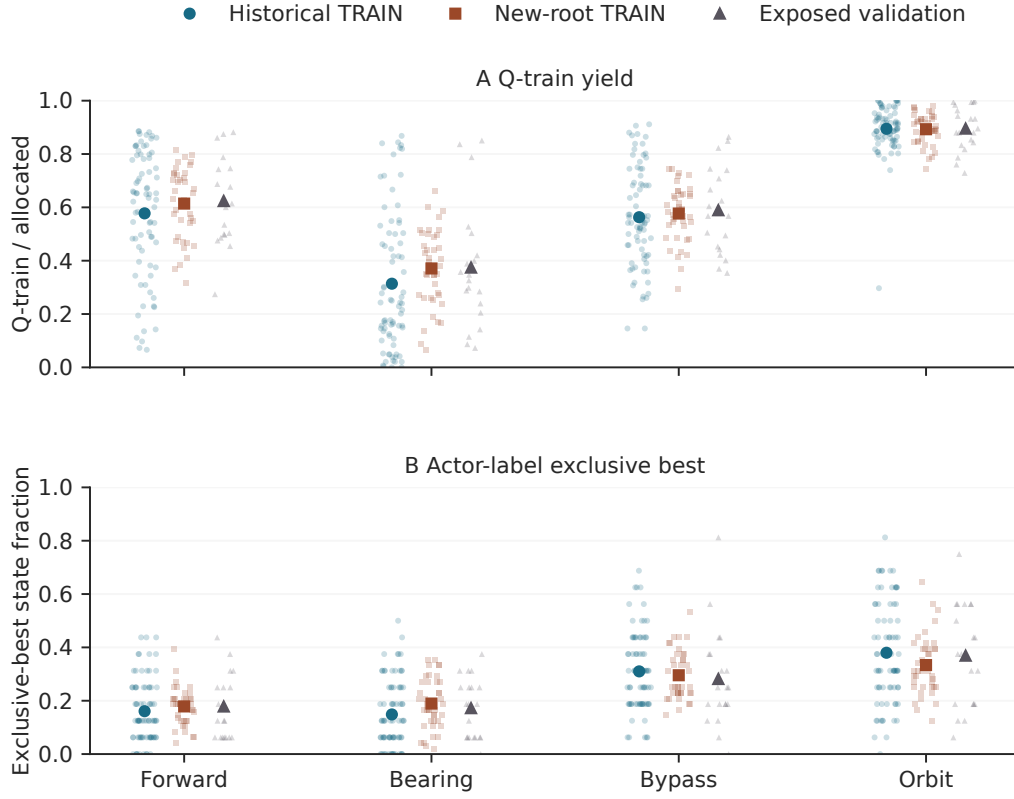


Figure 22: Candidate-family opportunity under the retained proposal recipe. Upper: training-label support divided by allocated candidates. Lower: fraction of recorded states at which a family supplies the uniquely best finite actor-action/oracle gain; ties are excluded from unique-best counts. The latter uses actor-label support rather than the training-support mask. Points show physical-scene summaries, and centers weight scenes equally. Historical TRAIN, incremental CUDA TRAIN and exposed validation remain separate; the Mac diagnostic is excluded. Unequal quotas and overlapping scene exposure preclude a causal family comparison.

The late-acquisition view asks how many supported alternatives remain at recorded states. [Figure 23](#) presents all persisted acquisition positions, with counts and finite label spans reduced within scene and averaged equally across available scenes. From the first to the sixteenth position, mean supported count falls from 44.64 to 24.27 in historical TRAIN, from 47.92 to 24.15 in incremental TRAIN, and from 46.7 to 23.85 in exposed validation. The label-span panel uses the 90th-minus-10th-percentile spread of finite actor-action/oracle gains at each state, requiring at least two finite gains; it is not restricted to training-supported labels. In exposed validation its scene-equal mean declines from 0.059607 to 0.004429.

The historical TRAIN label-span summary at acquisition twelve uses 76 available scenes and 1 missing scene. Other metric-position summaries retain their cohorts' recorded scene coverage. These late-position summaries are conditional on recorded states and finite label support. The incremental cohort includes a completed chain terminated for zero valid candidates after twelve recorded states. At the sixteenth position, 116 of 117 intended incremental chains remain recorded, while their metric summaries still cover 40 physical scenes. Reached and intended prefix counts at acquisitions 1, 2, 3, 5, 8, 12 and 16 expose the available continuation risk sets. Missing continuations remain unavailable rather than contributing zero; other positions have no inferred prefix denominator. The contraction characterizes this recorded experience. It does not establish that recording sixteen acquisitions is unnecessary or that a shorter prediction horizon would preserve performance.

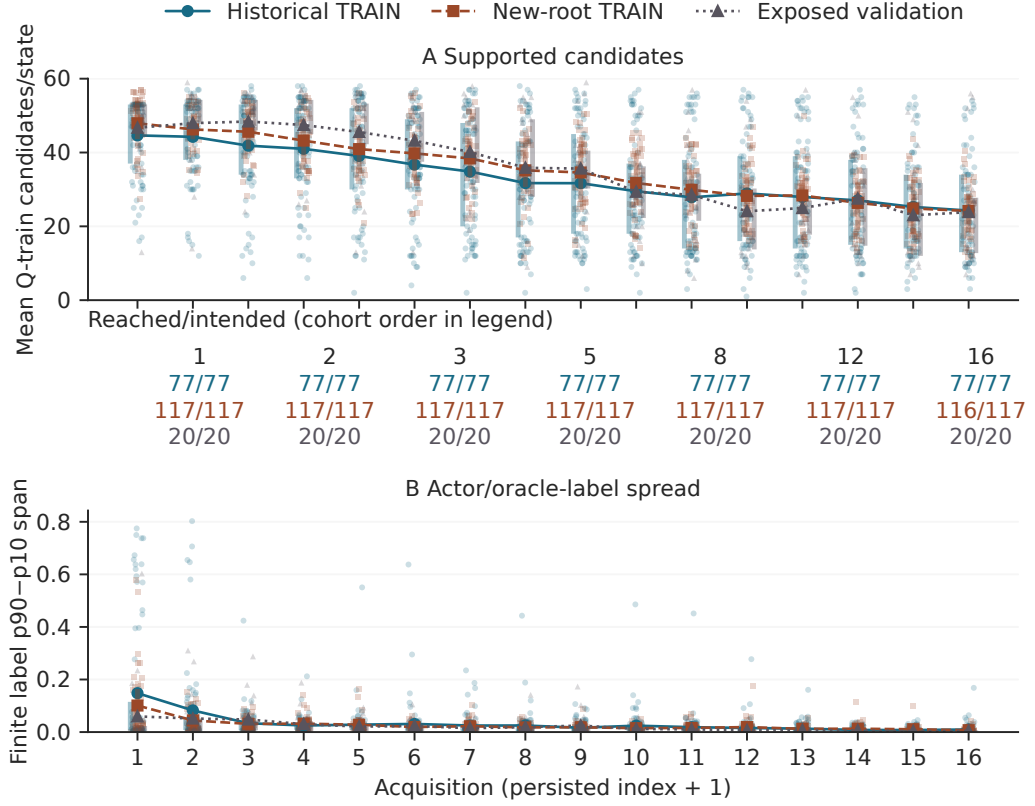


Figure 23: Supervision through recorded acquisition. Upper panel: training-supported candidate count. Lower panel: mean within-state 90th-minus-10th-percentile span of finite actor-action/oracle gains, requiring at least two finite gains; the span is not restricted to training support. Both panels use within-scene reductions and equal-weight available-scene summaries. Cohorts remain separate; metric-specific scene support is stated in the text. Reached/intended continuation counts appear only at the declared acquisitions. Missing continuations and nonfinite spans remain unavailable. Any quartile range describes scene spread rather than population uncertainty. Conditional curves do not establish an optimal recording or prediction horizon.

6.5 Did the nominated collector changes improve supervision?

The two TRAIN-nominated changes received a subsequent matched pilot on three physical scenes, with one root and one proposal/selection seed per scene. All nine Mac CPU cells completed sixteen acquisitions. The quota-only arm changed forward, bearing, lateral and orbit counts from 18/18/12/12 to 18/14/12/16. The separate radius arm retained the original counts and reduced only the bearing

radius from 1.1 m to 0.85 m. Initial root/target assignment, measurement and motion bounds remained fixed; the CUDA arm was prepared but not executed.

The incumbent retained 876 training labels, the quota arm 1,062 and the radius arm 1,095, each from 2,880 allocated candidates across 48 supported states. These are operational supervision counts. Endpoint effects differed by scene, and no learning comparison was performed. More retained labels therefore do not establish better value prediction or reconstruction by a learned policy. The three-scene, single-seed pilot provides descriptive matched evidence without population uncertainty or a selected collector winner.

The frozen pilot protocol and sealed report retain root identities, paired scene reductions, independent endpoint diagnostics and all-attempt costs. These results follow the census nominations; they are not an independent replication of the census or evidence for pooling hosts.

6.6 Where do the selected cameras move?

An orbit-family proposal need not produce a complete orbit, and an angular revisit need not repeat the same visible surface. We therefore inspect physical camera positions separately from proposal names. The second 97-scene cohort provides one complete 16-acquisition chain per scene. Its retained store-content digests were rechecked before extracting the root and selected camera poses. These are the same trajectories counted above, not additional independent data. The geometric definitions and full per-acquisition occupancy atlas are in [Section C.18](#). The atlas separates early concentration near the root bearing from later dispersion across scenes; it does not imply that every chain circles its target.

Each target is the coordinate origin. A fixed horizontal basis is aligned with the initial root-to-target direction, with world up retained; it is not rotated again at each acquisition. Position azimuth is reported relative to the target-to-root ray, so the root is at zero degrees. Positive changes are counterclockwise when viewed from above. The map in [Figure 24](#) uses azimuth and the sine of elevation: equal-sized cells then have equal solid angle. This avoids assigning the same area to

equally wide elevation bands near the equator and poles. Distance to the target and camera gaze are omitted by this projection, so it measures positional occupancy rather than visible target-surface coverage.

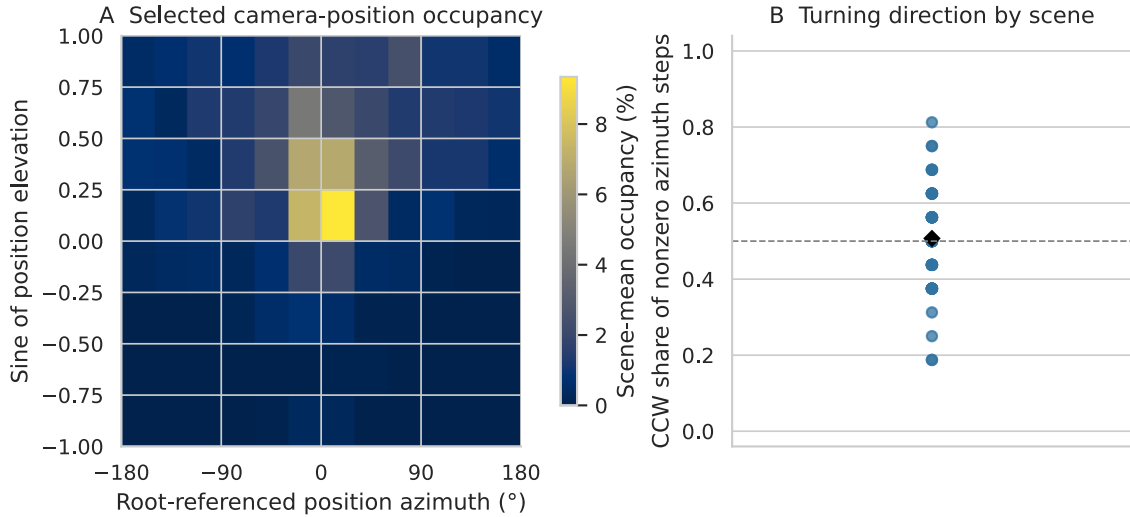


Figure 24: Selected physical-camera positions and turning direction in 97 successful scene trajectories. Left: scene-mean position shares in 12 azimuth by 8 sine-elevation cells; the root anchors azimuth but contributes no occupancy. Each chain is normalized before equal scene weighting. Right: each point is a scene’s fraction of nonzero azimuth changes that are counterclockwise; the diamond is the equal-scene mean. Changes use the shorter signed endpoint angle; degenerate bearings and exact half-turns are unavailable. These are observed motions under one collection policy, not assigned clockwise/counterclockwise treatments or surface coverage.

In this cohort, 66.3% of the scene-mean positional occupancy lies within 60 degrees on either side of the initial root bearing. The mean scene-level counterclockwise share is 0.507. Roughly balanced turning directions therefore coexist with a concentration of camera positions on the initial side of the target; frequent orbit-family choices do not imply uniform spherical coverage.

Direction increments use adjacent acquisitions; their counterclockwise share complements the position histogram but discards the order of those increments. Clockwise and counterclockwise increments can cancel, yielding little net angular progress despite substantial angular travel. A pair of sampled endpoints does not identify an intervening winding or the path taken between them. The collector used temperature-based selection rather than assigned orbit directions; the distribution diagnoses that collector, not the relative reconstruction quality of clockwise and counterclockwise policies.

Turning frequency and angular travel answer different questions. [Figure 25](#) reports both: many small turns can outweigh a few large turns in frequency while contributing less angular travel. The whole-chain comparison retains magnitudes, revealing cancellation that a direction count alone hides.

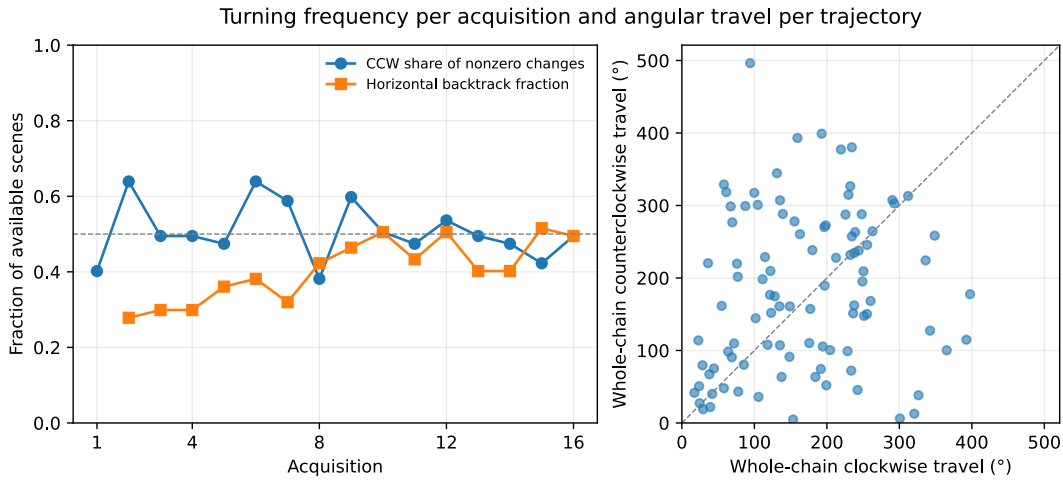


Figure 25: Turning in the same 97 scene trajectories. Left: fractions of available scenes with counterclockwise nonzero azimuth changes and with horizontal bends above 90 degrees, by acquisition. The first bend is undefined. Right: each point compares a chain’s total clockwise and counterclockwise angular travel; the diagonal denotes equal amounts. These are descriptive endpoint increments, not assigned direction treatments or continuous winding.

6.7 How continuous are the selected poses?

Continuity needs more than a plausible-looking path. Translation increments measure how far the physical camera moves between acquisitions, and geodesic rotation increments measure how much its orientation changes. Horizontal bend compares consecutive translation directions; a bend greater than 90 degrees is counted as backtracking. [Figure 46](#) retains these distinct quantities. Zero-length horizontal moves have no defined bend and remain outside its denominator. Rotation is an orientation change, not positional azimuth around the target.

Across scenes, the median within-chain median translation is 0.445 m, camera rotation is 48.9 degrees and horizontal bend is 71.2 degrees. The median backtracking fraction is 0.40. These substantial orientation and translation changes motivate checking the generated motion against a recorded egocentric reference; the spatial summaries alone do not determine whether it is natural.

Figure 26 retains acquisition position rather than collapsing each chain to one summary. The first acquisition involves the largest median reorientation. Median translation decreases later in the chain, while movement bends remain variable: smaller steps do not necessarily mean steadier movement direction.

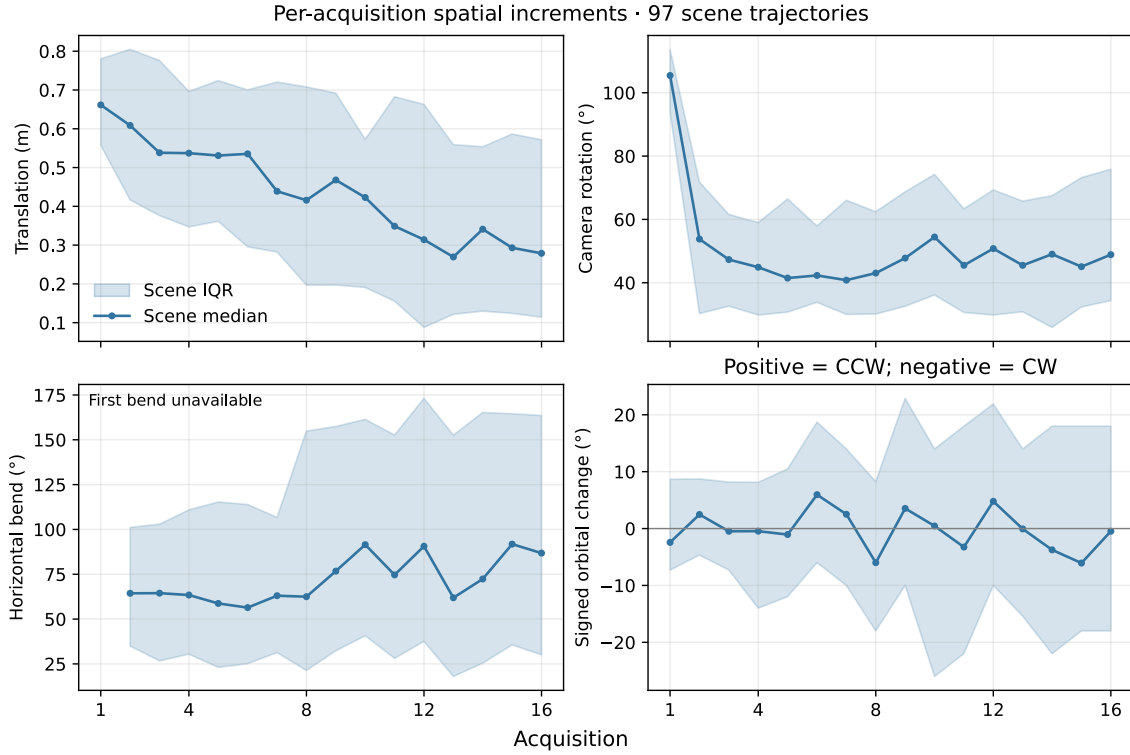


Figure 26: Across-scene medians and interquartile ranges of translation, geodesic camera rotation, horizontal movement bend and signed orbital increment, at each acquisition. There are 97 complete chains; the first bend is unavailable. Shading describes scene dispersion, not uncertainty in the median. Acquisition index carries no timing information.

Whole-chain totals provide a complementary view (Figure 27). The median sampled path length is 7.33 m, compared with a median root-to-final displacement of 2.92 m. Median absolute angular travel is 364.7 degrees, but median signed progress is only 15.3 degrees. The first angular quantity adds turns in both directions; the second allows them to cancel. Neither proves a completed revolution or identifies the physical route between acquisitions.

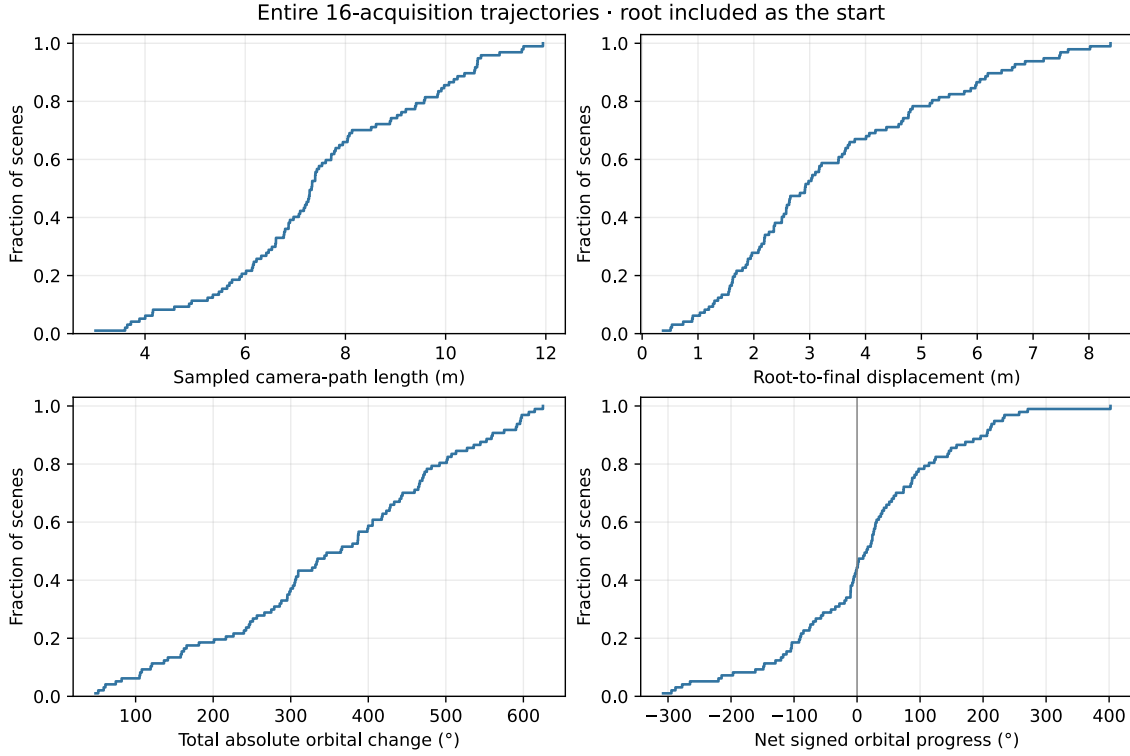


Figure 27: Empirical distributions across the same 97 scenes of sampled polyline length, root-to-final displacement, absolute angular travel and signed angular progress. Each scene contributes one complete chain. Angular totals sum shorter endpoint changes; absolute travel counts both directions, whereas signed progress permits cancellation. Sampled length is not a measured intervening physical route.

Generated acquisitions carry no step timestamps. The reported increments therefore cannot be converted into speed, acceleration or jerk. Nor does a small increment by itself establish human-like motion: that requires a named reference population and comparable timing, sampling and geometric conditions. A geometric path requires time parameterization to define a timed trajectory [41, Sec. 9.1]; movement smoothness also depends on task and measurement conditions [42, Eq. 1; Table 1].

Open decision: Choose a recorded egocentric trajectory reference and its sampling interval before interpreting these spatial diagnostics as evidence of natural motion. Physical jerk additionally requires justified acquisition timestamps or a specified motion model; do not infer either from acquisition index.

Source: Selected-pose continuity readback; canonical rollout steps have no timestamps.

Gate: A matched reference, timing model and statistical unit are specified; sealed trajectories remain unchanged.

6.8 Does the one-step learner clear the frozen practical-effect screen?

A separate development screen compared a root-moment learner with a constant-score baseline in four physical-scene-disjoint folds. It retained 336 paired states from 21 usable scenes under the original privileged label protocol. At each state, practical regret is zero when the gap between the best labelled candidate and the selected candidate is at most the per-state practical-tie tolerance of 0.005; above that boundary it equals the full gap. The reported improvement is baseline practical regret minus learner practical regret, averaged within each scene and then equally over the 21 scenes.

The frozen improvement was approximately +0.00264, below the predeclared +0.005 minimum scene-averaged improvement; 12 of 21 scenes improved. This aggregate effect criterion and the per-state tie tolerance have different roles even though their numerical values coincide. The screen therefore failed its frozen criterion. The constant baseline fits one scalar on training scenes and assigns it to every admitted candidate. Evaluation resolves ties by selecting the lowest admitted candidate index: changing the scalar can change prediction error but cannot change that action. A post-hoc sensitivity that averages regret across all tied admitted candidates gives approximately +0.00350, with 16 of 21 scenes improving. It retains the failed frozen decision.⁴

An exploratory reduction of the same fold rows found scene-equal pairwise ranking accuracy 0.591455, with 19 of 21 scenes above 0.5. Informative candidate pair counts were pooled within each physical scene before averaging the scene accuracies. A descriptive 95% percentile interval [0.551947, 0.629802] used 10,000 physical-scene bootstrap resamples. This is neither a per-state Spearman correlation nor an independent replication. Ranking signal can coexist with insufficient practical improvement at the selected maximum; it changes neither the frozen

⁴Original fold assignments, fit/evaluation receipts and readback: frozen one-step screen.

verdict nor the scope to multi-step or observed-target performance.⁵

6.9 Do action choice and prediction error agree in the larger one-step analysis?

The larger study fits the corrected privileged volumetric scorer on a 48^3 grid, using selected reference-mesh depth, pose-history mean pooling and the original centre-segment attention readout. Its root support excludes image-feature projection counts. This is a one-step-only fit with the auxiliary feasibility loss disabled; it is neither the compact root-moment screen above nor the proposed calibrated multi-ray representation comparison.

This development readback separates action choice from numerical prediction using 20 exposed validation physical scenes, each contributing paired original and cropped-rendering chains. States and acquisition recipes are reduced within scene, and scenes receive equal weight. The sole completed fit uses seed 41. The analysis retains the originally selected and final checkpoints with fitted weights fixed; its 95% intervals bootstrap physical scenes and omit training-seed and checkpoint-selection uncertainty. This supplements the earlier failed practical-effect screen without replacing its frozen verdict. Each checkpoint role evaluates 640 states and 22,753 candidate labels under the original admitted action-label subset. Its regret uses raw immediate gain, rather than the earlier screen’s thresholded practical regret. An exact maximum over this subset does not establish the maximum over hard-valid rows with missing labels.

Lower selected-action regret means choosing a candidate closer to the best labelled immediate gain. Against the TRAIN-fitted constant predictor, regret differences are -0.005685 $[-0.008978, -0.002702]$ for the selected checkpoint and -0.007507 $[-0.012096, -0.003619]$ for the final checkpoint. Brackets give the paired 95% interval. A constant score preserves the original deterministic action tie rule, whereas uniform choice averages over admitted candidates. The constant

⁵Retained exploratory reduction: scene-level ranking evidence. The interval resamples scenes with replacement, seed 2930, using linear 2.5th and 97.5th percentiles; it is descriptive analysis of the existing screen.

has 0.002789 [0.000269, 0.005608] more regret than uniform choice; improvement over the constant alone is therefore an insufficient test of useful ranking.

Against uniform choice, selected-checkpoint regret differs by -0.002896 $[-0.006162, 0.000177]$, whose interval includes zero. The final-checkpoint difference is -0.004718 $[-0.008551, -0.001345]$, giving a conditional descriptive advantage on these exposed scenes. Final-minus-selected regret is -0.001822 $[-0.004514, 0.000788]$, so this analysis does not resolve which checkpoint ranks better.

Mean absolute prediction error gives a different result. Selected-minus-constant error is -0.003503 $[-0.00462, -0.002384]$, whereas final-minus-constant error is 0.004087 $[0.001618, 0.00668]$. The final checkpoint’s action-choice advantage over uniform selection therefore coexists with larger mean absolute prediction error than the constant reference. [Figure 28](#) presents these contrasts with their original interval endpoints and a zero reference. This divergence motivates evaluating action ranking and gain prediction separately. It does not justify changing the frozen checkpoint selector or identify a superior representation or multi-step policy.

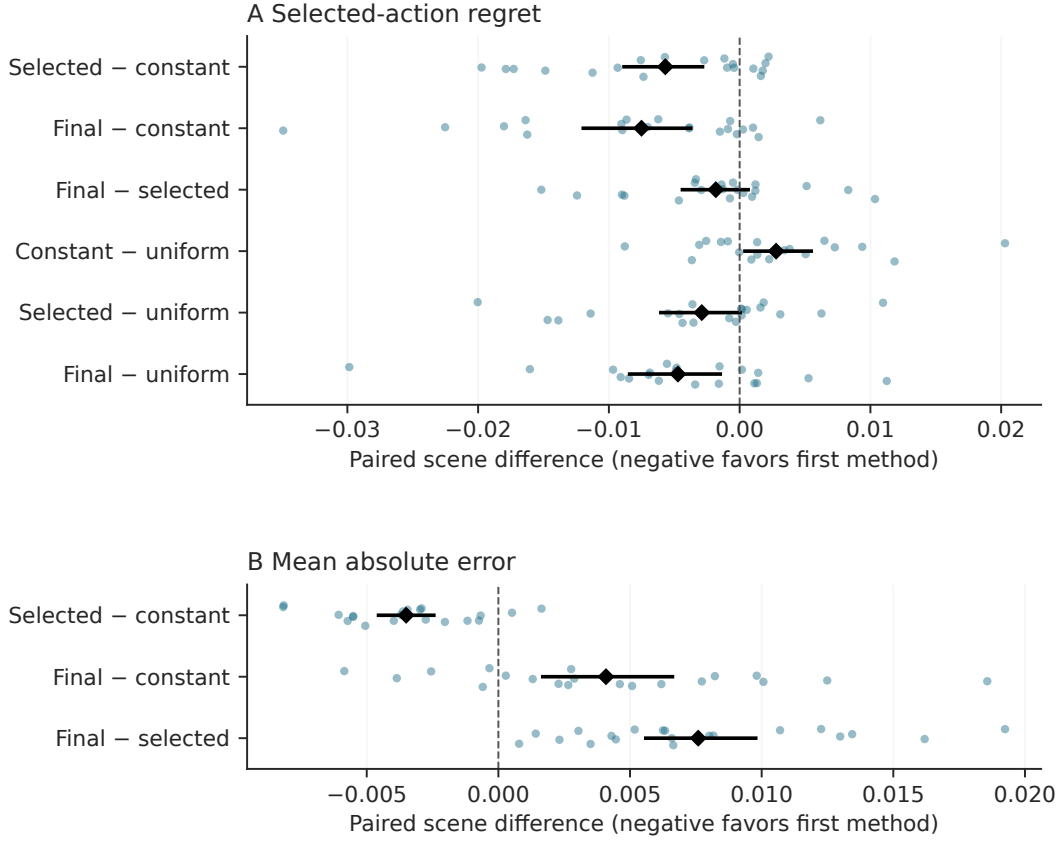


Figure 28: One-step action choice and numerical prediction error on exposed validation physical scenes, with paired original/cropped-rendering chains and one completed fit. Regret and mean absolute error are named paired differences; lower favors the first member. Scene observations and original paired 95% bootstrap intervals retain equal scene weighting after within-scene reduction. Zero denotes equal mean performance. Weights are fixed; intervals omit seed, model-selection and complete post-hoc-search uncertainty. The TRAIN constant uses the original deterministic tie choice; uniform choice averages admitted candidates. Final-versus-uniform regret is negative throughout its interval; selected-versus-uniform includes zero. Improved action choice does not imply accurate gain magnitudes.

6.10 Did additional trajectories improve the matched Mac learner?

A later matched Mac study expanded TRAIN from 77 to 145 R1 trajectories on the same 77 physical scenes. Its baseline and expanded-data fits each completed 3,120 successful updates with seed 41, direct scalar Huber supervision and zero auxiliary-loss weights. The exposed validation population comprised twenty physical scenes, 320 states and 11,444 exactly matched supported candidates. These fits are separate from the earlier one-step analysis above.

Expanded-data minus baseline final-checkpoint regret was -0.001331 $[-0.006233, +0.004134]$. Its 2.9% point reduction missed the frozen 5% practical-improvement gate. Mean absolute error instead increased by 0.006222 $[0.003721, 0.009062]$, exceeding the matched-control calibration cap. Neither final model’s regret-versus-uniform interval excluded zero. The paired 95% intervals resample physical scenes and condition on fixed weights; they do not include training-seed or checkpoint-selection uncertainty.

Initial states contributed 65.6% of the expanded fit’s final squared error and 92.5% of its selected-to-final increase. This localizes drift without identifying representation collapse or its cause. No auxiliary loss was active in either fit, so competition with an active feasibility loss cannot explain this result. A lower-learning-rate arm completed engineering checks only; it supplies no completed scientific learning-rate comparison.

The matched Mac publication retains the selected/final roles, native snapshot, scene contrasts and error localization. Full-TRAIN fixed-checkpoint evaluation remains unavailable, so the observations do not establish a matched generalization gap or a policy benefit.

6.11 What did the continuing-value comparison establish?

The registered continuing-value comparison did not complete. Joint seed 41 stopped at update 2,467 of 15,400 after the nonfinite-gradient guard; the other three planned cells have no launch receipt. The three retained checks at updates 770, 1,540 and 2,310 show fixed-label one-step Huber loss falling and then rising (0.003743, 0.003137, 0.005382). The mixed-horizon fitted-target loss increased (0.004862, 0.006586, 0.009170), but those recursive targets change during training.

Each check contains 640 one-step state contributions and 40 contributions per supported higher horizon. These are correlated state/query aggregates from one changing model, without scene-level continuing-value errors or intervals. They cannot establish worsening policy value. The retained update-770 checkpoint is not a completed matched final comparison; absent failure-time batch and scaler state

also limits diagnosis. This numerical interruption leaves the multi-step learning and endpoint questions open.⁶

6.12 Historical proposal-support and profile diagnostics

The historical no-render audit retained 3146 of 6000 proposed views, but 44 state-family combinations retained none. It failed its proposal-support gate. Because no views were rendered, the audit measures neither reconstruction gain nor policy utility. [Section C.21](#) preserves the counts, exclusions, failed predicates and family-survival heatmap. Later collection does not overturn this earlier profile-specific failure.

A paired radius diagnostic on 6 fixed training roots and 2 seeds removed proposals beyond the 1 m motion bound without eliminating support from any previously supported family at paired root-preccheck and step-zero states. Both profiles completed 10 of 12 intended chains; other attempts failed root admission. Selected label and successor checks agreed, but this is a support/parity result, not a reconstruction-quality benefit or an observed-target test. The aligned setting remains a named diagnostic; the principal 1.1 m radius is unchanged. Exact counts, settings and transition checks are retained in [Section C.19](#).

6.13 Target-Task Coverage

The implemented pipeline samples geometry-valid GT target tasks and records target and validity fields in the rollout schema. The loaded confirmatory bundle does not contain a validated held-out population and exclusion table for the study. Eligible-target exposure, class coverage and exclusion denominators for that population are unavailable. The bounded development scene counts and validity diagnostics reported above do not supply those held-out denominators.

⁶Frozen readout and native report: interrupted continuing-value analysis. The native report's snapshot digest is 00123084744210eae2b6cf95c15a8ee52e266fc94dd4ae3424f10903de53a0b5.

6.14 Candidate-Profile Development Diagnostic

The frozen radius challenger completed fewer intended root-seed tasks than the baseline: 9 versus 11. On 6 shared-complete pairs across 3 development scenes, its scene-mean endpoint gain difference was -0.0098 . This conditional comparison does not cover the incomplete tasks. Insufficient target-current support left admission preservation undefined for two scenes, so the challenger failed the development gate and was not tested on holdout scenes. It does not justify changing the declared 1.1 m profile. The full completion, exclusion and interval records remain in [Section C.19](#); three scenes do not establish population-level advantage.

6.15 Outstanding evidence

6.15.1 Representation and selected-observation comparisons

Validation TODO: Report the matched descriptor, volumetric readout and selected-observation fusion comparisons specified in the experimental design. Separate changes to representation from changes to observation access, and identify privileged selected-depth controls. Collection qualification and the root-moment screen do not supply these comparisons.

Source: PR 422 consolidated report, SR-02; downstream PRs 434 and 436; Representation commissioning before horizon claims; [Section 3.2](#)

Gate: Each reported contrast has matched information, scenes, labels and optimization, with prediction, ranking, regret and resource outcomes.

6.15.2 Independently evaluated finite-horizon predictions

Validation TODO: Report value-prediction error and calibration by requested horizon against independently qualified outcomes. Keep exact two-step backup agreement, fit to bootstrapped targets, selected-action regret and endpoint policy performance distinct. An improved endpoint policy does not establish unbiased value predictions.

Source: [Section 4.9.5](#); Replay Eligibility and Finite-Horizon Learning Gate

Gate: Name the prediction estimand, outcome source, supported rows, independent scenes and uncertainty; preserve unavailable horizons as gaps.

6.15.3 Policy comparison

No endpoint-policy effect is reported because the loaded evidence lacks a validated paired effect, interval or consistent positive scene denominator. This absence does not establish a measured null effect. The one-step development diagnostic above

addresses a separate action-selection question and does not supply the missing held-out policy comparison.

No validated oracle-headroom decision is available in the loaded bundle. Its absence does not prevent reporting an independently qualified policy comparison and is not evidence of zero oracle opportunity.

Open decision: Choose whether to include a matched learned-myopic reference and the optional recovery diagnostic in the confirmatory comparison. Freeze the named policies before inspecting their outcomes; the independent endpoint comparison does not require either diagnostic.

Source: RQ3; [Chapter 5](#)

Gate: The comparison manifest names the reference policies and distinguishes endpoint performance from evidence for a specifically non-myopic mechanism.

6.16 Confirmatory Report Interface

This table reports only the loaded bundle, not the independent development diagnostics discussed in Results. The bundle’s evidence class is *pilot*.

Result	Required evidence	Status
Population and targets	population facts with denominators in every validated store	not available
Candidate-support QC	five descriptive diagnostics with the frozen state–scene macro identities and exact scene denominators	not available
Paired policy effect	paired per-scene effect, interval and exact paired-scene count in every validated store	not available
Optional oracle headroom	a recorded Boolean decision with its own positive scene denominator and declared aggregation	not available
Resource feasibility	wall time, peak GPU memory, and storage in every validated store	not available

Table 14: Evidence availability in the loaded report bundle. A status of “not available” means that no thesis result is inferred from fixture values, train-only attempts, or an incomplete store.

7 Discussion

Remove or rewrite before submission: Complete this interpretation when the confirmatory comparisons are available. The present discussion interprets bounded renderer, runtime, collection and one-step results; the retained development hypotheses are not results.

Source: [Chapter 6](#) and the frozen analysis plan

Gate: Each interpretation identifies a reported observation, a justified scope, and any competing explanation; no missing outcome is treated as zero.

7.1 What the available evidence establishes

Measurement is conditional on the acquisition protocol. The completed renderer contrast supports candidate cropping only where full-scene selected observations, successor histories and endpoint scoring remain equivalent. Its earlier failure shows why faster rendering alone cannot qualify that change. The separate runtime diagnostic found too little preparation work to justify a larger mesh cache. Because the clocks cover different devices and nested operations, it identifies further measurements rather than a universal bottleneck or throughput guarantee.

Generation yield and independent experience differ. The historical collection demonstrates retained continuations under its original evaluator. The newer collections add targets and states within the same development scenes. Neither a longer suffix nor another target draw adds an independent scene, creates dense future labels, or admits a new observation protocol. Successful input joins are necessary for fitting the declared objective, but cannot establish predictive usefulness. Summed costs from collections with different recipes also cannot replace the matched renderer comparison.

The post-hoc census makes this distinction concrete: its incremental TRAIN chains broaden target and state exposure on already represented scenes. [Table 12](#) retains the independent-scene denominator. Candidate support varies by family, and supported choices and finite immediate-gain variation contract at later recorded acquisitions. These are measured properties of the incumbent recipe. Scene geometry, target exposure, unequal quotas and conditional continuation coverage can all contribute; none is isolated as a causal explanation. Repeating

supported choices within familiar scenes may help fitting, but unseen-scene usefulness remains a separate evaluation.

Ranking signal need not improve the selected action enough. The frozen one-step screen missed its practical-effect threshold, while exploratory analysis found a ranking signal. Ranking many pairs correctly can coexist with an error at the maximum that determines the chosen view. The constant reference’s tie rule further affects that choice. Post-hoc tie analysis helps design another screen; it does not overturn the frozen negative result. Neither absent useful signal nor representation loss has been established as its cause.

The larger fixed-seed analysis further separates action choice from numerical prediction error. The final checkpoint has less regret than uniform choice on the exposed scenes, while its mean absolute error exceeds that of the constant predictor. Its regret advantage over the selected checkpoint remains uncertain. A model can preserve a useful ordering while assigning inaccurate gain magnitudes, so neither prediction error nor action regret substitutes for the other. These conditional observations do not identify state aliasing, an architecture limitation or poor collection as the cause, and leave the frozen checkpoint selector unchanged.

Support and completion condition the conclusions. The earlier proposal audit remains a negative result for its own profile, even though later collection succeeded. Radius limiting retained inspected family support without proving wearer-motion realism or policy utility. Likewise, the B0/S endpoint contrast is conditional on shared-complete chains; its small population and undefined admission preservation leave the challenger unqualified. Retaining the baseline is therefore a bounded development decision, not proof of optimality.

The interrupted continuing-value fit supplies a numerical failure, rather than a comparison of successful multi-step estimators. A rising loss against a changing bootstrap target can reflect target movement as well as prediction error; it is not an independent measure of policy value. Repeating a qualified fit with fixed-label checks and independent continuation outcomes is needed before interpreting the learning method, representation or collector as its cause.

The principal held-out multi-step prediction and learned-policy endpoint results remain unavailable. An implemented observed-target route does not supply an evaluated actor-visible population.

7.2 What the unresolved comparisons must distinguish

Measured reconstruction and actor information can evolve differently. After an acquisition, the oracle evaluation cloud includes the selected surface, while an actor restricted to compact root features and pose history may receive no corresponding geometric evidence. Two histories that acquired different surfaces can then appear identical to the scorer. This is a possible source of state aliasing, not an observed explanation of a learned-policy result that has yet to be measured.

An oracle may also choose a continuation using information unavailable to the actor, so a gap to its choice need not be approximation error alone. On a fully labelled successor shell, comparing an actor-selected continuation with the oracle maximum isolates the consequence of the two selection rules under that actor. It does not identify either the best possible actor or an irreducible information gap. This distinction limits what a future finite-horizon result can establish without prescribing a larger architecture in advance. A policy may choose useful actions despite biased value estimates, so endpoint success requires a separate prediction and calibration analysis before it supports a claim about value accuracy.

7.3 Data collection before action-space replacement

Additional interaction is relevant when the available proposals are useful but their consequences are poorly represented in replay. DAgger motivates collecting supervision at learner-visited states [43], and Ball et al. demonstrate retaining offline data during online off-policy learning [44]. Oracle-labelled transition aggregation would use that distributional motivation, not implement the same imitation-learning algorithm or inherit its guarantees. The proposed comparison therefore includes an equal-cost exploratory-collection control on training scenes.

The TRAIN-only analysis froze two collector nominations before validation was exposed. The first is a matched quota test, retaining all four families and

the same proposal budget. Its frozen prescription is: “Test proposal counts 18/14/12/16 against incumbent 18/18/12/12; change only bearing/orbit allocation.” The second is a separate matched bearing-radius test: “Separately test bearing maximum radius $1.1 \rightarrow 0.85$ m with incumbent 18/18/12/12 quotas and unchanged 1.0 m step/0.25 m height safety bounds.” Both nominations subsequently received the bounded three-scene Mac pilot in [Section 6.5](#). More labels were retained, but heterogeneous endpoints and the absence of a learning comparison leave a collector winner unestablished. Further independent scene/seed tests remain prospective. Preserve root, target and assignment matching, complete attempted cost and independent endpoints; testing quota and radius separately distinguishes allocation from geometry. The validation readback selected neither intervention.

The matched Mac learning comparison in [Section 6.10](#) likewise shows why more trajectories on the same scenes cannot be equated with more independent experience. Its negative retention result and localized initial-state error motivate fixed-checkpoint calibration and feature-use controls before another architecture or optimizer claim. No active auxiliary-loss competition was needed for drift to occur in that comparison.

Late support contraction does not identify an optimal recording length. Selecting fewer training records can preserve the actual observations that produced each retained state. Skipping acquisitions changes those observations and may change subsequent actions. Such an altered history requires rebuilding the observed or fused state and recomputing affected rewards, targets, validity and returns. Joining factual transitions must retain elapsed-acquisition and horizon semantics together with intervening returns. The learning benefits of record subsampling and altered acquisition histories remain separate hypotheses.

Changing the action space addresses a different limitation. GenNBV’s continuous camera action and sequential coverage objective provide a precedent [\[7\]](#), not evidence that continuous control will improve this thesis’s target-specific metric. Such a comparison needs its own feasibility, observation, simulator, and

cost assumptions. New experience, updated model weights, causal scene updating, and continuous actions are distinct interventions. The optional fusion comparison associated with [Section 1.3.2](#) keeps model weights fixed.

Changing the simulator or observation model would test a broader claim than generalization to new ASE scenes. It can change candidate feasibility, calibration, depth noise and reconstruction behaviour together. Those changes need separate specification and comparison before transferring conclusions about policy utility or numerical value accuracy.

The next evidential step is therefore narrower than an architecture change: qualify the intended learning objective on the retained labels, evaluate predictions against independent references, and complete matched policy comparisons. The renderer endpoint validation already closes its bounded acquisition contrast; the collection audits verify stored endpoint bindings rather than independently recomputing their endpoint statistics. A supported negative result can answer the bounded planning question. Missing support, unreliable measurement, and unfinished runs cannot.

8 Conclusion

Validation TODO: Update these bounded answers when confirmatory evidence becomes available. The current chapter distinguishes defined methods, bounded generation and learning diagnostics, and unanswered empirical questions; it is not a final policy verdict.

Source: [Chapter 6](#); [Chapter 7](#)

Gate: One evidence-calibrated answer per principal research question, consistent with the reported population, comparisons and uncertainty.

This thesis studies learning to score candidate views by their expected contribution to a requested object’s reconstruction over several acquisitions. The decision problem is to choose views from a finite candidate set using the observations already acquired and the remaining acquisition budget. The formulation separates four quantities that must not be conflated: the measured target outcome, the actions available to the policy, the factual experience used for learning, and the information available at decision time.

For RQ1, the thesis defines target-specific measurement and a generation process that records immediate labels and selected continuations. Bounded renderer comparisons establish transition and endpoint agreement under the tested protocol, while historical and newer collections establish recorded coverage with explicit support failures. These findings qualify particular generation procedures; they do not establish unrestricted measurement validity, dense future supervision or a representative physical acquisition process. Repeated targets and states within the same scenes broaden coverage without increasing scene independence.

For RQ2, Method describes compact and spatial representations and the incorporation of selected observations. Their effect on prediction requires matched representation comparisons; implementation alone does not establish a benefit. Observed-target admission and the actor–oracle boundary are explicit, but the end-to-end actor-visible protocol has not yet been evaluated. Deriving the target descriptor from observations does not ensure that action feasibility, scene evidence and continuation selection obey the same information boundary.

For RQ3, the evidence reported here does not yet establish independently validated multi-step predictions or learned-policy endpoint improvements. The

one-step screen missed its practical-effect threshold on 21 development scenes. Exploratory ranking evidence narrows the next question but leaves that negative decision intact. The larger privileged volume shows that action-choice and numerical-error comparisons can diverge on exposed scenes; the interrupted continuing-value fit supplies no completed multi-step comparison. Neither result settles multi-step learnability. Agreement with fitted targets is not an independent value check. The paired endpoint comparisons are unavailable, so the absence of a result is not a measured null. Optional oracle comparisons would assess the learned policy against those bounded references, not against an unknown global optimum.

The central lesson is diagnostic rather than comparative: a missing useful view, an unobserved continuation, and an actor state that omits relevant evidence are different reasons why useful view sequences may not be learned, and they require different diagnostics. Independent value checks and reliable, matched endpoint comparisons on a declared task and action population would test prediction and decision benefit separately. A learned-myopic control and oracle lookahead can help explain a learned result, but neither that control nor positive oracle headroom is required to measure the learned policy’s outcome. Attributing an endpoint benefit to lookahead would require a matched mechanism comparison. The present contribution combines a defined method and experimental design with bounded evidence about generation and learning; it does not establish learned-policy superiority.

Reader-state ledger

This development-only ledger records the intended learning journey of the thesis. It is an editorial contract, not a generated summary and not an owner of scientific claims. Chapter prose, equations, figures, and evidence remain with their active sources. The ledger makes their pedagogical dependency explicit so that an authoring agent can preserve what the reader knows, what the chapter must teach, and what the next chapter may safely assume.

Chapter 1: Introduction docs/typst/thesis/sections/01-introduction.typ

Reader question

Can a finite-horizon scorer predict a candidate pose's target reconstruction improvement from acquired evidence, and what data and evaluation make that claim meaningful?

Enters knowing

- Images and geometric measurements provide partial information about a three-dimensional scene.
- A learned model can predict an outcome from available inputs; the thesis will specify the acquisition problem.

Teaching pointer

docs/typst/thesis/sections/01-introduction.typ: known proposed pose, unseen observation and constrained preparatory move

Leaves knowing

- The scientific target is the geometry of a requested object under a fixed view budget.
- RQ1 concerns how measurement, candidate generation and collection policy affect the reliability and coverage of training experience; RQ2 compares representation, fusion and architecture effects on prediction and candidate ranking, separating added information from its encoding; RQ3 compares numerical prediction with reconstruction achieved by learned view selection on unseen scenes. Independent value checks and acquisition outcomes answer different questions; neither alone establishes absence of pipeline bias.
- Historical and simulated acquisitions can provide different observations; a stopped recording leaves later outcomes unknown. Generalization concerns unseen scenes under fixed generation, observation and measurement conditions, allowing the declared policy comparison, and synthetic success does not establish physical transfer or unrestricted optimality.

Chapter 2: Foundations and Related Work

<docs/typst/thesis/sections/02-foundations/index.typ>

Reader question

What objective, action, information-state, temporal and geometric structure must a target-conditioned finite-horizon view scorer preserve?

Enters knowing

- The thesis asks whether a finite-horizon scorer can learn target reconstruction improvement from acquired observations, with measurement, experience and representation as connected research questions.
- The reader distinguishes target-specific reconstruction from scene-wide coverage and knows that numerical value prediction and useful view selection require different evidence.

Teaching pointer

<docs/typst/thesis/sections/02-foundations/02-03-candidate-support-and-motion-feasibility.typ>: semidense evidence illustration; 02-04 separates return and continuation

Leaves knowing

- View value depends on target utility, causal history, offered actions and continuation; actor-optimal value, privileged optimal-oracle value and an implemented policy's sampled return are distinct. A finite prediction horizon, real task budget and recorded length differ; a cutoff leaves unknown rewards even though a completed finite query has zero remaining return.
- Passive coordinate changes and candidate permutations require compatible predictions, while task-relevant abstraction must preserve decision-changing distinctions; recurrence or symmetry alone does not guarantee predictive sufficiency.
- Observed surface, ray-supported free space, unknown space and learned priors have distinct meanings; missing acquisition, finite extent, discretization and pooling cause different losses, which constrain learnable value and valid data generation. Feasibility and held-out synthetic performance do not establish wearer-motion realism or real-world transfer.

Chapter 3: Oracle and Data Generation

<docs/typst/thesis/sections/03-oracle-and-data-generation/index.typ>

Reader question

How are reconstruction improvement, candidate support and generated experience measured without leaking hypothetical outcomes into the actor's causal state?

Enters knowing

- Candidate value depends on target utility, feasibility, causal state, and horizon.
- The actor and the offline oracle have different information roles.

Teaching pointer

<docs/typst/thesis/sections/03-oracle-and-data-generation/index.typ>: complete acquisition walkthrough and measurement-to-collection chain; appendices own sampler probabilities, ordered preparation and diagnostic formulas

Leaves knowing

- Historical observations, selected counterfactual observations, and current candidate outcomes differ by acquisition provenance; actor and oracle access form a separate axis, finite S0/S1/S2 representations are modality profiles rather than distinct physical state spaces, and only the selected observation may extend the next actor state.
- Target sampling controls exposure, proposals define available interventions, and collection determines which continuations receive factual supervision beyond dense immediate labels; operational point-mesh error separates reconstruction-to-surface accuracy from equal-face completeness and requires a fixed mesh representation and scoring protocol; retessellation can lower the metric without improving the reconstruction.
- Source-dependent modalities and target descriptors offer complementary information; Method selects their encoding. Signed fixed-root rewards telescope only for complete, consistently measured chains. Local RRI, collector returns, fitted continuation and policy-relative endpoint reduction answer different questions; missing continuation is unknown.

Chapter 4: Method docs/typst/thesis/sections/04-method/index.typ

Reader question

How should scene evidence, target identity, action history, candidate geometry, and horizon interact to predict finite-horizon target value?

Enters knowing

- The actor-visible state and oracle supervision have already been separated.
- The reader understands why geometry, time, hierarchy, and finite candidate support matter.

Teaching pointer

docs/typst/thesis/sections/04-method/index.typ: contiguous evidence-to-score path; replay and value learning follow representation controls

Leaves knowing

- Scene, target, history, candidate, budget and horizon have distinct roles. Dense immediate labels and selected recursive targets have different support; online selection with delayed evaluation does not create missing transitions or establish agreement with independent actor outcomes.
- Candidate-relative geometry, shared-state conditioning, causal temporal summaries, and permutation-consistent candidate scoring express useful structural priors.
- A volumetric representation can combine historical features with newly acquired geometry in a common grid; compact representations provide controls. The larger Q1 fit uses corrected measured-root support with privileged selected depth, separately from the proposed calibrated readout. Which feature sources and spatial processing improve prediction remains a matched empirical comparison.

Chapter 5: Experimental Design docs/typst/thesis/sections/05-experimental-design/index.typ

Reader question

What independent evidence assesses value prediction, fair representation comparisons and acquisition outcomes, and which diagnostics are optional?

Enters knowing

- The selected method, oracle, actor-visible state, and candidate support are conceptually defined.
- The reader can distinguish immediate counterfactual labels from selected factual transitions.

Teaching pointer

RQ-to-evidence comparisons and independent backup checks; replay appendix owns batching and validation procedure

Leaves knowing

- Independent evidence is counted at the physical-scene level even when stride changes yield more factual windows or transitions; distinct store-manifest provenance groups do not establish distinct physical scenes.
- Matched populations, nested proposal pools and independent continuation or endpoint evidence distinguish exposure, opportunity and experience limitations; optional collector interventions respond to diagnosed gaps. Repeated fixed-policy returns estimate a policy expectation under declared sampling assumptions; dependent branches and beam maxima do not certify an optimal oracle value.
- Actor modalities, oracle-only assets, factual-counterfactual asymmetry, and simulator realism must be controlled before policy differences are interpreted; oracle headroom and legacy recovery are optional diagnostics, while relative endpoint-error reduction uses a named reference policy and requires a positive reference error.

Chapter 6: Results docs/typst/thesis/sections/06-results.typ

Reader question

What does the evidence establish about pipeline measurement and experience, representation and scorer design, and multi-step value learning?

Enters knowing

- The reader knows the estimands, comparisons, population, and validity conditions.
- The reader does not need to know report keys or artifact schema names.

Teaching pointer

Current measurement, generation and learning lessons, including the matched three-scene collector pilot and Mac trajectory-expansion comparison; historical collection tables are retained in the reproducibility appendix

Leaves knowing

- What measurement reliability and candidate or continuation coverage are established for the evaluated pipeline and population. Per-acquisition occupancy and pose increments explain progression; whole-chain totals distinguish angular travel from signed progress. Selected camera-position occupancy, orbital direction and discrete pose continuity describe the recorded collector; they do not establish visible-surface coverage, timed jerk or human-motion realism.
- What matched evidence establishes about representations and incorporation of acquired observations, separately from additional information access.
- One-step action choice can improve while numerical prediction error worsens; the matched Mac expansion misses its retention gates, and the three-scene collector pilot does not establish learning benefit. The interrupted continuing-value fit supplies no completed multi-step or policy outcome. Unavailable evidence remains distinct from a measured null.

Chapter 7: Discussion docs/typst/thesis/sections/07-discussion.typ

Reader question

What do the current generation and one-step diagnostics explain, which value-learning claims remain untested, and what change is justified next?

Enters knowing

- The principal empirical answers and unresolved outcomes are explicit.
- Diagnostics have been separated from speculative architecture explanations.

Leaves knowing

- Which conclusions are specific to the evaluated scenes, targets, candidates, horizons, and modalities.
- Which mechanisms are supported, which remain plausible alternatives, and what experiment would discriminate them.
- Why selecting factual training records differs from omitting acquisitions, and why state updating, additional collection, parameter learning, and continuous action generation address different diagnosed limitations; changing the simulator or observation model tests a separate transfer claim.

Reader question

What has the thesis taught about target-conditioned finite-horizon view planning, and what remains unresolved?

Enters knowing

- The reader knows the results, their mechanisms, and their limitations.
- No new protocol, architecture, symbol, or evidence source should be required.

Leaves knowing

- What the thesis establishes about reliable measurement and training experience in the evaluated synthetic acquisition setting.
- What the evidence establishes about scene representations, observation fusion and learning multi-step candidate values.
- Which geometric, temporal, hierarchical, representation, and data lessons transfer beyond the exact implementation.

Maintenance contract

`reader-state.toml` is authored rather than inferred. Update the affected chapter record in the same change whenever its central reader question, prerequisites, durable takeaways, or teaching pointer changes. Copy editing, citation repair, and layout-only changes do not require a ledger update unless they alter that learning journey. Development compilation validates the schema and renders these cards; submission mode omits them completely.

Development roadmap

This page is the compact strategic view of thesis development. It is omitted from submission output. Scientific claims remain owned by the active thesis; executable behavior and measurements remain owned by source, tests, configuration, and immutable evidence artifacts.

Author review and professor feedback

Author commitments agreed on 16 October 2026:

Target	Required outcome
26 Oct 2026	Aim to deliver Chapters 1-3 progressively by 26 October; the minimum packet is a newly human-reviewed Introduction and complete Foundations chapter.
30 Nov 2026	All manuscript content, including figures, captions, appendices and declarations, has been human reviewed and revised; the exact candidate is suitable for submission. Existing experimental and release gates still require their own evidence.
12 Dec 2026	Request final professor feedback and review by 12 December; this is the author's planning target, not a confirmed supervisor commitment.

Table 15: Human review and feedback targets; dates do not certify completed review or empirical evidence.

The experimental schedule below was last reviewed on 2 October and is not reval- idated by this author-review plan. Data-generation and training-run scheduling is deferred until LRZ throughput is established.

Implementation: partialEvidence: pending

Reviewed 2 Oct 2026 refresh due 16 Oct 2026. Target: reproducible full-draft freeze by 30 November 2026, followed by a safety buffer until submission on 7 January 2027.

Promotion gate: Close P1 before confirmatory policy claims.

Development source: docs/typst/thesis/development/roadmap.toml; active thesis, package, test, configuration, and evidence owners

TL;DR

Done

Pilot infrastructure

Candidate-benchmark smoke evidence, a bounded two-scene orbit MVP, and a one-scene, five-step pilot rollout report are committed. The October review retains P0 as complete for pilot infrastructure only; these artifacts do not establish population-scale support or policy gains.

Now

Evidence substrate

Freeze the multi-scene store, scene-level split, camera/depth and target-RRI fidelity, candidate-support admission, and throughput receipts while the thesis narrative and conceptual figures remain under review. The separate research publication PR #524 records a larger development collection; completion reconciliation, portable evidence verification and deep cohort audit remain pending before explicit admission. P1 remains in progress; subsequent milestone dates are planning targets conditional on its closure.

Blocked

Confirmatory claims

Lookahead-headroom, learned-value, and population-coverage claims remain blocked until paired multi-scene evidence and an immutable experiment registry exist. The results chapter still gates confirmatory claims, and its default report is a development fixture; neither that fixture nor the separate pilot rollout report discharges P1-P3.

Next

Paired evidence to freeze

Run equal-budget baselines, train and evaluate the finite-horizon value model under actor-visible inputs, synthesize results and limitations, and reach the reproducible full-draft freeze recorded below.

Critical path

The strategic dependency is evidence-first: close the substrate, establish paired baseline outcomes, evaluate the learned value model, synthesize the claims, and freeze the full draft. The December-to-January interval is a safety buffer, not planned scientific work. This is the scheduled order, not a scientific prerequisite chain: held-out Q1 prediction and matched representation comparisons can proceed without paired policy endpoints. Endpoint evidence is required for a policy-improvement claim; oracle headroom remains a separate optional diagnostic.

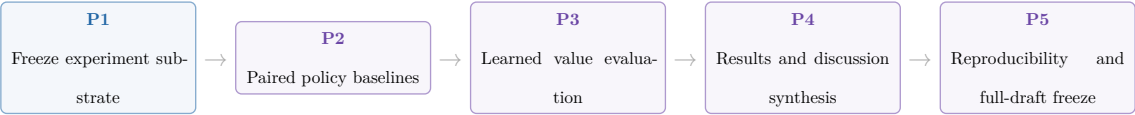


Figure 29: Scheduled development gates through the 30 November 2026 full-draft freeze. Arrows show planned order; policy endpoints and oracle headroom are not prerequisites for Q1 prediction.

Milestones

ID	Phase	Dates	State	Exit gate
P0	Pilot infrastructure	7 Jul–30 Aug 2026	done	Committed smoke and MVP artifacts compile and remain explicitly bounded as pilot evidence.
P1	Freeze experiment substrate	31 Aug–2 Oct 2026	in progress	Multi-scene immutable store, scene-level split, camera/depth and target-RRI fidelity, candidate-support admission, and throughput receipts.
P2	Paired policy baselines	3 Oct–20 Oct 2026	planned	Equal-budget random-valid, greedy, and bounded-lookahead comparisons report uncertainty, support, and oracle receipts.
P3	Learned value evaluation	21 Oct–10 Nov 2026	planned	Held-out ranking, calibration, and paired policy value are reported under actor-visible inputs.
P4	Results and discussion synthesis	11 Nov–22 Nov 2026	planned	Final tables, failure cases, coverage limits, and claim dispositions resolve to the immutable evidence registry.
P5	Reproducibility and full-draft freeze	23 Nov–30 Nov 2026	planned	Full draft, reproducible configurations, final figures, and release checks pass with no unresolved placeholders.
P6	Safety buffer and submission	1 Dec–7 Jan 2027	safety buffer	Advisor feedback, formal checks, final upload, and the authenticated submission receipt are closed without reopening the scientific core.

Table 16: Evidence-gated schedule from the completed pilot infrastructure to submission. Dependencies are checked against the adjacent roadmap data.

Blockers

- **B1 — Multi-scene experiment substrate** (affects P1): The committed artifacts do not yet establish a fresh multi-scene store, scene-level split, camera/depth fidelity, target-RRI robustness, and representative throughput in one immutable receipt chain.
- **B2 — Candidate-support scale admission** (affects P1, P2): The positive two-scene orbit MVP is explicitly not scale-up admission; family survival, framing, diversity, and proposal regret still require broader evidence.
- **B3 — Confirmatory experiment registry** (affects P2, P3, P4): The default report bundle is a development fixture; the separate pilot report covers one scene. Neither establishes confirmatory comparisons, which need immutable

run/config identities, coverage accounting, uncertainty, and claim-to-artifact resolution.

The primary evidence pointers remain beside each record in `roadmap.toml`; the roadmap contract checks their existence, while scientific review decides whether they are sufficient for promotion. Bounded null and negative findings are reportable under the same matched-evidence conditions as positive findings; positive oracle headroom is not a prerequisite for learned-policy comparison.

Promotion queue

Promotion queue — candidate: Promote bounded-lookahead and learned-value results only after paired held-out evidence is immutable.

Source: P2-P3 experiment registry and rollout reports target: Results and Discussion gate: matched held-out comparisons, actor-visible evaluation, uncertainty, and support coverage; oracle headroom is a separate diagnostic and is not required to report a null, negative, or positive policy difference

Promotion queue — blocked: Resolve candidate-support and substrate limits before claiming population coverage.

Source: P1 evidence receipts and candidate benchmark manifests target: Experimental Design and Results gate: multi-scene support, family survival, scene-level split, and representative throughput

Promotion queue — deferred: Keep online discrete and continuous target-then-pose extensions outside the core until the offline finite-candidate gates close.

Source: Discussion and roadmap P1-P3 target: Discussion gate: stable offline learned-value evidence and an explicit post-core scope decision

Maintenance contract

`roadmap.toml` is the single owner for the snapshot, dates, states, dependencies, blockers, release checks, and evidence pointers rendered above. Update it at least every 14 days and whenever a gate changes state. `make thesis-roadmap-contract` rejects stale review dates, broken dependencies, missing local evidence pointers, divergence from the thesis submission date, or references to the retired M1 snapshot. Internal trackers and hosted issue state may inform a refresh but are not imported as public thesis truth.

The native table, cards, and critical-path strip are deliberately sufficient: they show exact milestones and the consequential dependency chain without a second diagram data model. A date-dense Gantt can be reconsidered only if this compact projection no longer answers the planning question.

HM/FK07 release gate

These checks are human-owned and apply to the exact final candidate and the author’s authenticated records; the repository cannot satisfy them by itself.

- **R1 — Applicable rules:** Confirm the individually assigned programme and SPO version in PRIMUSS; public reading versions establish only the current baseline.
- **R2 — Registration record:** Retain the PRIMUSS confirmation, official start date, individual deadline, assigned examiners, supervisor-specific form requirements, and requested attachments.
- **R3 — Exact upload set:** Before the individual deadline, verify the exact final PDF and every PRIMUSS-marked mandatory document or attachment. A repository build, digest, or upload alone is not submission evidence.
- **R4 — Conditional paper copy:** Provide an identical paper copy only if the examiner or authenticated PRIMUSS record requires one under the current FK07 process.
- **R5 — Final-PDF review:** Check the declaration and AI/tool disclosure in the exact PDF prepared for upload; confirm any mandatory signature or attachment against the authenticated record.
- **R6 — Submission closure:** Finalize with the PRIMUSS action “Abschlussarbeit abgeben”, retain the authenticated timestamp, and archive the exact submitted PDF; only that receipt closes the gate.

Official baseline checked 31 August 2026: HM examination regulations, FK07 thesis process, PRIMUSS registration guide, PRIMUSS submission guide. At the full-draft freeze, recheck the official versions and the individually assigned instructions in authenticated PRIMUSS records.

Freeze and submission

By 30 November 2026, the full draft, experiment registry, configurations, figures, and release checks must be reproducible and free of unresolved placeholders. The interval from 1 December to 7 January 2027 is reserved for advisor feedback, formal HM/FK07 checks, final upload, and recovery from unexpected defects. Only the authenticated submission receipt closes the final gate.

9 Development Pilots

9.1 Candidate-Family Support

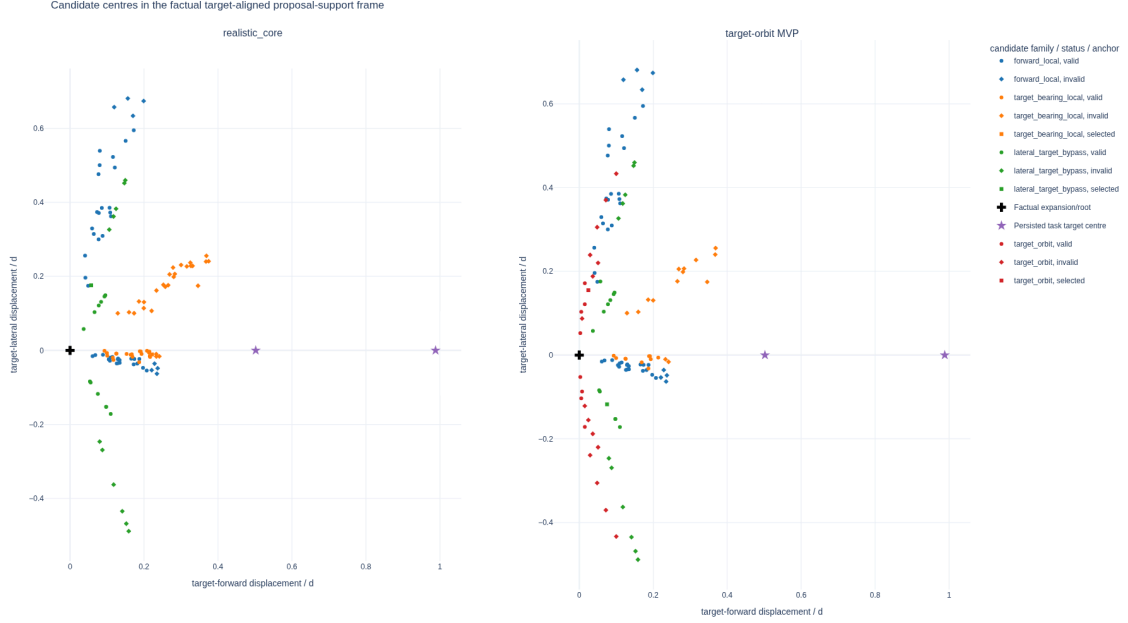


Figure 30: Attempted candidate centres from a two-scene, two-state real-data method audit. Each state is centred on its factual expansion pose, yaw-aligned to the oracle-evaluation target with world Z-up, and divided by its current three-dimensional target distance. The cross marks the expansion/root pose, the stars mark the two oracle task target centres, colour denotes proposal family, and marker shape denotes actor-valid or selected status. The figure diagnoses proposal geometry under the `v0_gt_input` target protocol; it is neither a population estimate nor evidence of actor-visible target discovery or policy quality.

The target-orbit challenger adds attempted support on both target-relative sides in this bounded pilot, while pruning still leaves invalid rows and one configured family/state pair without actor-valid support. The associated immutable evidence manifest records the common-frame reducer, store-manifest hashes, expected state index, committed reduced candidate rows, resolved challenger profile, reproduction commands, metric denominators, and plot hashes. No frame-mixed side-balance or orbit-span values are used. The interactive evidence and Streamlit view can optionally add short arrows for the valid candidates' camera optical axes; the static thesis figure omits them to preserve legibility.

9.2 Target-Frame $\mathcal{S}^2$ Diagnostics

The following frozen figures are generated by the same Python reporting and plotting owners as the stored-rollout application. For the selected real-data shard, the report records these support counts: source samples 1, snippets 1, scenes 1, targets 1, retained rollout chains 1, and selected steps 5. It remains pilot evidence: its purpose is to validate the representation and reporting contract, not to estimate policy quality. The two point-direction views contain 5 factual displacements and 5 factual optical axes. Colour denotes rollout-chain identity, marker shape denotes decision-step identity, and the sphere heat map is computed from the complete equal-solid-angle count grid rather than from the bounded incidence overlay.

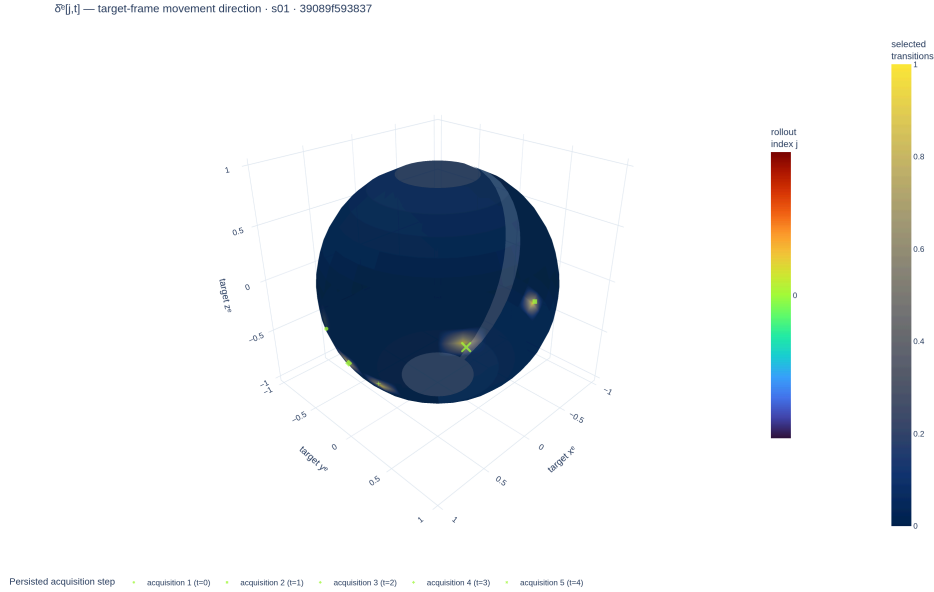


Figure 31: Target-frame $\mathcal{S}^2$ movement directions from the frozen pilot report. This view records how the camera moved; it is not a visibility measurement.

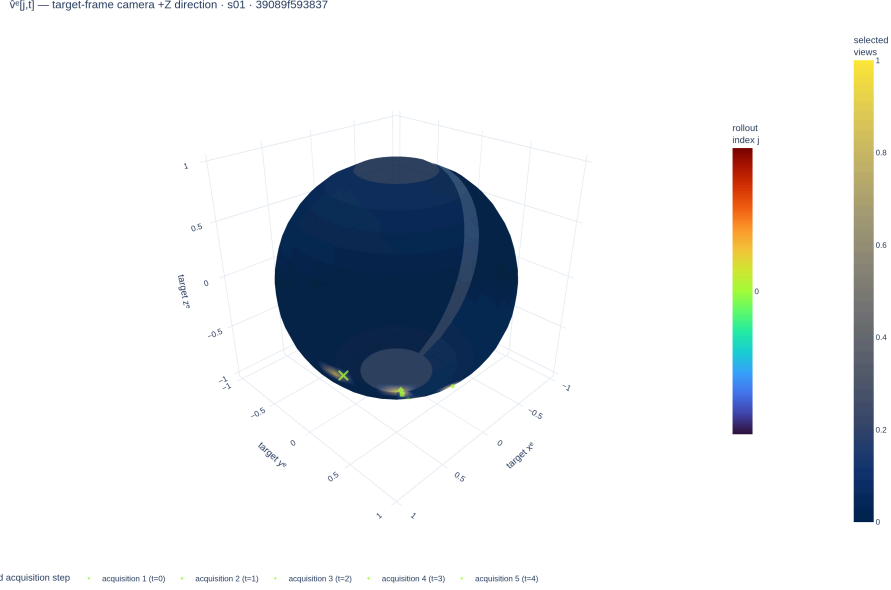


Figure 32: Target-frame $\mathcal{S}^2$ selected-camera optical-axis directions from the same frozen pilot report. This view records where the camera pointed; it is distinct from both camera motion and calibrated surface support.

The calibrated proxy diagnostic integrates 5 selected frusta. Their mean intrinsic field-of-view solid angle is 2.49 sr. On the geometric-mean-scale proxy sphere, one view covers a mean fraction of 0.168, while their union covers 0.312. These fractions express geometric potential support under the proxy and front-facing tests they exclude true target shape, depth ordering, and scene occlusion.

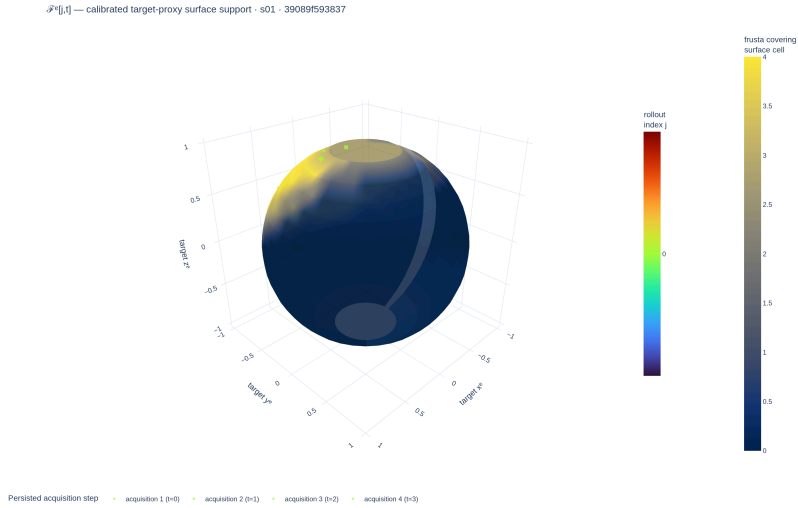


Figure 33: Calibrated selected-frustum support on the target-centred proxy sphere. The surface heat map is the complete equal-solid-angle coverage field; overlaid incidence samples preserve rollout and time heritage. This is not measured target-mesh visibility.

Development Register: Method Alternatives

This register preserves executable but unpromoted controls and future hypotheses outside the thesis-core Method chapter. Entry here means that an idea has a named comparison and promotion gate; it does not mean that the thesis evaluates or recommends it.

Implementation: exploratory Evidence: pending

The thesis-core selection remains A1-H0- $s_t^{S0\text{-pose}}$ with direct continuous regression. Executable controls, planned candidates, and exploratory ideas below remain available but are excluded from its central method claim.

Promotion gate: promote only a one-factor comparison with an immutable receipt and untouched scene-disjoint evaluation

Development source: [aria_nbv/aria_nbv/vin/](#); [aria_nbv/aria_nbv/lightning/](#); [docs/contents/theory/](#)

Target exposure and horizon-weighting alternatives

The selected target sampler draws uniformly from its declared eligible pool; the learning objective averages candidate losses within state, states within horizon, and non-empty horizons uniformly. The following alternatives would change exposure or weighting, rather than add missing actions or successors. They are preserved design options, not a requirement to run every comparison.

Weighted target exposure. Equal weights $w_e = 1$ recover uniform sampling. Positive weights correspond to softmax logits $\ell_e = \log w_e$; equal eligible logits, including zero, also give a uniform draw. Ineligible targets must be removed or assigned logit $-\infty$, and an empty eligible pool produces no draw. These logits allocate examples; they are neither value predictions nor a detector’s binary confidence log-odds. For sampling without replacement, the initial normalized weights are generally not marginal inclusion probabilities. Derive inclusion probabilities from the selected procedure before using them to weight the objective, and account separately for later filtering and failed rollouts.

Adaptive exposure. Automatic goal generation prioritizes intermediate difficulty [45], while Prioritized Level Replay uses estimated learning potential [46]. These results motivate a comparison, not an optimal NBV target score: high

prediction error or disagreement can reflect noise or persistent ambiguity. Mixing prioritized and reference draws can retain nonzero probability for every eligible target. Appropriate objective weighting requires known sampling probabilities; reweighting cannot reconstruct never-proposed actions, unrecorded successors or unresolved target identities. Keep scene diversity and the evaluation population fixed when testing an exposure change.

Dataset-history priority. A proposed allocator would first draw only from admissible targets without a completed record and reopen the full admissible pool when that subset is empty. Previously completed targets then have zero probability while the restricted subset is nonempty; this is a hard priority, not a small confidence penalty. It needs an identity scope across overlapping snippets and a definition of completion. Neither this history rule nor an EVL-field containment gate is implemented in the current selectors; a record’s presence is also different from geometric surface completeness.

Horizon balance. Alternatives include horizon-stratified sampling, per-horizon loss weights w_h , or a fixed number of admitted targets per horizon and scene. Staged backward induction, fixed-horizon networks, separate horizon heads and additional queries $1 < h < b_t$ are separate choices about optimization or parameter sharing. Every added query needs its own factual target support; a weighting change needs a frozen learning identity and per-horizon support counts. None fills unknown continuations with zero.

Research TODO: Select or explicitly defer one exposure or horizon-weighting comparison only when diagnostics identify an imbalance it can address. The options above do not create an obligation to evaluate the whole catalogue.

Source: Target-sampling and horizon-balanced replay contracts

Gate: Name the changed factor, sampling probabilities or horizon weights, supported target population, fixed evaluation population and learning identity.

9.3 Ordered Pose-History Control

The pooled history encoder above discards acquisition order. The exploratory H1 control instead binds relative age to each pose:

$$a_{t,k}^{\text{hist}} = \frac{t - 1 - k}{H_{\text{max}}} \quad (67)$$

The most recent selected view has age zero; older views have progressively larger ages relative to the configured maximum horizon. H1 preserves acquisition order in a token sequence:

$$\mathbf{X}_t^{\text{hist}} = \left[\mathbf{e}_{\text{empty}}, \left\{ \mathbf{p}_{t,k}^{\text{hist}} + g(a_{t,k}^{\text{hist}}) \right\}_{k < t} \right] \quad (68)$$

The learned empty token represents the realized root state. Every subsequent row combines current-camera-relative geometry with temporal distance. Causal self-attention then contextualizes this ordered sequence:

$$\mathbf{U}_t^{\text{hist}} = \text{CausalTransformer}(\mathbf{X}_t^{\text{hist}}) \quad (69)$$

Each row of $\mathbf{U}_t^{\text{hist}}$ can aggregate only the empty token and selected poses no later than itself. The last valid row therefore carries a fixed-width summary of the complete realized prefix:

$$\mathbf{u}_{t,\text{last}}^{\text{hist}} = \text{LastValid}(\mathbf{U}_t^{\text{hist}}) \quad (70)$$

The final projection only changes representation width so the summary can join the other shared state tokens:

$$\mathbf{h}_t^{\text{histH1}} = \text{HistProj}(\mathbf{u}_{t,\text{last}}^{\text{hist}}) \quad (71)$$

Only H0 belongs to the selected method; H1 remains an implemented, unselected pose-only control. Padded entries cannot affect either summary. Prefix cardinality and chronology are not substitutes for remaining budget b_t or requested horizon h , which remain separate tokens. Because both controls observe poses rather than selected surfaces or free-space evidence, they inherit the $s_t^{\text{S0-pose}}$ sufficiency limitation identified above.

Pose order and observation content answer different questions. H1 tests whether chronology within the same pose-only state matters. A signed first view-direction moment together with $\mathbf{M}^{\text{dir}}$ would instead test whether approach-direction coverage is lost by generic history pooling. Such a directional descriptor is

scientifically meaningful, but it remains a contingent idea until held-out errors are shown to depend on target-relative approach direction; neither H1 nor directional moments supply missing surface or free-space observations.

9.4 Optional local target shell

An alternative local target-shell design samples candidate centres directly around the target. For a sampled unit direction $\mathbf{d}_i$ and radius r_i , the centre is $\mathbf{c}_i = \mathbf{p}_e + r_i \mathbf{d}_i$. The radius is sampled uniformly in metres between the configured bounds, rather than uniformly in enclosed volume. An angular-box support samples azimuth uniformly and samples the sine of elevation uniformly within their configured bounds. The actor-facing spherical cap instead samples azimuth uniformly around the three-dimensional target-to-actor axis and samples the cosine of the cone angle uniformly. These choices make direction uniform with respect to solid angle on the selected angular support. This solid-angle construction is a local proposal design; no external NBV method is taken to prescribe these exact distributions.

Sampling a centre and choosing its gaze are separate operations. The same centre can be paired with either base gaze and its configured jitter, then checked for feasibility before RRI scoring. The reported target-shell example uses one source under the privileged V0/GT protocol. It illustrates the construction but does not establish population-level performance or realism of wearer motion.

9.5 Current admission and proposed allocation: integrated illustration

The example in [Figure 34](#) follows one scene from observed boxes to a target task and then to a camera action. Matching first fixes the reference identity against which reconstruction will be evaluated. The subsequent camera draw keeps that target fixed. Between these decisions, the figure illustrates optional volume-containment and dataset-history priority gates, followed by weighted target exposure. These are design alternatives retained in the development alternatives register,

rather than stages of current observed-target admission. The proposed IoU weighting expresses association confidence rather than predicted learning value.

The floor-mat pair explains why admission uses three-dimensional overlap: its reference box is only 7.8 mm high, whereas the detected box is 100 mm high. Their similar footprints therefore yield insufficient volumetric overlap. Invalid descriptors or unusable IoU, missing compatible references, and ambiguous matches also prevent admission.

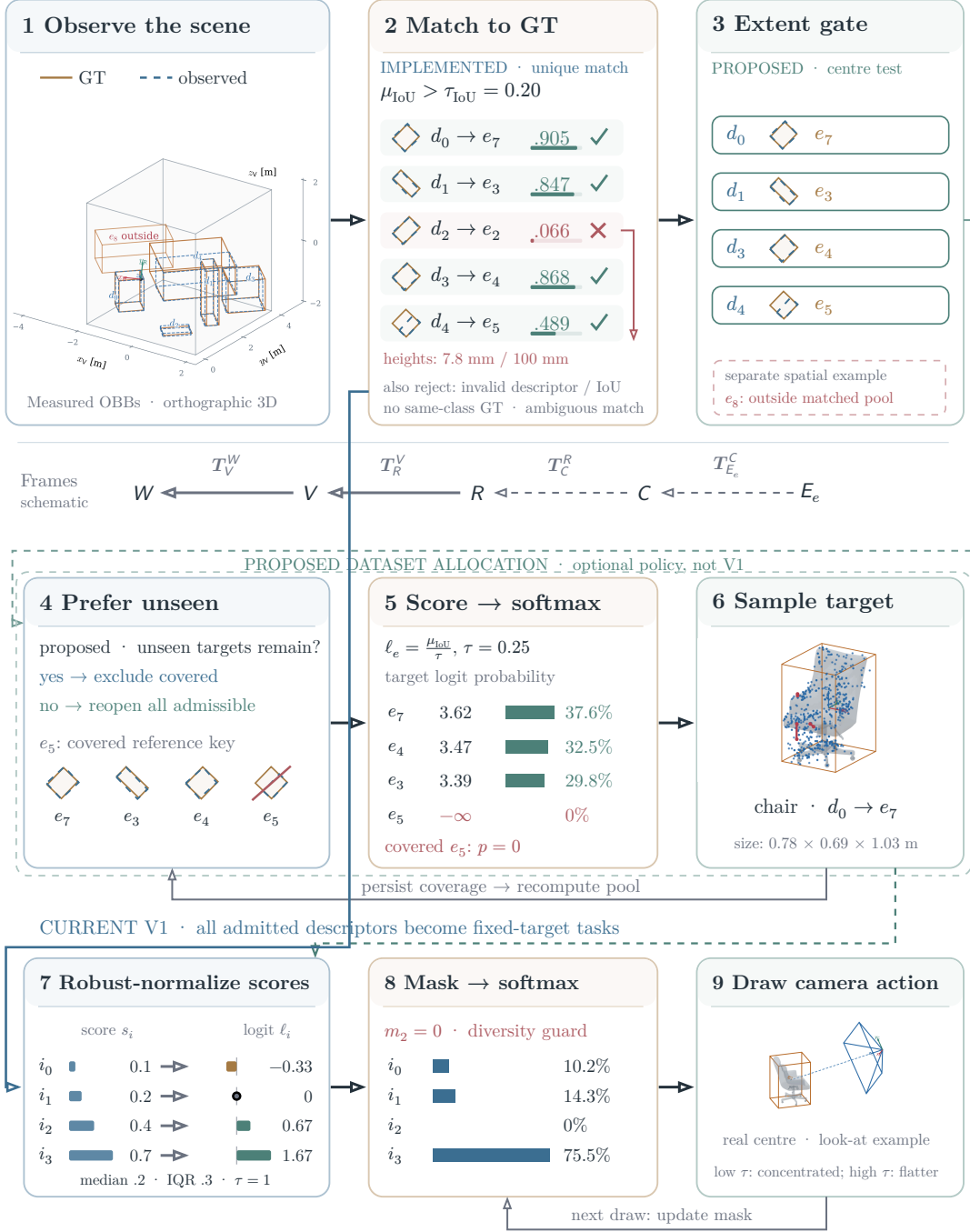
The admitted chair proposal d_0 is matched to reference e_7 and supplies the illustrative target instruction. The frame rail relates world W , voxel V , reference rig R , camera C , and object E_e . The chair’s observed dimensions are approximately $0.78 \times 0.69 \times 1.03$ m; its OBB also supplies the centre and orientation.

The translucent evaluation crop is privileged $\mathcal{M}_e^{\text{GT}}$. The overlaid points are observed semidense geometry, and the segments terminate at nearest points on mesh triangles. They illustrate point-to-surface distance, not a candidate gain or a reconstructed cloud.

The camera inset keeps the position of the sample’s one stored scene-level candidate and constructs a target-facing orientation; its recorded orientation points away from this chair. This is a geometric illustration, not a validated or recorded target-conditioned draw. The four-score example is constructed and does not represent four persisted camera poses.

Spatial containment is a separate choice. The EFM3D training filter keeps an OBB when its centre lies strictly inside every volume bound after a 0.1 m inset. A box may cross the boundary and still pass; overlap with an outside centre fails. Neither current target selector applies this filter as an admission gate.

Measured admission · proposed allocation · implemented camera draw



V0: non-padding, finite positive GT → uniform shuffle → $k = \min(K, N)$ tasks; inclusion $\frac{k}{N}$.

Figure 34: From observed boxes to a target-conditioned camera draw in ASE 83269/000056. OBBs, IoUs, mesh, observed points, and camera centre are measured; target-facing orientation, coverage, scores, and draws are illustrative. Stages 2 and 7–9 illustrate implemented matching and fixed-target action selection. Stages 3–6 show unselected containment, history-priority, and weighted-exposure alternatives, which are not part of the current admission procedure. The scene is an orthographic 3D view; the frame rail is schematic, with unsourced camera/object links dashed.

9.6 Failure-driven representation ladder

Information loss	Smallest intervention	Decisive comparison	Promotion gate
Selected observations are absent from the dynamic state.	Consume the existing selected-surface summary in the privileged S1 control.	Compare source-matched CF+ H0 with CF+ S1 to test whether selected-surface signal survives the current summary.	Require a frozen matched receipt; interpret the result as the complete S1 package rather than an actor-visible or capacity-matched effect.
Root moments discard spatial correspondence within the persisted local EVL chart.	Query the persisted chart at target support and candidate-frustum support.	Compare candidate-relative pooling with the same root evidence reduced to global moments.	Require held-out value evidence under identical labels and action support before changing the selected encoder.
The fixed chart has finite support or insufficient resolution for the diagnosed target and ray queries.	Only after measuring that failure, test a sparse or re-centred carrier with the same evidence semantics and readout.	Compare carriers with identical admitted observations, physical query support, labels, and decoder.	Require deterministic fusion, extent/resolution diagnostics, source-dropout, no-future-observation, leakage, and held-out task-sufficiency evidence.
Target evidence lies outside the root chart or lacks usable appearance cues.	Re-center logged evidence around the target or attach actor-visible appearance to supported points.	Test each extension separately against candidate-relative reads of the original persisted evidence.	Promote only after the simpler readout establishes that missing support or appearance, rather than root pooling, limits value prediction.

Table 17: Failure-driven representation ladder. Each row begins with an identifiable information loss, restores one distinction, names the comparison that isolates it, and states the evidence required for promotion. The table orders tests; it does not claim that any unselected carrier is superior.

9.7 Conditional decoder comparison

Decoder	Estimated quantity	Meaningful comparison
Direct Huber regression	continuous finite-horizon return in root-gain units	frozen per-horizon support, calibration, ranking, and endpoint evaluation
CORAL decoder-objective control	the same underlying continuous return represented by ordinal thresholds and decoded representatives	matched input plus frozen discretization, support artifact, CORAL loss, representative-value rule, and saturation diagnostics

Table 18: Selected direct-Huber path and implemented CORAL decoder-objective control. Both use the same input and underlying continuous fitted-Q target; the control changes parameterization, loss, discretization/support artifact, and representative decoding together.

CORAL preserves ordinal order but requires an additional support artifact and representative-value rule before its output again has return units [38]. Hard-invalid rows remain excluded from both objectives. The complete estimator catalogue remains available in the development register; its alternatives become eligible only

after a measured failure identifies the estimand, parameter sharing, or uncertainty assumption that should change.

9.8 Optional capacity and data pilot

9.8.1 Capacity and data pilot

Open decision: H434-CAPACITY01 — Retain this pilot only if capacity or data scarcity prevents interpretation of the selected representation comparison. The settings below describe an optional study, not a prerequisite for every Q1 result.

Source: Failure-driven representation comparisons; HREV-M03

Gate: A diagnosed limitation and an author-selected bounded comparison.

Capacity is selected only within the retained Q1 volume family. Nested 25%, 50%, and 100% subsets of the training scenes share fixed validation scenes; snippet windows, candidate rows, and counterfactual futures do not count as additional independent environments. The first sweep varies only the 3D encoder width through 8, 16, and 32 channels while holding volume resolution, readout width, candidate samples, decoder, optimization budget, and validation procedure fixed. Trainable parameters, compute, latency, and peak memory are reported with calibration, ranking, and regret.

These settings are a bounded pilot, not a scaling-law prediction. A model is not promoted because it has more parameters or fits training labels more closely. The selection rule and practically acceptable validation-regret difference are fixed before examining the comparison; promising finalists are repeated across seeds. The smallest model within that margin is retained and evaluated once on untouched test scenes. If the runs have not reached the declared optimization criterion, the comparison is reported as unconverged rather than interpreted as a capacity effect.

State and history alternatives

- $s_t^{\text{S1-surface}}$ **selected surfaces:** the executable privileged control backprojects only factual selected mesh depth, pools the causal point set, and initializes its residual at zero so a fresh S1 scorer matches H0. It lacks free/unknown

evidence, is density weighted, and has no confirmatory receipt; it therefore cannot support a geometry-benefit or deployment claim.

- $s_t^{\text{S2-ray}}$ **ray-aware memory**: a planned state would distinguish observed surface, observed free, unknown, support, uncertainty, source, and recency. It requires deterministic fusion and no-future-observation tests before a model comparison.
- **H1 ordered history**: an executable causal Transformer adds relative age to the same selected-pose tokens used by H0. It remains a pose-only capacity ablation until repeated-seed endpoint evidence justifies chronology.
- **Other carriers**: target-centred EFM3D re-lifting, appearance-on-point, sparse TSDF/SDF encoders, object-aware 3D Gaussian splats, and SceneScript context are hypotheses for diagnosed representation failures, not parallel thesis methods.

Object-aware 3D Gaussian splatting is an exploratory response to a diagnosed need for renderable memory or target-conditioned primitive confidence. It can attach soft object membership, opacity, and confidence to renderable primitives, but per-scene optimization and mask supervision change both state construction and computational cost; promote it only after a matched experiment identifies renderability or primitive uncertainty as the limiting distinction [2]. SceneScript is a separate exploratory response to a diagnosed need for broad layout or object hypotheses. Its structured commands can express global architectural and object relations, but those semantic hypotheses do not replace causal local surface, free-space, and unknown-space evidence; activate it only after the local relational state is shown insufficient [47].

Interaction alternatives

The ladder below is a development plan rather than a family of co-equal thesis methods. It orders feature integration by the smallest additional dependency introduced after the current A0/A1 controls. “Implemented” means that the tensor path and acceptance properties exist; “eligible” names the first one-factor comparison that may be frozen after its activation evidence is present; and

“possible extension” preserves a conditional test whose promotion still depends on a diagnosed failure.

Stage and status	Integrated feature	Scientific purpose	Principal risk and promotion gate
F0 — implemented	candidate-relative geometric query	Complete candidate pose relative to root and current camera, plus target pose relative to the candidate.	Already present; retain transform-direction and local-frame tests. It supplies geometry, not selected-observation state.
F1 — implemented	A0/A1 state fusion	Compare identical-input independent-row fusion with candidate-to-state cross-attention.	The controls are not parameter matched; report parameters and runtime and require a frozen held-out comparison.
F2 — first eligible promotion	candidate-relative relation embeddings	Embed each candidate’s relative transform to the target and pose-bearing causal history or state elements before A1 reads them.	Activate only after state, support, label, and optimization failures are ruled out and held-out residuals retain relation structure.
F3 — separate support-conditioned hypothesis	permutation-invariant candidate-set summary [27]	Test whether a versioned context-row set supplies ranking information beyond the physical state and query-local relations.	Freeze context rows, masks, duplicate semantics, and the score estimand. An action value conditioned on the offered set requires that set in its state and learning target; a ranking score alone does not establish expected return.
F4 — separate support-conditioned hypothesis	candidate self-attention [28]	Test whether pairwise candidate relations explain residual ranking error beyond an invariant set summary.	Adds quadratic interaction and support sensitivity; promote only after a frozen F3 profile fails under matched support.
F5 — possible extension	iterative or recurrent state reading	Test whether one candidate query needs repeated access to an already adequate causal state.	May hide missing state information behind capacity; promote only after the state, F2 relations, and simpler fusion controls pass.

Table 19: Development-only architectural feature-integration ladder. Status records implementation maturity, while promotion remains conditional on a named scientific failure and a one-factor comparison.

Candidate-relative relation embeddings are the first eligible architectural promotion because they can preserve the current per-candidate value interface while making query-to-context geometry explicit. Query-centric models use local coordinate systems and relative positional embeddings so that the active query can read other elements through their relation to it [26]. In ARIA-NBV, the same construction can attach a shared embedding of candidate–target and candidate–history transforms to the A1 key/value tokens. Applied independently to every

candidate, it preserves candidate-row equivariance and does not make an action’s value depend on unrelated candidate rows.

F0 already exposes complete relative poses, so F2 adds an inductive bias rather than new raw information. It can redundantly parameterize the same transform, overfit one proposal generator, increase cost with candidate–history pairs, or encode an inappropriate symmetry if gravity, metric scale, camera direction, or target orientation is suppressed. Geometric priors reduce the hypothesis space only when the chosen symmetry matches the task [25]. F2 therefore activates only when held-out residual error varies systematically with the relevant relations, a non-learned probe predicts that structure, and the pattern survives matched seeds, capacity, state, support, label, and optimization controls. The existence of relation embeddings elsewhere is not activation evidence.

F3 and F4 deliberately change the estimand’s conditioning context. Their natural output is

$$U_{\theta,h}^{\Xi}(z_t^{\sigma}, e, \mathcal{C}_t^{\text{ctx}}, q_{t,i}) \quad (72)$$

where $\mathcal{C}_t^{\text{ctx}}$ is a separately versioned context-row set. The score may depend on which other candidate rows are present. An independent $Q(s, a)$ conditions on the chosen state and action; a set-conditioned value instead uses the augmented state $(s, \mathcal{C}_t^{\text{ctx}})$. For example, adding a duplicate pose leaves the independent model’s prediction unchanged, whereas a set summary may change unless its duplicate semantics prevent that change. Such a model can represent an action value when the context set is part of the state and expected-return target. Set conditioning alone neither rules out an action value nor establishes that a fitted ranking score estimates one.

Its profile must state whether context rows are materialized, action-valid, or label-supported; which mask enters the model; whether duplicate poses are one semantic action or distinct proposals; and which A0/A1 invariants are retained or replaced. Until permutation, invalid-row, duplicate, and mask-sensitivity tests exist and the expected-return target is defined and evaluated, F3/F4 remain

development hypotheses reported as support-conditioned ranking models rather than as $\hat{Q}_{\theta,h}^{\sigma,\Xi}$.

Full SE(3)-Transformer and geometric-algebra Transformer variants remain outside the core study. The current local frames already remove arbitrary origin dependence while retaining gravity-aligned and metric quantities that may affect sensing. No measured symmetry failure presently warrants the representation, implementation, and compute commitments of an exact equivariant backbone. Candidate-set interaction remains a separate hypothesis because it changes the conditioning context rather than only how physical relations are encoded.

Objective and estimator alternatives

Direct Huber regression and the implemented Double-Q backup define the selected path. The remaining entries change the decoder-objective package, parameter sharing, estimand, decomposition, or uncertainty representation. Their activation conditions are proposed falsification gates rather than evidence that a failure has occurred.

Alternative	Scientific role	Quantity estimated	Condition for a meaningful comparison
Direct Huber regression	selected	continuous finite-horizon return in root-gain units	frozen per-horizon support, calibration, ranking, and endpoint evaluation
CORAL decoder-objective	implemented matched-package control	the same underlying continuous fitted-Q target represented by ordinal thresholds and decoded representatives	freeze input, discretization, support artifact, CORAL loss, representative-value rule, and saturation diagnostics; hard-invalid rows remain excluded
Separate or fixed-horizon heads	contingent parameter-sharing control	the same horizon-indexed values without the shared scalar-horizon interface	compare only after each horizon has sufficient matched support
Double-Q backup	implemented selected estimator	greedy finite-support continuation with separated selection and delayed evaluation	report support and max-bias diagnostics; it cannot repair state aliasing
Behavior-return regression	distinct-estimand control	return under the chain-generating behavior policy	must not be called greedy value unless behavior and target continuation coincide
One-step base + residual	contingent decomposition	immediate gain plus learned continuation residual	use additive units and no candidate-table mean centering
Quantile value	exploratory idea	a return distribution rather than one scalar expectation	justify with measured stochasticity or aliasing and freeze projection, decoding, and calibration

Table 20: Development catalogue of value-model alternatives by scientific role and estimand. Executability, selection, and evidence are distinct; activation requires a diagnosed failure and a one-factor comparison.

The CORAL ordinal Q decoder is an executable decoder-objective path over the same input and underlying continuous fitted-Q target rather than a different action-value definition. It represents that target with ordinal threshold labels, optimizes the CORAL loss, and requires a frozen discretization/support artifact plus representative decoding to return to continuous units [38]. It is therefore a matched-package control rather than an output-only ablation. Fixed-horizon heads isolate parameter sharing only after each horizon has matched support. Double Q is already the selected backup estimator; it may limit max bias but cannot repair state aliasing or missing support.

Behavior-return regression changes the estimand more fundamentally. Without off-policy correction it follows the continuation policy that generated the retained chain rather than the greedy continuation used by fitted Q [15, Secs. 5.2 and 5.5, pp. 96–97 and 103–109]. A one-step residual remains additive and may

not be mean-centred across a candidate table because unrelated or duplicate rows would then redefine an action’s target. Quantile regression becomes meaningful only after stochastic return variation or state aliasing is measured and a distributional Bellman projection, support, decoding rule, and calibration protocol are frozen.

The selected finite-horizon path retains population, action and label support, actor-visible input qualification, and the exact two-step and recursion checks described in Experimental Design. An independent learned-versus-named-reference endpoint comparison does not require positive oracle headroom or a separately trained learned-myopic control. Headroom and gap recovery can diagnose a non-myopic mechanism when their own comparison data are available. A representation, optimizer or decoder control becomes useful only when it addresses a measured failure under matched support.

Conflict: Specify the target and observation sources for the core RQ2 comparison and the optional fixed-scorer observation-update comparison. Align the selected method description and certification scope with that decision; this register changes neither the exact two-step nor recursion support gates.

Source: RQ2/I12 information-profile decision; Experimental Design learning-readiness table

Gate: Declare the core actor state, optional intervention and claimed horizons before freezing policy evaluation and certification.

Ranked scientific-core work

Architecture and data TODOs use one extension of the thesis draft-marker contract rather than a separate backlog. Priority is ordinal and claim-centred: C0 protects a validity or information boundary shared by the central claims; C1 blocks the next principal-RQ inference; C2 discriminates a diagnosed alternative explanation; C3 is an exploratory extension outside the core claim. Readiness (*ready*, *blocked*, or *contingent*) is orthogonal, so a blocked C0 remains scientifically more critical than a ready C2 even though only the latter may be immediately executable. Every marker must name its affected claim, source, promotion gate, and blocker when blocked. The policy deliberately avoids weighted scores whose apparent precision would become stale as evidence changes.

Scientific-core TODO C0: Freeze and audit the actor-visible target corpus before interpreting a learned result as an end-to-end actor-visible target-conditioned comparison. Controlled privileged-input comparisons require their declared source and leakage audit, but do not establish deployment.

Rank C0 (validity boundary) — domain data — readiness ready

Affected claim: actor/oracle separation and deployability of the target-conditioned value claim

Source: $v1_observed$ target protocol and frozen corpus manifest

Gate: scene-disjoint corpus, explicit matching failures, source dropout, descriptor identity, and leakage audit

Scientific-core TODO C1: Implement the smallest observation-updated state that preserves surface, free, unknown, support, uncertainty, source, and recency.

Rank C1 (core inference) — domain architecture — readiness blocked

Affected claim: whether actor-visible state retains information needed for target-specific bounded return

Blocked by: an admitted actor-visible selected-observation data path

Source: s_t^{S2-ray} candidate realization and selected-observation replay contract

Gate: deterministic fusion, no-future-observation test, source masks, and matched held-out comparison against $s_t^{S0-pose}$

Scientific-core TODO C1: Produce the immutable evidence chain that separates implementation parity, learned recursion accuracy, and policy benefit. The optional recovery ratio requires a sufficiently positive learned-myopic-to-oracle-lookahead reference gap under its declared conditions.

Rank C1 (core inference) — domain data — readiness blocked

Affected claim: supported finite-horizon learning and independent held-out endpoint policy comparison

Blocked by: validated rollout stores under the selected target, state, candidate, and reward protocols

Source: exact- Q_2 census, recursion validation and paired named-reference endpoint receipt

Gate: frozen eligible population, objective-specific support, error tolerances, actor information contract and untouched scene-disjoint evaluation

Scientific-core TODO C2: Promote exactly one control that tests the diagnosed limitation; do not bundle state, interaction, and objective changes.

Rank C2 (diagnostic discriminator) — domain architecture — readiness contingent

Affected claim: alternative explanation for residual value error after state and support gates pass

Source: feature-integration ladder F2–F5, H1, CORAL, residual, and separate-head controls

Gate: one measured failure mapped to one factor-matched comparison

Scientific-core TODO C3: Retain these ideas as scoped extensions; do not treat conceptual interest as evidence for inclusion in the core method.

Rank C3 (exploratory extension) — domain architecture — readiness contingent

Affected claim: post-core generalization beyond the finite-candidate offline study

Source: 3DGS memory, SceneScript context, continuous policy, and quantile-value ideas

Gate: completed core evidence chain plus a failure that requires the changed state, action, or estimand contract

Additional conditional hypotheses

The Introduction and Discussion share this register rather than duplicate their design catalogues. Each alternative needs a diagnosed limitation and a comparison that can distinguish its cause.

Relations and symmetry. Heterogeneous entity, observer, target, candidate and observation-event relations can encode transforms, visibility, uncertainty and elapsed time. Coordinate-consistent augmentation and known object symmetries must retain gravity, scale, occlusion, viewing direction, target orientation and physical action order. Root-, target- and candidate-local frames are alternatives, not proven equivalent learners.

Spatial memory. Local EVL evidence can be complemented by sparse point-and-ray memory beyond its extent and coarse entity or room context [6], [47]. Target, frustum and intersection pools test spatial correspondence before a new carrier or renewed perception. Signed first and second direction moments, with spherical harmonics as a further option, test approach-direction information lost by history pooling. Calibrated-frustum reads require actual calibration rather than merely sampling a candidate-to-target segment.

Learning efficiency. Shared parameters, dense immediate labels, horizon curricula, candidate permutations, target resampling, causal sub-prefixes and source-aware modality dropout are candidate comparisons. Broader ASE geometry, semantics and temporal consistency could support pretraining without gain labels. More overlapping snippets still do not add independent scenes.

Motion and multiple targets. A prior on natural pose increments with an exploratory tail could test wearer-motion relevance [4]. Evaluating independent target-conditioned queries in one batch need not change the single-target task. Choosing which target to acquire next instead requires a cross-target objective. Reusing cached scene evidence must respect target-dependent encodings and coordinate frames. A hierarchical target-then-pose policy needs per-target candidate availability, descriptors and endpoint aggregation; shared scene encoding alone supplies none of these. A target-then-pose query could condition on the

target descriptor, remaining budget, requested horizon and acquisition step. Per-target candidate tables and an explicit target-by-candidate availability mask are alternative constructions. Continuous actions remain future work, with distinct feasibility, safety, cost and simulator assumptions and a matched discrete reference. Recurrent reading is an alternative only when errors remain after explicit causal state and time conditioning; trajectory-shaped storage alone does not justify it.

Optional target-allocation rationale

9.8.2 Reliability and causal observation support

Confidence is useful here as an estimate of detection reliability, provided it is calibrated for that task. Guo et al. study classification calibration and temperature scaling [48]; calibration of the present three-dimensional detector remains to be established. Confidence stored with a ground-truth box is a separate field and cannot substitute for the detector’s estimate.

Observed-to-ground-truth IoU describes localization overlap. It can associate a detection with the correct evaluation object, but does not measure how much its reconstruction could improve. Grouping training examples by ground-truth IoU is a privileged data-selection procedure, even when the actor never receives IoU. Such a procedure must leave the reference evaluation population unchanged and cannot serve as an observation-only target-selection policy.

Visible mask area describes how much of an image the target occupies. Adding this area across frames measures accumulated exposure: repeated frontal views increase the sum without revealing the cabinet’s side. A projected OBB footprint is less informative about visibility because it ignores occlusion. Viewing-direction diversity and the union of observed surface elements offer complementary descriptions of acquired evidence. These quantities could define training strata, with repeated observations counted according to a declared rule. They describe **observed support**, not full-object completeness unless the total surface is known. Only evidence acquired by the current decision time may contribute; missing masks or geometry remain unavailable measurements.

Jeong et al. use object probabilities and opacity to weight information gain towards poorly supported Gaussian primitives of a requested object [2]. That criterion prioritizes views **within** one target. Choosing which targets to train on is a separate problem; the method does not prescribe a product of confidence, IoU and accumulated image area for that purpose.

9.8.3 Training allocation without changing the evaluation question

A library diagnostic allocator can group admitted observed targets by supplied causal observation support and balance exposure across these groups, or **strata**. It validates the support against the current decision and source population; it does not produce the support measurements or establish their availability for a corpus. A proposed model comparison would contrast this allocation with uniform sampling while preserving scene diversity. Both models would be evaluated on the same population, so the comparison would test training allocation rather than a change in which objects count as successes.

Adaptive curricula and completed-record priority remain unselected extensions. Their design questions are retained in the development alternatives register; neither changes the fixed evaluation population used here.

Bibliography

- [1] N. Frahm *et al.*, “VIN-NBV: A View Introspection Network for Next-Best-View Selection.” [Online]. Available: <https://arxiv.org/abs/2505.06219>
- [2] S. Jeong, E. Lee, J. Kim, and A. Kim, “Informative Object-centric Next Best View for Object-aware 3D Gaussian Splatting in Cluttered Scenes.” [Online]. Available: <https://arxiv.org/abs/2602.08266>
- [3] A. Guédon, T. Monnier, P. Monasse, and V. Lepetit, “MACARONS: Mapping And Coverage Anticipation with RGB Online Self-Supervision,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. [Online]. Available: <https://arxiv.org/abs/2303.03315>
- [4] J. Engel *et al.*, “Project Aria: A New Tool for Egocentric Multi-Modal AI Research.” [Online]. Available: <https://arxiv.org/abs/2308.13561>
- [5] Meta Platforms Inc., “Aria Synthetic Environments Dataset.” [Online]. Available: https://facebookresearch.github.io/projectaria_tools/docs/open_datasets/aria_synthetic_environments_dataset
- [6] J. Straub, D. DeTone, T. Shen, N. Yang, C. Sweeney, and R. Newcombe, “EFM3D: A Benchmark for Measuring Progress Towards 3D Egocentric Foundation Models.” [Online]. Available: <https://arxiv.org/abs/2406.10224>
- [7] X. Chen, Q. Li, T. Wang, T. Xue, and J. Pang, “GenNBV: Generalizable Next-Best-View Policy for Active 3D Reconstruction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16436–16445. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2024/html/Chen_GenNBV_Generalizable_Next-Best-View_Policy_for_Active_3D_Reconstruction_CVPR_2024_paper.html
- [8] C.-Y. Lu *et al.*, “Hestia: Voxel-Face-Aware Hierarchical Next-Best-View Acquisition for Efficient 3D Reconstruction,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2026. [Online]. Available: https://openaccess.thecvf.com/content/WACV2026/papers/Lu_

Hestia_Voxel-Face-Aware_Hierarchical_Next-Best-View_Acquisition_for_Efficient_3D_Reconstruction_WACV_2026_paper.pdf

- [9] Z. Jia, Y. Li, Q. Hao, and S. Zhang, “PB-NBV: Efficient Projection-Based Next-Best-View Planning Framework for Reconstruction of Unknown Objects,” *IEEE Robotics and Automation Letters*, vol. 10, no. 7, pp. 7444–7451, 2025, doi: 10.1109/LRA.2025.3573631.
- [10] A. Guédon, P. Monasse, and V. Lepetit, “SCONE: Surface Coverage Optimization in Unknown Environments by Volumetric Integration,” in *Advances in Neural Information Processing Systems*, 2022. [Online]. Available: <https://arxiv.org/abs/2208.10449>
- [11] W. Jiang, B. Lei, and K. Daniilidis, “FisherRF: Active View Selection and Uncertainty Quantification for Radiance Fields using Fisher Information.” [Online]. Available: <https://arxiv.org/abs/2311.17874>
- [12] M. Strong, B. Lei, A. Swann, W. Jiang, K. Daniilidis, and M. K. III, “Next Best Sense: Guiding Vision and Touch with FisherRF for 3D Gaussian Splatting.” [Online]. Available: <https://arxiv.org/abs/2410.04680>
- [13] K. Ashutosh, S. Kumar, and S. Chaudhuri, “3D-NVS: A 3D Supervision Approach for Next View Selection.” [Online]. Available: <https://arxiv.org/abs/2012.01743>
- [14] L. Liu *et al.*, “Object-Aware Guidance for Autonomous Scene Reconstruction,” *ACM Transactions on Graphics*, vol. 37, no. 4, 2018, doi: 10.1145/3197517.3201295.
- [15] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA: The MIT Press, 2018. [Online]. Available: <http://incompleteideas.net/book/the-book-2nd.html>
- [16] M. Lauri, D. Hsu, and J. Pajarinen, “Partially Observable Markov Decision Processes in Robotics: A Survey,” *IEEE Transactions on Robotics*, vol. 39, no. 1, pp. 21–40, 2023, doi: 10.1109/TRO.2022.3200138.

- [17] K. De Asis, A. Chan, S. Pitis, R. S. Sutton, and D. Graves, “Fixed-Horizon Temporal Difference Methods for Stable Reinforcement Learning,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 3741–3748. doi: 10.1609/aaai.v34i04.5784.
- [18] T. Schaul, D. Horgan, K. Gregor, and D. Silver, “Universal Value Function Approximators,” in *Proceedings of the 32nd International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 37. PMLR, 2015, pp. 1312–1320. [Online]. Available: <https://proceedings.mlr.press/v37/schaul15.html>
- [19] F. Pardo, A. Tavakoli, V. Levdik, and P. Kormushev, “Time Limits in Reinforcement Learning,” in *Proceedings of the 35th International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 80. PMLR, 2018, pp. 4045–4054. [Online]. Available: <https://proceedings.mlr.press/v80/pardo18a.html>
- [20] M. Lauri, J. Pajarinen, J. Peters, and S. Frintrop, “Multi-Sensor Next-Best-View Planning as Matroid-Constrained Submodular Maximization,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5323–5330, 2020, doi: 10.1109/LRA.2020.3007445.
- [21] S. Fujimoto, D. Meger, and D. Precup, “Off-Policy Deep Reinforcement Learning without Exploration,” in *Proceedings of the 36th International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 97. PMLR, 2019, pp. 2052–2062. [Online]. Available: <https://proceedings.mlr.press/v97/fujimoto19a.html>
- [22] A. Kumar, A. Zhou, G. Tucker, and S. Levine, “Conservative Q-Learning for Offline Reinforcement Learning,” in *Advances in Neural Information Processing Systems*, 2020, pp. 1179–1191. [Online]. Available: <https://papers.nips.cc/paper/2020/hash/0d2b2061826a5df3221116a5085a6052-Abstract.html>
- [23] D. Ernst, P. Geurts, and L. Wehenkel, “Tree-Based Batch Mode Reinforcement Learning,” *Journal of Machine Learning Research*, vol. 6, pp. 503–556, 2005, [Online]. Available: <https://jmlr.org/papers/v6/ernst05a.html>

- [24] T. Ni *et al.*, “Bridging State and History Representations: Understanding Self-Predictive RL,” in *International Conference on Learning Representations*, 2024. [Online]. Available: https://proceedings.iclr.cc/paper_files/paper/2024/hash/666c1861d709bd84e20b6e0e02a2c223-Abstract-Conference.html
- [25] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, “Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges.” [Online]. Available: <https://arxiv.org/abs/2104.13478>
- [26] Z. Zhou, J. Wang, Y.-H. Li, and Y.-K. Huang, “Query-Centric Trajectory Prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [27] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. J. Smola, “Deep Sets,” in *Advances in Neural Information Processing Systems*, 2017. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/hash/f22e4747da1aa27e363d86d40ff442fe-Abstract.html
- [28] J. Lee, Y. Lee, J. Kim, A. R. Kosiolek, S. Choi, and Y. W. Teh, “Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks,” in *Proceedings of the 36th International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 97. PMLR, 2019, pp. 3744–3753. [Online]. Available: <https://proceedings.mlr.press/v97/lee19d.html>
- [29] H. Zhao, L. Jiang, J. Jia, P. H. S. Torr, and V. Koltun, “Point Transformer,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2021/papers/Zhao_Point_Transformer_ICCV_2021_paper.pdf
- [30] C. Choy, J. Gwak, and S. Savarese, “4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. [Online]. Available: <https://arxiv.org/abs/1904.08755>

- [31] V. G. Satorras, E. Hoogeboom, and M. Welling, “E(n) Equivariant Graph Neural Networks,” in *Proceedings of the 38th International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 139. PMLR, 2021, pp. 9323–9332. [Online]. Available: <https://proceedings.mlr.press/v139/satorras21a.html>
- [32] D. Vutetakis and J. Xiao, “Active perception network for non-myopic online exploration and visual surface coverage,” *The International Journal of Robotics Research*, vol. 44, no. 2, pp. 247–272, 2025, doi: 10.1177/02783649241264577.
- [33] S. Pan, H. Hu, X. Huang, B. Wingender, and M. Bennewitz, “ObjView-Bench: Rethinking Difficulty and Deployment for Object-Centric View Planning.” [Online]. Available: <https://arxiv.org/abs/2605.10707>
- [34] Meta Platforms Inc., “Aria Training and Evaluation toolkit (ATEK).” [Online]. Available: <https://github.com/facebookresearch/ATEK>
- [35] Meta Platforms Inc., “ATEK Data Store documentation.” [Online]. Available: https://github.com/facebookresearch/ATEK/blob/main/docs/ATEK_Data_Store.md
- [36] C. Boutilier *et al.*, “Planning and Learning with Stochastic Action Sets,” in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018, pp. 4674–4682. doi: 10.24963/ijcai.2018/650.
- [37] H. van Hasselt, A. Guez, and D. Silver, “Deep Reinforcement Learning with Double Q-learning.” [Online]. Available: <https://arxiv.org/abs/1509.06461>
- [38] W. Cao, V. Mirjalili, and S. Raschka, “Rank consistent ordinal regression for neural networks with application to age estimation.” [Online]. Available: <https://arxiv.org/abs/1901.07884>
- [39] N. Kallus and A. Zhou, “Confounding-Robust Policy Evaluation in Infinite-Horizon Reinforcement Learning,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2020, pp. 22293–22304. [Online]. Available: <https://proceedings.neurips.cc/paper/2020/hash/fd4f21f2556dad0ea8b7a5c04eabebda-Abstract.html>

- [40] A. Baisero and C. Amato, “Unbiased Asymmetric Reinforcement Learning under Partial Observability.” [Online]. Available: <https://arxiv.org/abs/2105.11674>
- [41] K. M. Lynch and F. C. Park, *Modern Robotics: Mechanics, Planning, and Control*. Cambridge University Press, 2017. [Online]. Available: <https://hades.mech.northwestern.edu/images/2/25/MR-v2.pdf>
- [42] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, “On the analysis of movement smoothness,” *Journal of NeuroEngineering and Rehabilitation*, vol. 12, p. 112, 2015, doi: 10.1186/s12984-015-0090-9.
- [43] S. Ross, G. Gordon, and D. Bagnell, “A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning,” in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, in Proceedings of Machine Learning Research, vol. 15. PMLR, 2011, pp. 627–635. [Online]. Available: <https://proceedings.mlr.press/v15/ross11a.html>
- [44] P. J. Ball, L. Smith, I. Kostrikov, and S. Levine, “Efficient Online Reinforcement Learning with Offline Data,” in *Proceedings of the 40th International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 202. PMLR, 2023, pp. 1577–1594. [Online]. Available: <https://proceedings.mlr.press/v202/ball23a.html>
- [45] C. Florensa, D. Held, X. Geng, and P. Abbeel, “Automatic Goal Generation for Reinforcement Learning Agents,” in *Proceedings of the 35th International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 80. PMLR, 2018, pp. 1515–1528. [Online]. Available: <https://proceedings.mlr.press/v80/florensa18a.html>
- [46] M. Jiang, E. Grefenstette, and T. Rocktäschel, “Prioritized Level Replay,” in *Proceedings of the 38th International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 139. PMLR, 2021, pp. 4940–4950. [Online]. Available: <https://proceedings.mlr.press/v139/jiang21b.html>

- [47] A. Avetisyan *et al.*, “SceneScript: Reconstructing Scenes With An Autoregressive Structured Language Model.” [Online]. Available: <https://arxiv.org/abs/2403.13064>
- [48] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On Calibration of Modern Neural Networks,” in *Proceedings of the 34th International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 70. PMLR, 2017, pp. 1321–1330. [Online]. Available: <https://proceedings.mlr.press/v70/guo17a.html>
- [49] N. De Cao, W. Aziz, and I. Titov, “The Power Spherical distribution.” [Online]. Available: <https://arxiv.org/abs/2006.04437>
- [50] M. Pharr, W. Jakob, and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*, 4th ed. MIT Press, 2023. [Online]. Available: <https://pbr-book.org/4ed/>

A Reproducibility Record

The report bundle is the numerical interface for the store, parameter, and confirmatory-result tables loaded through the document's report loader. Its schema version, evidence status, store manifests, typed parameter rows, failure records, and sidecar hashes preserve provenance without duplicating analysis logic in Typst. The loader rejects schema or status mismatches before those tables are rendered. Compact labels and digest prefixes are used below; complete store identifiers, paths, and hashes remain in the machine-readable bundle.

Results also draws on independently frozen development diagnostics in `docs/contents/evidence/`. Their source map is:

- **Phase-A proposal support:** `candidate_family_phase_a_wp02.json` is the authenticated source. Its generated `candidate_family_phase_a_wp02_projection.json` supplies the Results counts and proposal-support table, while `candidate_family_phase_a_wp02_audit_heatmap.svg` supplies the candidate-family survival figure. The same projection supplies the bounded experiment contract below.
- **Proposal-radius and transition consistency:** `candidate_scale_h2_diagnostic_20260918.json` supplies the paired-radius and successor-agreement paragraphs under Historical proposal-support and profile diagnostics in [Chapter 6](#).
- **B0/S profile comparison:** `rollout_profile_development_confirmation_20260920.json` supplies the completion, admission, and conditional endpoint results under Candidate-Profile Development Diagnostic in [Chapter 6](#).
- **Historical broad collection:** the archive `rollout_recipe_archive_20260928/historical-results/broad-scene-development-20260926/` retains the projection, original manifest, ledger, queue and writer configuration. Its family-validity SVG is in the same archive's `historical-results/recipe-terminal-20260926/broad-readout-v2/` the roster receipt is under `source-`

receipts/overnight-recipe-collection-20260924/. The archive manifest maps these original files. The projection’s SHA-256 is 9820fb0d5046fd4d49 (prefix); it supplies [Section C.20](#), not the later exact-box cohorts. Frozen attempt summaries do not revalidate child stores or queue completion.

- **Camera-stable rendering and component timing:** the frozen October generation diagnostics linked beside their Results paragraphs own the paired scenes, endpoint audits and clock definitions. The earlier failed contrast is retained separately from the successful one.
- **Later development cohorts and input compatibility:** the original deep audit, second collection and actor-root increment at source revision 20c51d8a5e58bcbf77b1efc65a9339bbc52a36d7 own their separate rosters, seals and input joins. [Table 11](#) summarizes those populations; [Section C.2](#) records their admission limits.
- **Selected-camera coverage and continuity:** the new derivative manifest under docs/figures/thesis/vector/06-results/trajectory-statistics/ binds all 97 current store-content digests to the second collection’s frozen census. Its captured statistics supply [Figure 24](#) and [Figure 46](#). The fixed initial root-target frame differs from a target-object coordinate frame; the root anchors azimuth and is excluded from occupancy. The 12 by 8 cells are uniform in azimuth and sine elevation. Canonical geodesic rotation and translation increments remain untimed. The adjacent collection-figure provenance record binds [Figure 20](#) and [Figure 21](#) to byte-identical sealed figures, without editing the experiment archive.
- **Generation supervision re-analysis:** the frozen compact projection under docs/typst/thesis/data/generation-posthoc-20261011/ supplies [Table 12](#), [Figure 22](#) and [Figure 23](#). Its authenticated historical member map binds reuse of the second R1 cohort; the standalone report and new print export retain distinct identities.
- **Larger one-step development readback:** the closed native-v3 pilot under docs/typst/thesis/data/scorer-q1-posthoc-20261007/ supplies the paired

scene contrasts and unchanged [Figure 28](#). The dedicated native-v3 loader preserves the existing confirmatory and v2 interfaces.

- **One-step learning:** the same immutable revision’s `experiments/2026-09-research/development_corpus_readiness_20260929/e23` owns the four folds and frozen practical-effect screen. The independently linked exploratory reduction in Results retains the original failed verdict.

These diagnostics retain their own populations, source identities, and failure records. The report loader’s validation does not validate them or promote their bounded development findings to confirmatory results.

The implementation anchors behind the report-loader interface are the immutable VIN offline source store, the task-specific rollout replay store, and the derived finite-horizon training cache. Their exact schemas, normalized joins, chunking, codecs, receipt versions, and execution commands remain manifest-owned reproducibility metadata rather than scientific entities in Chapter 3. A row reported through this interface is admissible only when the bundle resolves these identities and preserves source role, missingness, and failure records.

Evidence class	Store	Manifest digest	Validated
Development fixture	store S1	not-a-real-m...	yes

Table 21: Loaded development-fixture provenance. The fixture verifies the document interface only and is not scientific evidence.

The bundled development fixture contains synthetic values solely to test typed loading and missingness. Its numerical rows are deliberately not rendered or interpreted.

A.1 Bounded Phase-A experiment contract

The Phase-A proposal-support diagnostic is projected independently of the confirmatory report bundle because it is a historical, non-confirmatory audit. Its claim is limited to authenticated proposal-support counts and gate predicates. The table below is generated from the same authenticated artifact as the Results counts and figure. “Not applicable” is a factual stage boundary, whereas a missing gaze-variant identity is retained as a reproducibility limitation.

The receipt binds two support floors. For a candidate width N_q , the root floor is $\max(12, \lceil 0.25N_q \rceil)$ hard-valid rows; the Phase-A width of 60 therefore requires 15 rows. Independently, every applicable diagnostic family must contribute at least one row to the final shell, and the applicable non-forward target-aware families must contribute at least three rows in total. A family marked inapplicable does not fail this gate. Missing applicability is an unknown provenance state and fails closed. These predicates characterize proposal support; they neither assign utility to an invalid row nor establish that an admitted candidate is useful.

Phase A stops before rendering, reward labeling, or policy execution. Its flat gain statistic is therefore unavailable with label denominator zero and eligible-state denominator zero, rather than being inferred from geometric dispersion. The authenticated counts and failed predicates are reported in [Chapter 6](#); they are a bounded proposal-support diagnostic and do not constitute reconstruction or policy evidence.

All persisted Phase-A geometry uses $d_{t,e}^{\text{current}} = \|\mathbf{p}_e^W - \mathbf{c}_{R_t}^W\|_2$, $\tilde{\mathbf{c}}_{t,i}^{\text{support}} = (\mathbf{B}_{r,t}^{\text{Z-up}})^\top \frac{\mathbf{c}_{t,i}^W - \mathbf{c}_{R_t}^W}{d_{t,e}^{\text{current}}}$, $\tilde{\mathbf{p}}_{t,e}^{\text{support}} = (\mathbf{B}_{r,t}^{\text{Z-up}})^\top \frac{\mathbf{p}_e^W - \mathbf{c}_{R_t}^W}{d_{t,e}^{\text{current}}}$, $\|\tilde{\mathbf{p}}_{t,e}^{\text{support}}\|_2 = 1$. The columns of the gravity-aligned basis are the horizontal expansion-to-target direction, its world-up left direction, and world up. Candidate centres and target-relative vectors are divided by the current three-dimensional target distance, while camera-forward directions are rotated by the same basis and remain unit vectors. Target-forward and target-lateral coordinates are therefore invariant to factual rig yaw rig-frame components cannot carry those axis labels directly.

Contract field	Receipt-bound value
Candidate width	60 rows per factual state
Configured families	forward_local, target_bearing_local, lateral_target_bypass
Family-support rule	family-support-floor-v1 known applicability required: yes
Candidate identity	scene_key, state_key, candidate_id
Family attribution	component_name
Coordinate convention	phase-a-target-aligned-z-up-v1
Target source	gt_obbs_oracle
Random streams	bound by writer-config and generation-revision hashes not applicable: Phase-A stops before policy selection
Reward and execution	not applicable: no rendering, reward labels, or executed transition
Action support	final_valid_action_shell with attempted, valid, and selected rows retained
Label support	absent by design; target-root gain and diagnostic RRI denominators are zero
Gaze-variant identity	historical v4 bundle retains candidate rows but no non-null gaze-jitter identity
Generation command	make phase-a-thesis-projection

Table 22: Generated Phase-A experiment contract. The source artifact SHA-256 begins `6d33e9e3d68737c8` and the projection SHA-256 begins `c1f8791c1f09f71e`. Full digests, per-row identities, and values remain in the machine-readable evidence bundle.

B Development Diary

Archived source note: This marked diary is a development-only record. It is excluded from submission mode; only evidence-backed definitions, protocols, and results graduate into the thesis body.

Source: thesis mode contract

B.1 Decisions Retained

The thesis keeps the privileged oracle and actor-visible policy as separate contracts, evaluates finite candidate sets under a shared validity mask, and treats target-specific reconstruction improvement as the common comparison signal. Rollout stores retain selected transitions and provenance rather than becoming a second implementation of the oracle or training pipeline.

Open decision: Freeze the scene-disjoint analysis population and the minimum evidence scale only when the resolved manifests and exclusion ledger are available.

Source: experimental-design contract

Gate: confirmatory bundle freeze

B.2 Failures That Changed the Plan

Early rollout attempts reached the mesh-rendering path but exceeded available accelerator memory when candidate views were rendered without a bounded batch. This failure narrowed the immediate claim to implementation readiness and made peak memory, throughput, failure provenance, and completed-store validation explicit scale gates.

Validation TODO: Require completed stores, matching manifests, resource summaries, and recorded failures before extrapolating rollout bandwidth or storage demand.

Source: rollout attempt artifacts

Gate: cluster-readiness evidence

B.3 Rejected Scope

Further scene representations and learning methods remain conditional comparisons within the finite-candidate study. A representation change need not change actions or supervision. Continuous control and alternative simulator or observation regimes remain future scope and require separately declared task contracts.

Remove or rewrite before submission: Delete bridge material that cannot be tied to a concrete limitation or a matched follow-up experiment after the confirmatory analysis is complete.

Source: scope review

Gate: discussion freeze

B.4 Directional-Memory Hypothesis

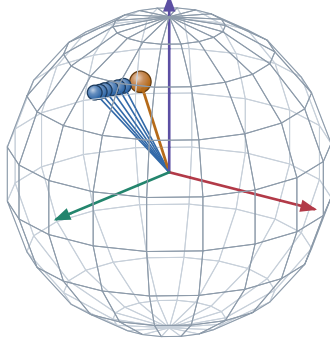
Research TODO: Treat target-centred directional memory as an ablation hypothesis, not an actor feature or contribution. The representation must be implemented, mask-audited, and compared against pose distance and overlap before any predictive claim enters the manuscript.

Source: RQ2 representation hypothesis

Gate: paired held-out ablation

A. Target-centred directions on $\mathbb{S}^2$

actual logged camera centres, expressed in the target-object frame

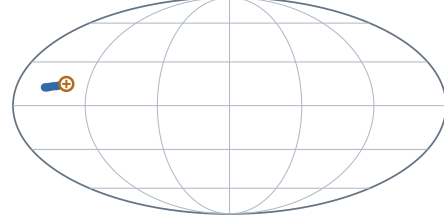


● logged history direction ● example query

Scene 81286, sample ASE_81286_Atek_000024, target instance 128 frames 0, 2, 4, 6, 8, 10 and query frame 15.

B. Complete-domain view and prospective compression

Mollweide is equal-area; no smooth field or SH lobe is implied



$+x_e$ at map centre · $+z_e$ north · wrap seam at $\pm 180^\circ$

$$\mathbf{M}_{\text{dir}(v)} = \sum_{k < t} w_{k(v)} \mathbf{d}_{k(v)} \mathbf{d}_{k(v)}^\top$$

$$\nu_{i(v)} = 1 - \frac{\mathbf{d}_i^\top \mathbf{M}_{\text{dir}} \mathbf{d}_i}{\text{tr} \mathbf{M}_{\text{dir}} + \epsilon}$$

$$\nu(\mathbf{d}) = \nu(-\mathbf{d})$$

Opposite directions have the same score. This second moment alone cannot distinguish front from back.

This fixture uses uniform weights and a logged future pose only to make the second-moment query inspectable. It does not claim that the learner currently stores $\mathbf{M}_{\text{dir}}$ or that this novelty predicts target RRI.

HYPOTHESIS / ABLATION MATERIAL. Keep outside submission claims until the representation exists and paired evaluation tests whether it adds information beyond pose distance and overlap.

Figure 35: Development-only directional-memory fixture. The sphere and Mollweide panels show the same logged ASE camera directions in a target-object frame; the orange query is another logged pose used to inspect one prospective second-moment novelty definition. No smooth field, spherical-harmonic representation, or learned benefit is claimed.

B.5 Next Evidence Gates

The next admissible evidence is a validated report bundle that fixes the study population, records target and candidate exclusions, supplies paired endpoint outcomes with uncertainty, and joins runtime and storage statistics to the same manifests. Only then can the manuscript replace the present availability statements with policy comparisons.

Implementation TODO: Export the confirmatory report bundle, compile both thesis modes against it, and inspect the resulting tables and failure cases before freezing claims.

Source: reporting contract

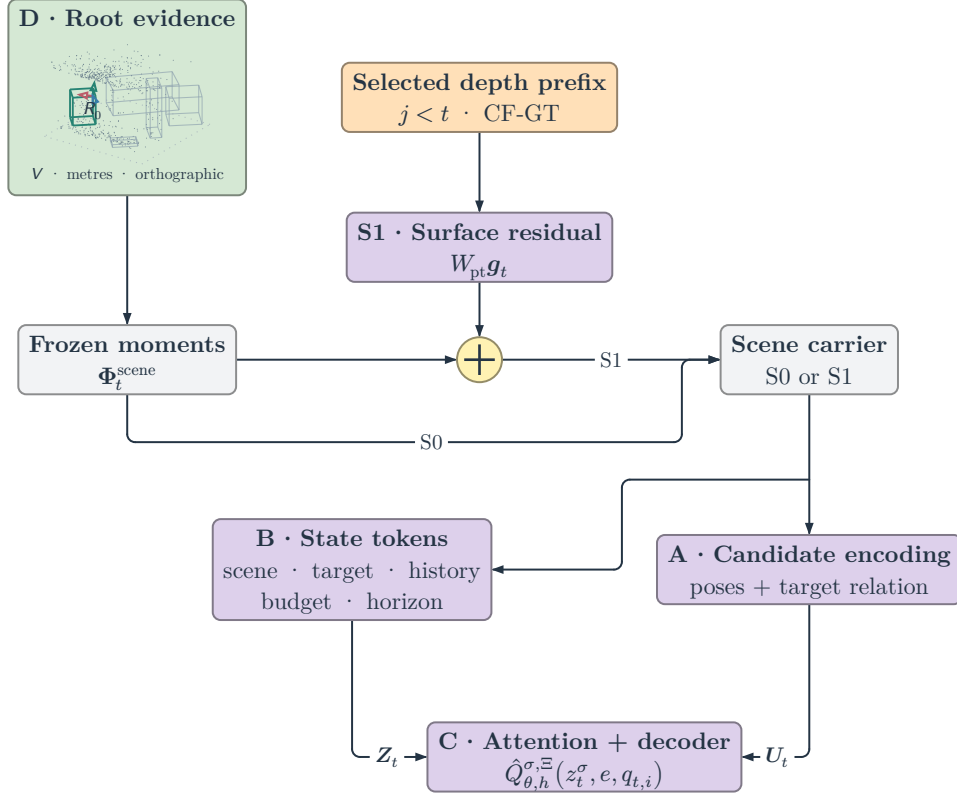
Gate: claim freeze

C Architecture Atlas

The method figures distinguish the proposed evolving-volume computation from two implemented scorer families. The state-composition figure in [Figure 9](#) is the proposal map: selected observations update source-qualified geometric evidence, a small shared encoder produces spatial features, and each target/candidate query reads the field. Weighted geometric pooling and geometry-biased attention are alternatives to compare, while a frozen EVL root neck field is the first optional pretrained arm, fused after a fixed geometric read during Q1 without requiring Q2. The existing compact and volumetric figures below describe current executable paths, not the proposed calibrated readout or a measured improvement.

Within the compact control, the figures describe different resolutions of one computation. [Figure 36](#) locates the scene evidence, candidate encoding, state-token formation and decoder within that computation. The main overview ([Figure 10](#)) supplies its principal dependencies; the expanded foldout [Figure 40](#) exposes their internal operations. Each root computation appears once; the pose encoder shares its weights across relations. Detailed views A–C and D/S1 remain beside their explanations in the Method chapter. A–D are figure-panel labels; they do not name the architectural controls A0 and A1. S1 identifies the selected-surface control where it appears in both the figure and the method.

One model, complementary levels of detail



S1 adds a privileged residual to the same frozen root evidence.
The overview composes A + B + C; D and S1 expand its scene input.

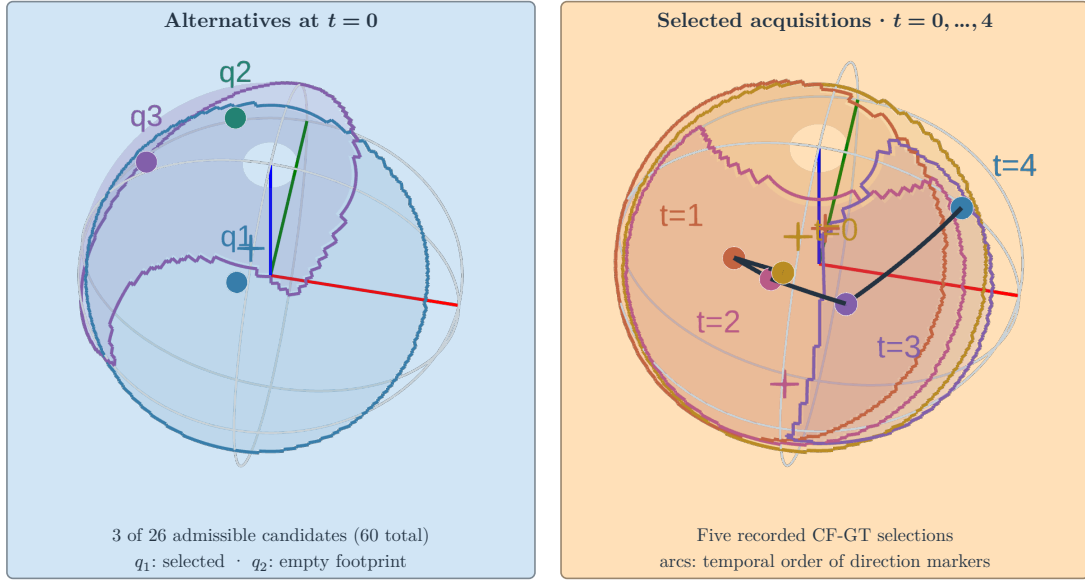
Figure 36: Interfaces between the detailed architecture views. D supplies frozen root evidence; optional S1 adds a privileged selected-surface residual. A and B consume the same resulting scene carrier, while C combines candidate queries with the five shared state tokens. S0 follows the direct frozen-root path.

The foldout [Figure 40](#) preserves physical text size rather than reducing the expanded computation to an A4 page. Tensor dimensions are supplied through the shared QH contract. The validation gate checks configured PyTorch structure and covered source identities; changes to those owners require explicit diagram review before refreshing the receipt. This mechanism detects implementation drift but does not replace scientific or visual review.

Direction normalization retains position bearing but discards standoff and camera orientation. The separate analytic spherical-harmonic study remains a VIN alternative in the gallery, rather than another view of the selected scorer.

Figure 37 contrasts alternatives at one factual state with the recorded acquisitions of the same CF-GT rollout. Its footprint calculation retains translation, rotation and rectified intrinsics. An admissible action can still have an empty target-proxy footprint, as candidate q_2 illustrates.

Candidate views and the selected counterfactual sequence



$$\mathbf{x}^e = r_e \mathbf{d}, \quad \mathbf{d} \in \mathbb{S}^2 \rightarrow \pi_{\kappa} \left(T_{E_e}^{c_{t,i}} \mathbf{x}^e \right)$$

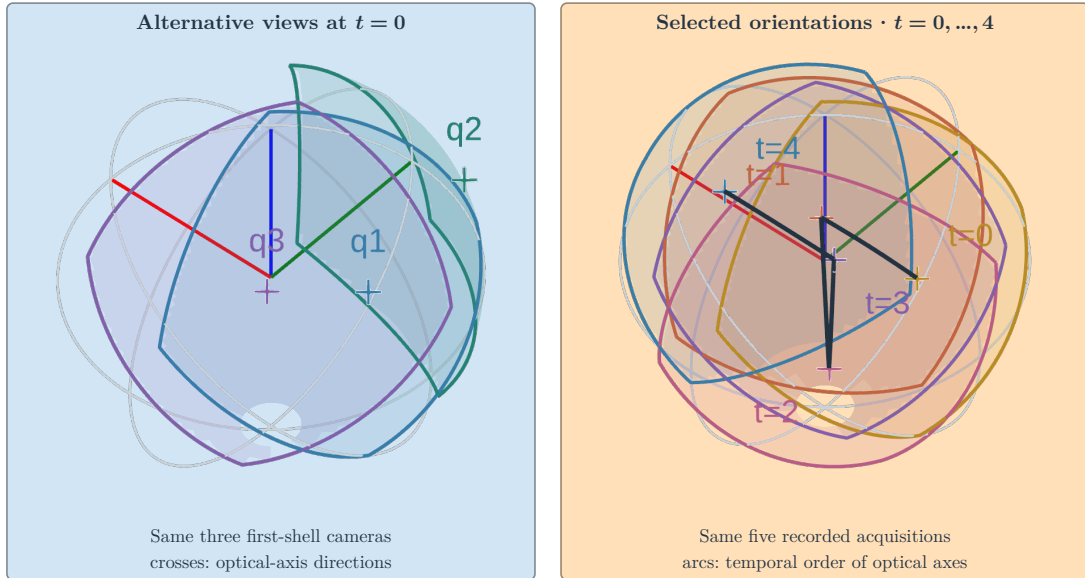
Target frame E_e · RGB axes · same projection in both panels. Coloured patches and outlines: calibrated, front-facing proxy-sphere footprint; no scene occlusion test. Discs: normalized camera positions, not focal points. Crosses: optical-axis hits, when present.

Target-centred footprints need not be symmetric about either marker. ASE 82254/000008, stored five-step temperature-softmax sequence ($\tau = 0.5$). This is potential geometric coverage, not mesh coverage or learned-policy evidence.

Figure 37: Candidate alternatives and ordered selected views, using stored ASE 82254/000008 poses and calibration. Three spatially separated admissible candidates are displayed from the first shell, including its selected candidate. On the right, colours and time indices identify five recorded acquisitions; arcs indicate temporal order of normalized camera directions, not executed paths. Discs mark normalized camera positions, and crosses mark optical-axis intersections where the principal ray hits the sphere. These target-centred footprints need not be symmetric about the position markers. Footprints mark calibrated potential visibility on a target-centred proxy sphere. Both panels use the same frame and camera.

The camera-ray sphere in Figure 38 represents a different quantity from Figure 37. Each camera is translated to the sphere centre, and its calibrated rays are rotated into the target basis. The patches therefore describe orientation and field of view, not target-surface visibility. With a centred principal point, the angular footprint is reflection-symmetric about its optical axis; the stored calibration has only a small principal-point offset. The oblique projection can still make its outline appear asymmetric. Candidate q_2 has a valid field of view here even though its target-proxy footprint is empty in the surface view.

Camera-ray sphere · orientation and field of view



$$\hat{d}^c = R_{ec} \frac{\pi_{\kappa}^{-1}(u, 1)}{\|\pi_{\kappa}^{-1}(u, 1)\|} \in \mathbb{S}^2$$

Each camera origin is placed at the sphere centre; directions are expressed in the target basis. Patches: calibrated rectangular pinhole fields of view. Crosses: optical axes, not camera positions. Camera translation and target visibility are discarded. Intrinsic offsets are small, so each patch is approximately angularly symmetric about its cross; an oblique 2D view need not look symmetric. ASE 82254/000008 · same calibration, candidates, sequence and colours as the target-surface view.

Figure 38: Orientation-only spherical projection of the same stored views. Each camera origin coincides with the sphere centre; patches retain calibrated ray directions in the target basis and discard translation. Crosses identify optical-axis directions. Colours preserve candidate or temporal identity from the target-surface view. Approximate angular symmetry follows the nearly centred calibration, not target coverage or learned performance.

The implemented diagnostic-volume detail in Figure 39 expands the two spatial stages of Figure 15. The fine-scale residual stream is concatenated with upsampled coarse context. Sampling then retains deterministic evidence alongside the learned field rather than asking learned features to encode source availability implicitly.

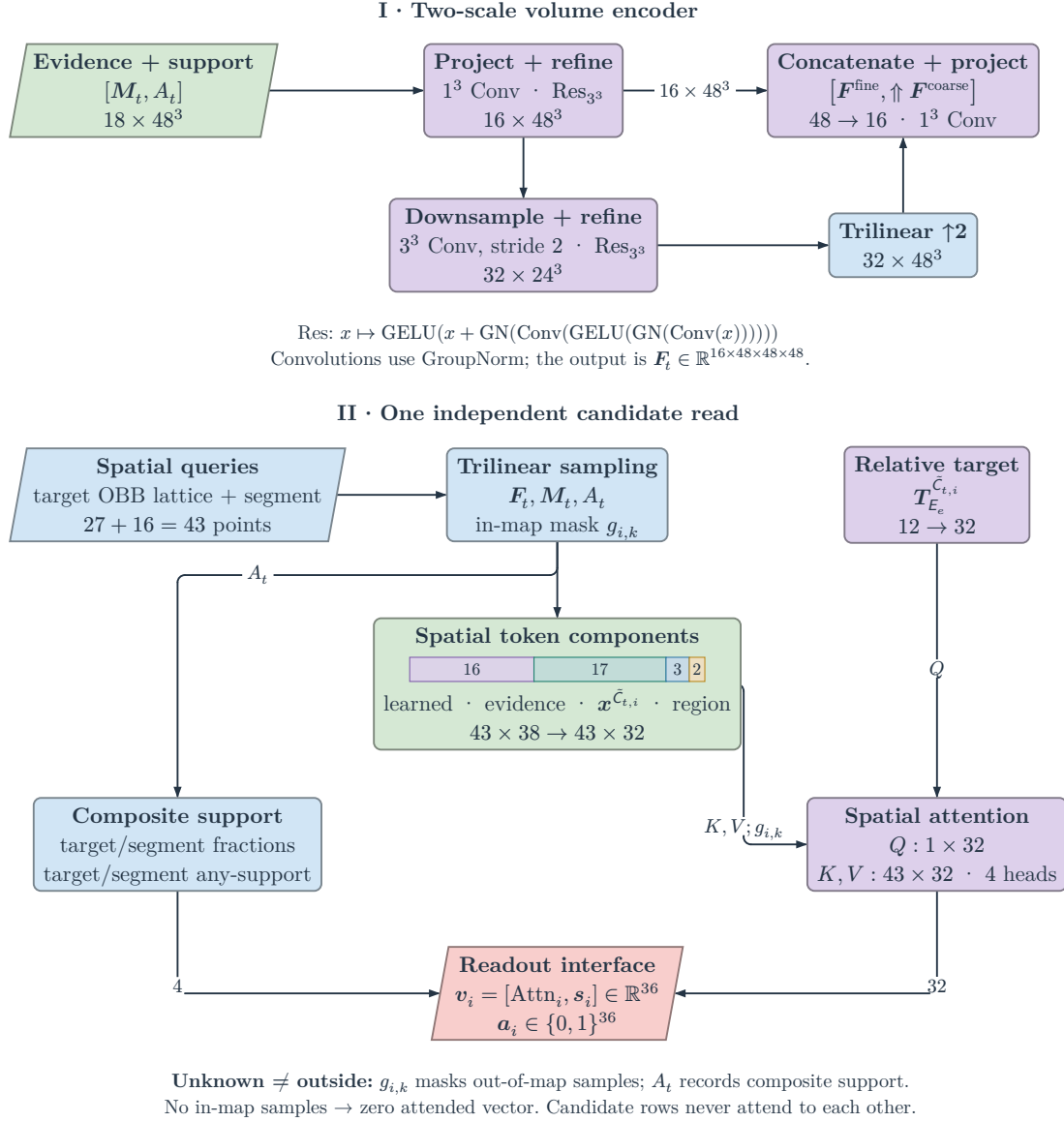
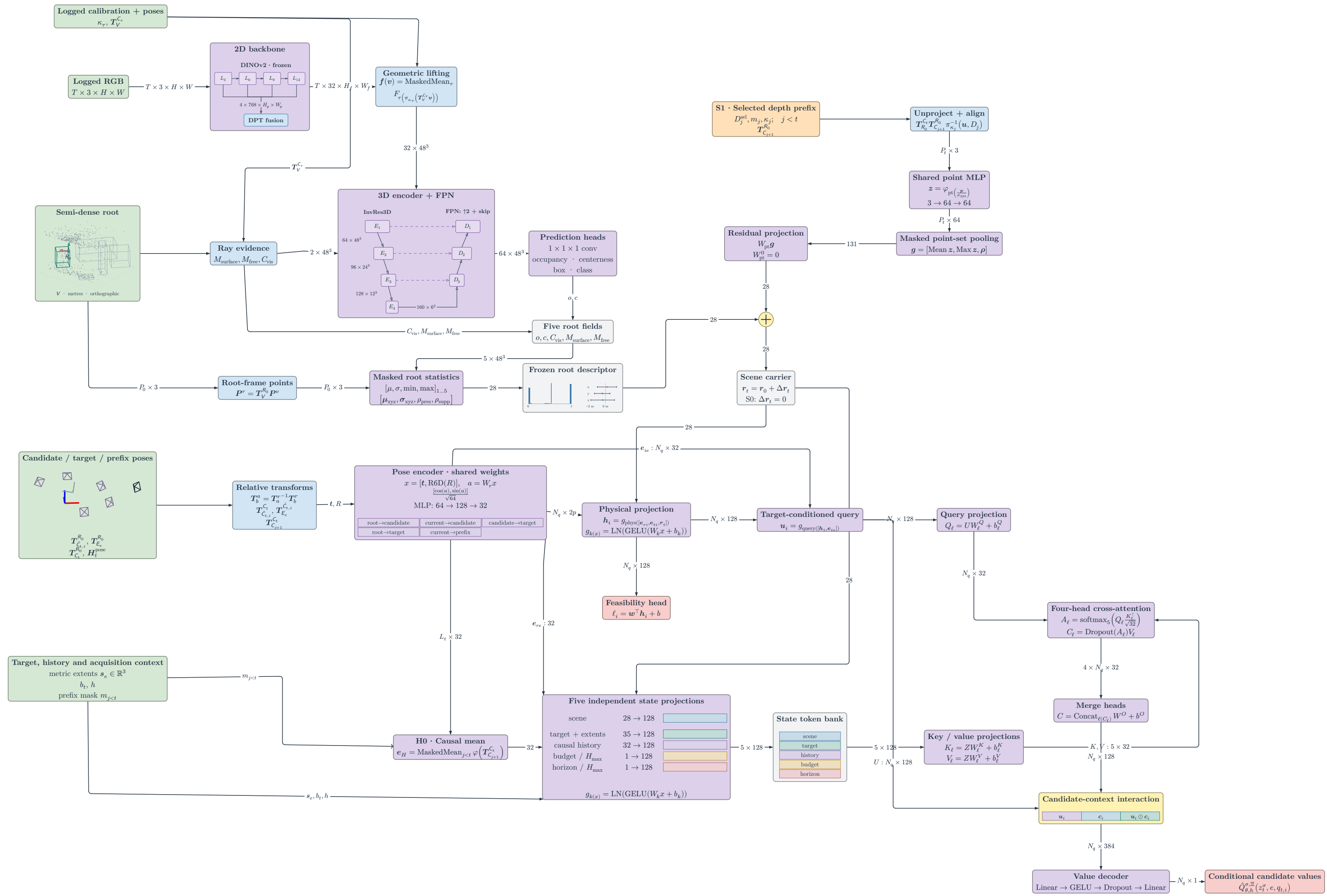


Figure 39: Volumetric encoder and candidate-local readout. Tensor dimensions omit batch and state axes. The token bar is proportional to input channel counts, with region encoded by two indicator channels; projection maps the 38 inputs to width 32. In-map support $g_{i,k}$ determines attention admission, while the current supplied composite support mask A_t supplies four target-lattice and segment statistics that inherit the root projection-count conflation until the evidence-semantic patch. An entirely out-of-map read yields a zero attended vector. The 36 features and their availability bits enter the value-query interface in Figure 15. All tensors and channel proportions are schematic configured structure.

QH · From logged observations to target-conditioned candidate values

One connected computation graph · D / S1 / A / B / C · default scorer widths; batch and factual-step axes suppressed



Reading the graph
Green: observations and geometry inputs. Blue: explicit geometry. Purple: learned maps. Orange: privileged S1 input. Pink: predictions. $p = 32$, $d = 128$; shapes annotate edges. Pose relations occupy separate streams with shared encoder weights.

Masks and downstream use
The candidate materialization mask zeros physical/query tokens and both outputs. Authoritative action admissibility and action selection remain outside the scorer. Feasibility is an auxiliary output; it does not silently replace the action mask. Candidate rows never become attention keys or values.

Evidence and scope
Root inset and statistics: stored ASE 83269/000056. Camera shell: seeded, unfiltered reconstruction, not a measured rollout. EVL topology: current RGB configuration. S1 is an optional selected-surface control. SH belongs to a separate VIN alternative and is not on this QH path.

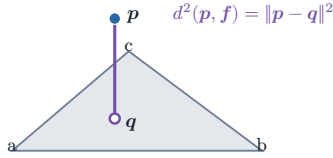
Figure 40: Expanded QH computation with shared operations shown once. Colour encodes observation, geometry, learned-map, privileged-input and prediction roles. Distinct pose relations share encoder weights, while the five state maps have independent weights. Tensor widths correspond to the checked default scorer. Root insets are measured; the camera shell is reconstructed. Candidate materialization masks apply inside the scorer, while authoritative action masking and selection remain external.

C.1 Point–mesh metric validity

The controlled fixture isolates directional reductions and the dependence of equal-face completeness on tessellation. It complements the measured retained-depth example in Figure 5. In panel C, subdividing one part of the mesh gives that region more weight in the equal-face average. The surface and reconstructed points are unchanged, yet the measured error decreases. The change therefore reflects mesh weighting rather than better reconstruction. Comparisons must retain the same reference mesh as well as the same point-preparation procedure.

A. Exact point–triangle primitive

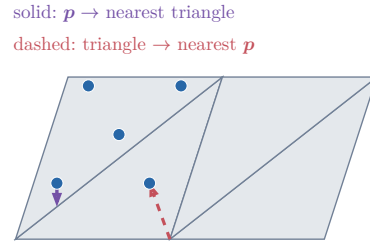
the closest point may lie on the face, an edge, or a vertex



The backend evaluates squared Euclidean distance to the triangle itself, not distance to its supporting plane or centroid.

B. Two directional reductions

the arrows reverse which population contributes summands



One exact witness is shown per direction; the implementation reduces over every point or every face.

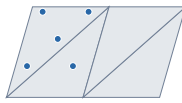
C. Controlled tessellation check

same planar surface + same points; only tessellation changes

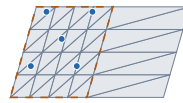
4 equal-area

40 non-uniform

left region: 32/40 equal-weight faces



$$D_{M \rightarrow P} = 0.03640 \text{ m}^2$$



$$D_{M \rightarrow P} = 0.02284 \text{ m}^2$$

The left region contributes 32 of 40 equally weighted face summands.

Controlled conclusion. $D_{P \rightarrow M} = 0.00640 \text{ m}^2$ in both; $D_{M \rightarrow P}$: $0.03640 \rightarrow 0.02284 \text{ m}^2$ after retessellation.

Figure 41: Controlled point–mesh validity fixture computed with the repository metric and exact witnesses. The construction isolates the asymmetric reductions and equal-face tessellation sensitivity while holding surface support and points fixed. It provides synthetic mechanism evidence, not an ASE performance result.

C.2 Recorded Replay Compatibility

Input and target	Recorded support	Compatibility and proof
V10, privileged CF+, GT target	legacy/subset labels; selected GT depth and successor links	Legacy and H1 batch admission measured on four old/new V10 chains; dense-valid objective rejected.
V11, privileged CF+, GT target	actor-only roots; target replay labels, selected GT depth and successors	Five real H16 joins validated; legacy/subset contract retained. Actor-only preparation adds no dense-valid admission.
Required observed CF0	observed target and actor-permitted inputs, with objective-specific labels and transitions	Recorded GT-target cohorts rejected by the observed-target requirement. Substrate validation alone is insufficient.

Table 23: Observed compatibility depends on actor inputs, target provenance and objective-specific label support. Partial labels do not certify dense supervision over all physically valid actions.

The current model reader accepts qualified V10 and V11 actor roots and checks their compact visual evidence and source identities. The cited measurements establish a mixed old/new V10 batch load and separate V11 joins. They do not establish a combined V10/V11 training batch or a model fit. Reader capability, measured batch integration, and scientific corpus admission are separate claims.

C.3 Generation re-analysis contract correspondence

The post-hoc product retains a V10 root contract with a 48^3 voxel grid and the cropped-rendering R1 acquisition control, sixty candidates and a maximum recording length of sixteen acquisitions. Full-scene selected observations remain the causal successor input. Allocation, actor validity, oracle label availability, training support, selection and continuation are separate predicates. The compact report preserves cohort, stage, physical-scene and qualified chain/state identities, retaining overlapping scene exposure and missing continuations.

Correspondence	Contract / runtime role	Admission limit
Root input	V10; 48 ³ voxel grid	Root context is separate from rollout and scene denominators
Acquisition	R1 cropped candidates; sixty proposals; maximum sixteen recorded acquisitions; full-scene selected observations	Recording length does not complete every future query
Recipe	18 forward / 18 bearing / 12 bypass / 12 orbit; selector temperature 1; zero heading weight and score floor	Two frozen nominations remain unapplied, separate tests
Mac runtime	EVL/MPS roots; CPU/PyTorch3D-Mojo rollout lane	One census diagnostic; no host-speedup or general numerical-equivalence claim
CUDA runtime	Generation for incremental TRAIN acquisition	Overlapping scenes; not a matched host-treatment arm
Mixed admission	Preserve source-specific root and rollout identities	Reader execution and individual checks do not qualify a mixed-root training corpus

Table 24: Protocol and runtime correspondence for the bounded generation re-analysis. Runtime roles do not establish interchangeable numerical results or joint corpus admission.

The separate Mac V11/64³ readiness result remains outside this V10/48³ census. The single-source V10/48³ root comparison supports a pilot observation only; it establishes neither joint corpus admission nor R1 rollout equality.

C.4 Replay lineage and stored relations

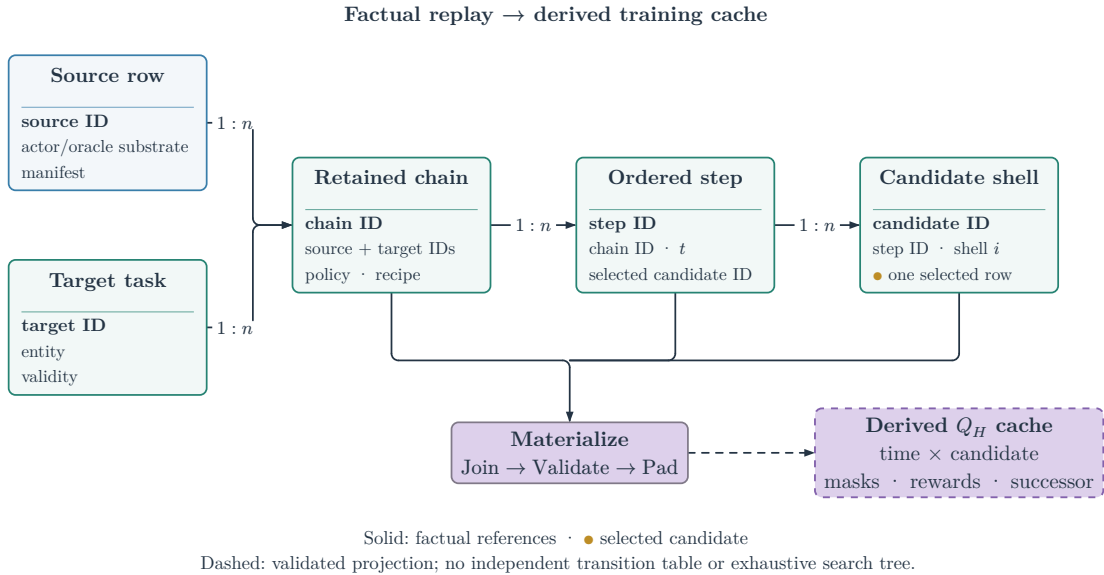


Figure 42: Persisted replay lineage and repository-validated Q_H materialization. Solid links are persisted row references or ownership relations; the selected marker identifies one member of each full candidate shell.

Stored relation	Mathematical role	Scientific meaning
Source–target	s_0, e	Binds a task to one immutable logged substrate and one declared target source.
State–candidate	$s_t, q_{t,i}$	Preserves the full proposal support, hard feasibility, label availability, and proposal provenance.
Selected transition	$(s_t, a_t, r_t^e, s_{t+1})$	Records the causal edge created by the selected action.
Retained chain	$(s_0, a_0, \dots, s_H)$	Identifies the recipe-specific trajectory from which finite-horizon returns may be derived.
Derived training view	$(\mathbf{X}, \mathbf{m}, \mathbf{y})$	Pads and batches factual relations without changing their information or missingness semantics.

Table 25: Scientific roles of the normalized replay relations. State, target, action support, supervision, and lineage are bound before model-specific tensorization.

C.5 Replay batching and backup validation

C.5.1 Training unit, causal batching, and prediction support

A retained rollout chain supplies several realized decision states, not one monolithic recurrent sample. The root is one state. Each selected action produces a factual successor with an updated current pose, selected prefix, remaining budget, and candidate table. The batching procedure may place several such states in one padded tensor batch, but the prediction at state t uses only the causal prefix available before action a_t . Later states in the same batch cannot contribute their poses, observations, or candidate tables to that prediction.

The scorer evaluates these states and their candidate rows in parallel. For one scalar horizon query per state, it emits one conditional value for every materialized candidate. Parallel batching does not introduce recurrence or temporal attention: the scorer is invoked again only after the selected action has produced a new causal state and candidate table. A future encoder that jointly processes several states would require causal masking across that axis; the present independently materialized state representation does not.

Prediction support and loss support therefore differ. The scorer may emit a value for every materialized candidate, whereas dense one-step loss uses only

hard-valid candidates with finite immediate labels. For $h > 1$, recursive loss uses the selected factual transition; padding creates no multi-step supervision for unselected actions. Admission is profile-specific. The legacy countdown profile can retain a missing-successor row as an immediate-only target. The continuing profile requires an actual next-state record and finite, hard-valid successor reward labels, excluding rows without them from recursive loss. Neither missing records nor empty stored action sets establish a real task endpoint. Whether the actor stops or regenerates actions at an empty set remains a protocol choice. Method records the remaining boundary and budget–horizon admission decisions.

The commissioning mode uses only the dense $h = 1$ surface, including later realized states whose causal volumes already contain selected observations. It retains the factual remaining budget, the complete causal prefix, and the corpus maximum horizon; setting every budget to one would change the represented state rather than isolate the objective. Recursive targets and delayed-network evaluation are omitted in this mode. The implemented pure-Q1 diagnostic also omits feasibility supervision. Joint immediate-value and physical-validity commissioning therefore requires a separate explicit objective profile and its admission checks; the existing recursive objective does not isolate that experiment.

C.5.2 Scalar requested-horizon targets

The learning protocol uses the immediate and recursive targets defined in [Section 4.9.5](#). The remaining budget b_t belongs to the factual state; the requested horizon h determines the return being predicted. A shorter query $h < b_t$ is admitted only with an explicitly constructed truncated target. In the legacy countdown profile, the saved terminal flag or an empty actor-valid successor set gives zero continuation; actor-valid rows without finite reward-label support are excluded. In the continuing profile, a recording cutoff does not complete a prediction query: recursive supervision requires a recorded successor with finite hard-valid reward-label support. That state is queried at $h - 1$, independently of its scheduled query; missing or unsupported successors give no recursive loss. The lower-horizon

prediction is held fixed by stop-gradient, a frozen stage checkpoint, or a delayed target copy.

The joint objective fits one shared horizon-conditioned network on two supported query domains. Immediate supervision queries $h = 1$ at every realized state and uses each finite hard-valid candidate reward. Recursive supervision queries the factual remaining horizon $h = b_t$ and uses only selected transitions with $h > 1$, selecting the successor action online and evaluating it with delayed parameters at $h - 1$. The loss balances the query domains by factual state and horizon using the aggregation in [Section 5.2.3](#); it does not change their factual support.

The objective does not enumerate off-diagonal pairs $1 < h < b_t$. Admitting more state–horizon pairs requires explicitly constructed targets and a frozen weighting rule. Transaction batching is an implementation choice; staged backward induction, separate heads and broader horizon enumeration remain unselected alternatives in the development register.

Method defines the exact two-step target and its complete-successor-label requirement ([Section 4.9.6](#)). The executable check has two layers: exact-table injection tests backup arithmetic; the frozen-bundle check compares the learned backup with that factual reference using

$$\varepsilon_t^{(2)} = \left| y_t^{(2,\text{recursive})} - y_t^{(2,\text{exact})} \right|, \quad \varepsilon_t^{(2)} \leq \tau_{\text{abs}} + \tau_{\text{rel}} \left| y_t^{(2,\text{exact})} \right| \quad (73)$$

The latter is evaluated on a deterministic bounded sample drawn from a meta-data-only census. Sampling strata record scene, target row, configured horizon, candidate-width bin, candidate and rollout configurations, and behavior policy. Within each stratum, sampling prefers a previously unseen scene before another chain from an already represented one. This improves scene coverage without treating multiple rows from one scene as independent replications.

The exact two-step validation fixes the model, learning and actor-state contracts, ordered test-store provenance, sampling bounds and error tolerances. It checks agreement of the learned backup target with the factual exact control; an executable check alone does not establish a qualifying population result. Its

groups combine store provenance and scene identity: exporting one physical scene under two manifests can create two groups. Confirmatory splits and aggregation must therefore follow physical scene identity across stores. Cross-store physical-scene deduplication remains unverified here. The current promotion check requires at least five selected provenance groups as a minimum-support gate, at least one factual selected-action exact Q_2 row in every group, and the same error gate to pass in every group. Passing this check alone does not establish five independent scenes, statistical power or generality; scenes from one dataset or physical environment may remain correlated. Pooled row-level error remains diagnostic and cannot compensate for a failing group. The validation record preserves the denominator from eligible chains through materialized successors and complete immediate labels to factual exact two-step rows. Per-stratum support and ordinal-decoder saturation remain separate from recursion error in that record. Missing support is a failed gate rather than zero error. Receipts using the earlier development schema remain readable historical evidence but cannot promote a claim. Agreement with the exact control checks the fitted backup target, not learned Q_2 prediction accuracy or policy benefit. Longer-horizon policy effects require independent held-out endpoint evaluation; positive oracle-lookahead headroom is not a prerequisite for that comparison. A persisted terminal-step role contrast is retained as a diagnostic proxy and is manifest-bound, but it cannot promote an $h > 2$ claim.

C.6 Expanded replay update and seed lineage

The compact relation identifies the implemented S0 information-state transition. Its stored component updates are

$$\begin{aligned}
\mathbf{T}_{C_{t+1}}^{R_0} &\leftarrow q_{t,a_t}, \\
\mathbf{H}_{t+1}^{\text{pose}} &\leftarrow \left(\mathbf{H}_t^{\text{pose}}, q_{t,a_t} \right), \\
b_{t+1} &\leftarrow b_t - 1, \\
(\mathcal{Q}_{t+1}, \mathbf{m}_{t+1}^{\text{act}}) &\leftarrow g_{\Xi}(q_{t,a_t}, \mathbf{H}_{t+1}^{\text{pose}}, e)
\end{aligned} \tag{74}$$

Immutable root evidence and target identity pass through unchanged. The selected endpoint becomes the realized camera pose, is appended to the strictly causal pose prefix, and consumes one acquisition. The seeded generator g_{Ξ} then regenerates both the next candidate table and its hard action-mask vector $\mathbf{m}_{t+1}^{\text{act}}$ from the selected endpoint, updated prefix, fixed target e , and the immutable scene, calibration, and generator settings carried by Ξ . Its proposal seed is derived from the selected full-shell index path, so no additional free variable ξ_t is required. Reordering retained trajectories preserves that lineage; reindexing the candidate shell need not preserve the seed even if the selected physical pose is unchanged. This is distinct from scorer equivariance on a fixed shell.

C.7 Measurement and evaluation details

The following records preserve exact constructions and historical controls referenced by the main argument. They are reproducibility details, not additional required comparisons.

C.8 Applied sampler construction

Each row of the three local families begins with a random unit direction in the gravity-aligned sampling frame $\mathcal{S}$, whose origin is the reference rig position and whose forward axis follows its horizontal heading. The frozen profile uses a forward-biased Power Spherical distribution with $\kappa = 8$ [49]:

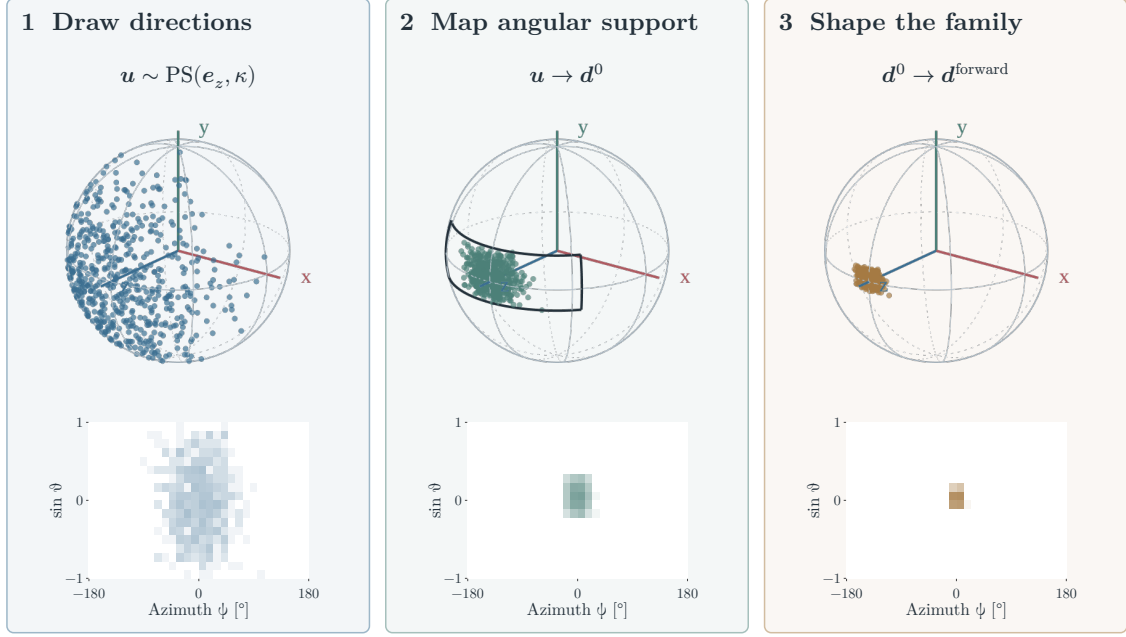
$$\mathbf{u}_i \sim \text{PS}(\mathbf{e}_z, \kappa), \quad p(\mathbf{u}) = c_{\kappa}(1 + \mathbf{e}_z^T \mathbf{u})^{\kappa}, \quad \kappa = 8 \quad (75)$$

The sampled direction is then mapped into the configured azimuth and elevation caps. With $\psi = \text{atan2}(u_x, u_z)$ and $u_y = \sin \theta$, the transform is

$$\psi' = \psi \frac{\Delta_{\psi}}{2\pi}, \quad y' = \sin \theta_{\min} + \frac{u_y + 1}{2} \cdot (\sin \theta_{\max} - \sin \theta_{\min}) \quad (76)$$

$$\mathbf{d}_i^0 = \left(\sqrt{1 - y'^2} \sin \psi', y', \sqrt{1 - y'^2} \cos \psi' \right), \quad \|\mathbf{d}_i^0\|_2 = 1 \quad (77)$$

Figure 43 follows the same draws through both transformations. The angular bounds constrain the intermediate directions; the family map determines the final directional support.



Same 600 draws · $\kappa = 8$ · gravity-aligned S

$$\psi' = \psi \frac{\Delta_\psi}{2\pi}, \quad y' = \sin \theta_{\min} + \frac{u_y + 1}{2} \cdot (\sin \theta_{\max} - \sin \theta_{\min})$$

$\Delta_\psi = 100^\circ$, $\theta \in [-12^\circ, 18^\circ]$ before family shaping.

Lower panels: empirical frequency per equal-solid-angle bin; logarithmic colour, 10^{-3} –1.

Figure 43: From Power Spherical sampling to forward-local proposals. The 600 seeded draws use the actual position sampler. Spherical panels show direction; lower panels show empirical frequency in equal-solid-angle bins on a common logarithmic scale, floored at 10^{-3} . Radius is sampled subsequently.

The cap deterministically remaps every draw and therefore changes angular support without changing the row count. This is not rejection sampling from the original distribution conditioned on the configured caps. Family-specific maps then reinterpret the same bounded random variable relative to rig forward, the target bearing, or the lateral direction. This factorization separates **diversity within a family** from the **geometric meaning of the family**.

Let $\mathbf{f} = \mathbf{e}_z$ denote sampling-frame forward, $\mathbf{b}_e$ the supplied target bearing, $\mathbf{l}_e$ its normalized horizontal lateral direction, and $\mathbf{e}_y$ world up in the sampling frame. The three core maps are

$$\begin{aligned}
\mathbf{g}_i^{\text{forward}} &= \mathbf{f} + \alpha_f(\mathbf{d}_i^0 - (\mathbf{d}_i^0 \cdot \mathbf{f})\mathbf{f}), \quad \mathbf{d}_i^{\text{forward}} = \frac{\mathbf{g}_i^{\text{forward}}}{\|\mathbf{g}_i^{\text{forward}}\|_2}, \quad \alpha_f = 0.45 \\
\mathbf{g}_i^{\text{target}} &= \mathbf{b}_e + \alpha_t(\mathbf{d}_i^0 - (\mathbf{d}_i^0 \cdot \mathbf{b}_e)\mathbf{b}_e), \quad \mathbf{d}_i^{\text{target}} = \frac{\mathbf{g}_i^{\text{target}}}{\|\mathbf{g}_i^{\text{target}}\|_2}, \quad \alpha_t = 0.4 \\
\mathbf{g}_i^{\text{bypass}} &= 0.55\mathbf{b}_e + 0.85 \text{sign}(\mathbf{d}_{i,x}^0)\mathbf{l}_e + \text{clip}(\mathbf{d}_{i,y}^0, -0.35, 0.35)\mathbf{e}_y, \\
\mathbf{d}_i^{\text{bypass}} &= \frac{\mathbf{g}_i^{\text{bypass}}}{\|\mathbf{g}_i^{\text{bypass}}\|_2}
\end{aligned} \tag{78}$$

The forward and target maps remove the component parallel to their main direction, then scale the remaining transverse perturbation around that direction. The bypass sign chooses a lateral side; clipping limits its vertical perturbation, and normalization gives a unit direction before radius selection.

A family-conditioned radius turns direction into translation; the reference sampling pose then maps that offset into world coordinates:

$$r_i \sim \mathcal{U}(r_{\min}^{k(i)}, r_{\max}^{k(i)}), \quad \mathbf{c}_i^W = \mathbf{T}_S^W(r_i \mathbf{d}_i^{k(i)}) \tag{79}$$

C.8.1 Target population and sampling priority

A confidently detected cabinet can still have an unobserved side. Low confidence, by contrast, may reflect an ambiguous object estimate rather than a need for further reconstruction. Detection reliability and reconstruction need are different quantities. Admission determines which objects belong to the study; sampling determines how often those objects contribute training examples. The value model then predicts the benefit of a view for the specified object. Goal-conditioned value functions make this last relationship explicit by conditioning on the supplied goal [18].

For a nonempty ground-truth pool, selecting $k = \min(K, N)$ of N eligible targets with cap K gives each target inclusion probability $\frac{k}{N}$. For example, selecting three of ten targets gives each a $\frac{3}{10}$ chance of appearing in the sample, although its chance of being drawn first is only $\frac{1}{10}$. This is the difference between

inclusion probability and **draw probability**. Here N counts geometry-valid reference objects for the legacy control and support-eligible objects for the applied development collection. The same uniform-subset rule applies after either population has been defined.

Implementation correspondence. The sampler implements the subset through a random permutation and stores the marginal inclusion probability in `selection_probability`. The current-root support receipt retains the pre-sampling population and its exclusions.

The observed route includes every uniquely matched descriptor at task construction. Subsequent execution limits and failures are recorded separately: being selected for generation is not the same as producing a usable rollout. Neither construction gives uniform exposure to unique objects or scenes, because an object can recur in overlapping snippets. Evaluation must therefore specify its scene, root-context, and target populations. Eligibility is fixed before gains are measured, and the target identity and root normalization stay fixed during each rollout.

The implemented sampler uses the uniform rule above. Weighted target exposure is an unselected alternative discussed in the development alternatives register; its sampling weights would require a separate inclusion-probability definition.

Confidence can indicate detection reliability, subject to calibration; the current detector’s calibration remains unestablished. Image area and observed surface coverage describe acquired support. None directly measures remaining reconstruction need. Ground-truth IoU can support offline association or sampling diagnostics, but cannot enter an observation-only selection rule. The detailed calibration and optional allocation alternatives are retained in the development register; the implemented admission is specified in [Section 3.3.3](#).

C.8.2 Sibling-heading guard

Sibling diversity uses the camera-forward direction projected onto the world xy plane, with world $+z$ as up. A nearly vertical view has no stable horizontal heading: it may be retained as the first sibling, but cannot support further heading-

constrained branches. Stochastic fallback may relax angular separation, but does not waive heading availability; fewer siblings than requested may therefore be retained when the yaw guard is enabled.

Normalization uses the original finite score vector; the sampling probabilities are recomputed over the remaining eligible rows after each draw.

C.9 Historical pruning controls

The earlier control used

$$\|\mathbf{o}_i\|_2 \leq 1.0 \text{ m}, \quad |\Delta h_i| \leq 0.25 \text{ m}, \quad \max(0, -o_{i,z}) \leq 0.25 \text{ m}, \quad \Delta\psi_i \leq 70^\circ$$

The applied profile retains a maximum step of 1 m and a maximum absolute height change of 0.25 m, together with the snippet, mesh-clearance and path checks. It disables the optional backward-motion and heading-change caps; the displayed 0.25 m backward and 70° heading limits describe the earlier control. This relaxation changes action support without establishing wearer-motion realism.

Here $\mathbf{o}_i$ is the reference-to-candidate translation expressed in the reference-pose frame, and Δh_i is the height change along world up. The term $\max(0, -o_{i,z})$ measures backward motion along reference forward. The angle $\Delta\psi_i$ compares the horizontal camera headings of the sampling orientation and the candidate; when a horizontal direction is undefined, the implementation compares their full forward axes instead.

The earlier screen required at least 15 feasible rows in its 60-row shell:

$$N_{\text{valid}} \geq \max(12, \text{ceil}(0.25N_q)) \quad (81)$$

This threshold is historical; the applied collection instead requires three hard-valid rows across bearing, bypass and orbit, without a per-family minimum.

C.10 Shared label and commit procedure

Shared label and commit procedure

1. Build the fixed canonical root once; retain every selected observation as its full acquired, strided world cloud.
2. At step t , concatenate the root and selected history to obtain the current observation multiset.
3. For each feasible candidate, render against the full scene, mask and backproject its depth, then append its full cloud to a copy of that multiset.
4. Apply Φ_e to the current and candidate-appended multisets; compute their errors against the same cropped target mesh.
5. Store a finite label only when both support and measurement are valid. After selection, append that exact acquired cloud to history and rebuild through the same Φ_e .

Figure 44: Conceptual pseudocode for transition-consistent oracle supervision with default full-scene rendering. “Append” means row concatenation with multiplicity, not set union or independent source-wise averaging.

C.10.1 Ordered fusion and point capping

For reproducibility, samples in each sorted voxel group are reduced in their retained input order before division by n_v . Identical ordered inputs are reproducible in the pinned runtime; this does not imply invariance to input permutation or bit equality across devices and library versions.

After fusion, an optional cap acts on the ordered representatives. For $K > N$, the implementation keeps the deterministic indices $\lfloor j \frac{K}{N} \rfloor$, $j = 0, \dots, N - 1$:

$$\mathcal{K}_N \left((\mathbf{p}_k)_{k=0}^{K-1} \right) = \begin{cases} (\mathbf{p}_k)_{k=0}^{K-1} & \text{if } K \leq N \\ (\mathbf{p}_{\lfloor j \frac{K}{N} \rfloor})_{j=0}^{N-1} & \text{if } K > N \end{cases} \quad (82)$$

This is deterministic thinning. It is neither random sampling nor an optimization of geometric coverage. The ordered representative sequence is therefore result-affecting; after voxel fusion it follows the sorted voxel-key order, whereas with voxel fusion disabled it inherits point-row order. A mask or crop can exclude evidence; a cap can retain only part of the eligible evidence. The cap reserves no capacity for the newest observation and can discard all representatives contributed by it.

These operations must remain visible in the measurement definition rather than being summarized as generic “downsampling.”

C.11 Alternative reconstruction measurements

Metric family	Question answered	Dependency and thesis role
Equal-face bidirectional point-mesh error	Are reconstructed points close to GT faces, and does every GT face have nearby point evidence?	Depends on point fusion and mesh tessellation; selected operational RRI error.
Point-sampled Chamfer / accuracy-completeness	Are two sampled surface point sets mutually close?	Depends on sampling density, randomness, norm, and squaring convention; alternative operationalization.
Thresholded precision, recall, and F-score	What fraction of predicted and reference samples fall within tolerance τ , and how do the two fractions balance?	Interpretable at a physical tolerance but threshold-dependent; secondary diagnostic alongside scalar RRI.
TSDF, occupancy, coverage, or information gain	How much volumetric surface, free space, or previously unknown space is reconstructed or observed?	Depends on grid, truncation, and visibility definitions; answers a different question from target surface quality.

Table 26: Reconstruction measures answer different geometric questions. The equal-face point-mesh error is the selected outcome; the other families are diagnostics or alternative operationalizations.

Point-sampled Chamfer distance draws a finite reference set Q_e from the mesh and averages nearest-neighbour distances in both directions:

$$D_{\text{PS-Chamfer}}(\mathcal{P}, \mathcal{Q}) = \frac{1}{\|\mathcal{P}\|} \sum_{\mathbf{p} \in \mathcal{P}} \min_{\mathbf{q} \in \mathcal{Q}} \|\mathbf{p} - \mathbf{q}\|_2 + \frac{1}{\|\mathcal{Q}\|} \sum_{\mathbf{q} \in \mathcal{Q}} \min_{\mathbf{p} \in \mathcal{P}} \|\mathbf{q} - \mathbf{p}\|_2 \quad (83)$$

Sampling density, randomness, norm, and squaring convention determine its finite value despite the symmetric formula. Thresholded precision, recall, and F-score add a physically interpretable tolerance τ , but discard error magnitude on either side of that threshold. Volumetric coverage and information gain measure observed space or uncertainty reduction. These are useful complementary quantities. Each operationalization maps a different geometric construct.

C.12 Secondary proportional-error diagnostic

The log-gain variant

$$J_{e,\log}^{(H)} = \log(\Delta_0^e + \varepsilon_{\text{err}}) - \log(\Delta_H^e + \varepsilon_{\text{err}}) \quad (84)$$

expresses the proportional reduction from initial to final reconstruction error. Its interpretation near zero error depends on the stabilizer ε_{err} . $\text{RRI}_{t,i}^e$ describes local improvement, $G_{t,e}^{(h)}$ describes accumulated discounted reward, and $J_e^{(H)}$ measures the fixed-budget outcome. Endpoint gain and undiscounted cumulative root gain agree at horizon H because both use $\Delta_0^e + \varepsilon_{\text{err}}$. State-relative RRI instead uses $\Delta_t^e + \varepsilon_{\text{err}}$, which changes along the rollout. Each policy comparison fixes the discount, clipping, crop, point update, and evaluation geometry.

C.13 Secondary proposal geometry diagnostics

For attempted target-conditioned rows, side balance

$$\begin{aligned} n_+(s) &= \sum_{i \in \mathcal{J}_s^{\text{target}}} \mathbb{1}[y_{s,i}^{\text{target}} > \tau_{\text{side}}], \\ n_-(s) &= \sum_{i \in \mathcal{J}_s^{\text{target}}} \mathbb{1}[y_{s,i}^{\text{target}} < -\tau_{\text{side}}], \\ n_0(s) &= \sum_{i \in \mathcal{J}_s^{\text{target}}} \mathbb{1}[|y_{s,i}^{\text{target}}| \leq \tau_{\text{side}}], \\ b_{\text{side}}(s) &= 1 - \frac{|n_+(s) - n_-(s)|}{n_+(s) + n_-(s)}, \\ n_+(s) + n_-(s) &> 0 \end{aligned} \quad (85)$$

compares positive and negative target-relative sides after excluding the neutral band $|y_{s,i}^{\text{target}}| \leq \tau_{\text{side}}$. Here $y_{s,i}^{\text{target}}$ is the dimensionless lateral displacement in the normalized proposal-support frame, and $\tau_{\text{side}} = 10^{-9}$ is a spatial classification tolerance, distinct from the reward tolerance. Circular orbit span

$$\begin{aligned} \alpha_{s,1} &\leq \dots \leq \alpha_{s,n_s}, \quad \alpha_{s,n_s+1} = \alpha_{s,1} + 2\pi, \\ s_{\text{orbit}}(s) &= 2\pi - \max_{1 \leq j \leq n_s} (\alpha_{s,j+1} - \alpha_{s,j}) \end{aligned} \quad (86)$$

measures the smallest angular arc covering the attempted directions. These statistics characterize proposal diversity before pruning. Family survival and oracle measurements describe the later feasibility and utility stages.

For evaluated candidates, the projection fraction

$$f_{\text{proj}}(s) = \frac{\sum_{i \in \mathcal{J}_s^{\text{evaluated}}} \mathbb{1}[i \in \mathcal{J}_s^{\text{projected}}]}{|\mathcal{J}_s^{\text{evaluated}}|}, \quad |\mathcal{J}_s^{\text{evaluated}}| > 0 \quad (87)$$

records whether the target centre falls inside the calibrated image domain. This measures framing; visibility also depends on object extent and occlusion.

C.13.1 Candidate-support populations and aggregation

Learning can rank only the alternatives supplied by candidate generation. The proposal profile is therefore part of the experimental design, and its support must be characterized before reward or policy results are interpreted. Let $\mathcal{J}_s$ contain all attempted non-padding rows, let $\mathcal{V}_s \subset \mathcal{J}_s$ contain the feasible rows selected by $m_{t,i}^{\text{act}}$, and let $\mathcal{L}_s \subset \mathcal{V}_s$ contain feasible rows with finite oracle labels. These nested populations separate three questions: what was proposed, what could be selected, and what could supervise learning.

The actor-valid fraction

$$f_{\text{actor}}(s) = \frac{\sum_{i \in \mathcal{J}_s} m_{s,i}^{\text{act}}}{|\mathcal{J}_s|} \quad (88)$$

measures pruning severity, whereas

$$n_{\text{valid}}(s) = |\{i \in \mathcal{J}_s : m_i^{\text{act}} = 1\}| \quad (89)$$

counts the number of alternatives that survive pruning. The distinction matters because two profiles can retain the same fraction while presenting very different numbers of choices. Family collapse is measured separately by

$$z_{\text{family}}(s) = \frac{1}{|\mathcal{F}_s|} \sum_{f \in \mathcal{F}_s} \mathbb{1}[n_{\text{valid}}(s, f) = 0] \quad (90)$$

which is one exactly when every configured family has zero feasible rows and zero only when every family survives. It detects a failure that a pooled valid fraction can hide: a large forward family may dominate the row count even when all target-conditioned families disappear.

Because many states can originate from one scene, row-level pooling would let large scenes dominate the study. Every state statistic $q(s)$ is therefore averaged within scene and then macro-averaged across scenes:

$$\begin{aligned}\mathcal{S}_{c,q} &= \{s \in \mathcal{S}_c : q(s) \text{ is defined}\}, \quad \bar{q}_c = \frac{1}{|\mathcal{S}_{c,q}|} \sum_{s \in \mathcal{S}_{c,q}} q(s), \\ \mathcal{C}_q &= \{c \in \mathcal{C} : |\mathcal{S}_{c,q}| > 0\}, \quad \bar{q} = \frac{1}{|\mathcal{C}_q|} \sum_{c \in \mathcal{C}_q} \bar{q}_c\end{aligned}\tag{91}$$

The metric-specific set $\mathcal{S}_{c,q}$ contains only states in scene c for which the statistic q is defined. Each included scene therefore contributes one scene mean, regardless of how many states it supplies. Failed generation roots contribute zero support and complete family collapse, while a directional statistic is undefined if no relevant direction was attempted. Undefined states and scenes are reported separately, and paired profile contrasts use the shared eligible-scene intersection. This preserves the difference between absence of support and absence of a defined measurement.

Support geometry is compared in the target-relative normalized frame $d_{t,e}^{\text{current}} = \|\mathbf{p}_e^W - \mathbf{c}_{R_t}^W\|_2$, $\tilde{\mathbf{c}}_{t,i}^{\text{support}} = (\mathbf{B}_{r,t}^{\text{Z-up}})^\top \frac{\mathbf{c}_{\tilde{c}_{t,i}}^W - \mathbf{c}_{R_t}^W}{d_{t,e}^{\text{current}}}$, $\tilde{\mathbf{p}}_{t,e}^{\text{support}} = (\mathbf{B}_{r,t}^{\text{Z-up}})^\top \frac{\mathbf{p}_e^W - \mathbf{c}_{R_t}^W}{d_{t,e}^{\text{current}}}$, $\|\tilde{\mathbf{p}}_{t,e}^{\text{support}}\|_2 = 1$. The factual expansion pose is the origin, the target lies at unit distance, and world up fixes the vertical axis. The normalization removes absolute standoff while preserving whether candidates move towards, around, above, or behind the target. This coordinate system describes candidate motion; calibrated projection and visibility are measured in their respective image and observation domains.

Whole-chain inspection uses a second, fixed frame so that changing expansion poses cannot hide accumulated motion. Its origin is the physical camera at the initial factual root, its horizontal forward axis points from that camera to the target centre, and world up completes the frame. The plot is reported both in metres and after division by the initial horizontal camera–target distance. Candidate families, feasibility, camera-forward headings, selected acquisition order, calibrated selected frusta, and the horizontal target-box footprint share this one frame. It is a geometric viewpoint-support diagnostic rather than a measurement

of visible target-surface coverage. Because the physical camera origin cannot be recovered from the corrected sampling reference alone, legacy stores without the persisted physical-camera anchor are reported as unavailable rather than aligned to a guessed origin.

Side balance describes the two lateral sides offered before pruning; circular orbit span describes their angular spread. Target-centre projection measures image framing rather than object visibility. The exact coordinate, tolerance and denominator definitions are retained in [Section C.13](#).

The immediate gain in these diagnostics is the decrease in target point-mesh error after adding the candidate observation, divided by the initial target error plus a numerical stabilizer. Both errors use the same reconstruction-preparation operator; the operational measurement and reward definitions are given in [Section 3.4.1](#) and [Section 3.4.4](#).

To assess available immediate gain independently of the policy’s choice, we compute the maximum over feasible candidates with finite labels:

$$o_{\text{oracle}}(s) = \max_{i \in \mathcal{L}_s} g_{s,i}^{\text{target-root}}, \quad |\mathcal{L}_s| > 0 \quad (92)$$

This is the best target-root gain in $\mathcal{L}_s$. Complete labels establish the exact maximum over $\mathcal{V}_s$. With missing labels, the observed maximum is a lower bound; with no labels, it is undefined. A small gain can reflect a nearly complete target or a need for a preparatory step, as well as poor proposals. The nested-pool comparison in [Section 5.4](#) tests whether additional proposals offer useful alternatives that the original generator missed.

Jitter compliance checks the configured angular perturbations only where bounded jitter is active. Its formula and excluded support class are retained in [Section C.14](#); it is generator quality control, not a utility or visibility measure.

Quantity	Population	Unit	Interpretation boundary
Actor-valid fraction	Attempted rows per state	fraction	Hard feasibility before label and training filters.
Valid support	Actor-valid rows per state	rows	Report the distribution, lower tail, and failed roots.
Configured-family zero rate	Configured families per state	fraction	Retains absent and zero-support families; state values are scene-macro reduced.
Side balance / circular span	Non-neutral attempted target-conditioned rows	fraction / angle	Neutral rows and undefined states are reported separately; proposal coverage precedes pruning.
Target-centre projection	Evaluated candidate views per state	fraction	Calibrated framing; zero-evaluated states are reported as undefined.
Oracle opportunity	Finite actor-valid rewards per eligible state	gain	Best labelled one-step gain; a lower bound if feasible labels are missing; undefined with no label.
Jitter compliance	Bounded-jitter rows per state	fraction	State-scene-cohort macro; report undefined, nonzero, and uncapped support separately.

Table 27: Candidate-generation diagnostics and their aggregation populations. Every state-level quantity is reduced before scene and cohort aggregation.

C.14 Jitter compliance

Finally, jitter compliance checks whether within-family perturbations obey their declared angular caps:

$$\begin{aligned}
c_i^\psi &= \mathbb{1}[|\Delta\psi_i| \leq |(\psi)_i|], \quad c_i^\theta = \mathbb{1}[|\Delta\theta_i| \leq |(\theta)_i|], \\
f_{\text{jitter}}(s) &= \frac{\sum_{i \in \mathcal{J}_s^{\text{jitter}}} \mathbb{1}[b_i = 1] c_i^\psi c_i^\theta}{\sum_{i \in \mathcal{J}_s^{\text{jitter}}} \mathbb{1}[b_i = 1]}, \\
\sum_{i \in \mathcal{J}_s^{\text{jitter}}} \mathbb{1}[b_i = 1] &> 0
\end{aligned} \tag{93}$$

Here c_i^ψ and c_i^θ are the per-row yaw- and pitch-cap indicators; the denominator retains only rows for which bounded jitter is active. Uncapped spherical profiles form a different support class and are reported separately. Jitter compliance provides generator quality control; visibility and policy quality use their own measurements.

C.15 Ranking, association and selected-action statistics

C.15.1 Ranking and calibration do not certify the selected maximum

On a fully labelled shell, let x_i be a model score and y_i the measured gain for the same candidate row. Spearman's coefficient is the Pearson correlation of their rank vectors, $\rho_s = \text{corr}(\text{midrank}(\mathbf{x}), \text{midrank}(\mathbf{y}))$, where $\mathbf{x} = (x_i)_{i \in \mathcal{A}}$ and $\mathbf{y} = (y_i)_{i \in \mathcal{A}}$. Ties receive their average rank. Fewer than two rows, or a constant score or gain vector, makes this correlation undefined; a tied predicted maximum also requires a declared selection rule. Spearman assesses the ordering of all rows, while regret evaluates the selected maximum in gain units.

For a separate affine fit, ordinary least squares estimates an intercept and slope from the same paired rows:

$$(\hat{\alpha}, \hat{\beta}) = \underset{\alpha, \beta}{\operatorname{argmin}} \sum_{i \in \mathcal{A}} (y_i - \alpha - \beta x_i)^2, \quad \tilde{y}_i = \hat{\alpha} + \hat{\beta} x_i \quad (94)$$

For n_A rows in the shell, when measured gains vary, its coefficient of determination is

$$R_{\text{aff}}^2 = 1 - \frac{\sum_{i \in \mathcal{A}} (y_i - \tilde{y}_i)^2}{\sum_{i \in \mathcal{A}} (y_i - \mu_y)^2}, \quad \mu_y = \frac{\sum_{i \in \mathcal{A}} y_i}{n_A} \quad (95)$$

The intercept matters: with one predictor, nonconstant scores and gains, this fitted R_{aff}^2 is squared Pearson correlation and can be high even for a negative slope, which reverses the ranking. If gains are constant, the denominator is zero and R_{aff}^2 is undefined; if only scores are constant, the intercept-only fit gives zero when gains vary. Even a positive slope and near-perfect global fit can miss the best view. With measured gains $(0, 1, \dots, 8, 9, 10)$ and scores $(0, 1, \dots, 8, 10, 9)$, Spearman is $\frac{109}{110} \approx 0.991$ and affine $R_{\text{aff}}^2 = \left(\frac{109}{110}\right)^2 \approx 0.982$, yet the score maximum chooses gain 9 rather than 10: within-shell regret is 1. Report ranking, fitted association, selected-action regret, and baseline improvement as distinct diagnostics, each with its defined shell and support count.

A common additive offset leaves within-state ordering and the chosen candidate unchanged, although it can change numerical prediction error.

The affine coefficients above are fitted retrospectively on the same candidate rows used to compute R_{aff}^2 ; they measure association within those rows. Held-out predictive calibration would instead freeze the fitted transform before evaluating new data. Per-state correlations use the candidates of one supported shell; a pooled correlation uses all admitted rows and also captures between-state offsets and scale differences. These have different denominators and need separately reported row, state, and defined-correlation counts. The surface-reference report computes shell-wise rank and affine association, then aggregates defined strata within scene and gives scenes equal weight. It marks constant-score shells uninformative; the mathematical intercept-only zero described above is therefore not an observed association in that report.

The retained one-step analysis averages paired state practical-regret reductions within each physical scene, then averages scenes equally. Its pair-ordering accuracy uses only pairs whose label difference exceeds the declared tolerance: correctly ordered pair counts are pooled within each scene and divided by that scene’s informative-pair count, before the scene accuracies are averaged. Predicted ties do not count as correct order. This pair statistic is distinct from Spearman and from an equal mean of state accuracies. Each regret contributes one value per admitted state; excluded states remain in planned and admission counts, while undefined statistics retain their support counts. Candidate rows, overlapping steps, and repeated seeds are not independent scene replicates. Uncertainty for the cited exploratory ranking analysis resamples physical scenes, and its result leaves the frozen practical-effect screen unchanged.

C.15.2 Policy advantage is a different quantity

Neither quantity is the reinforcement-learning **advantage** $\text{Adv}^{\pi(s,i)} = Q^{\pi(s,i)} - V^{\pi(s)}$, where $V^{\pi(s)} = \sum_{j \in \mathcal{A}_s} \pi(j|s) Q^{\pi(s,j)}$. Advantage compares an action with the value of the **same continuation policy** at that state; it is not a maximum-over-shell regret or an improvement against a separately chosen incumbent. At one step, complete immediate labels support the action comparison. At longer horizons, factual selected transitions alone do not provide $Q^{\pi(s,i)}$ for unselected rows.

C.16 Extended literature comparison

Family	Utility and target scope	State evidence and representation role	Support, feasibility, and decision
PB-NBV [9]	Projected frontier and occupied coverage; reconstruction object	Analytic classified voxel clusters and compact projection proxies	Reachability- and camera-conditioned partial hemisphere; greedy score
3D-NVS [13]	Pairwise voxel reconstruction; object	ImageNet-pretrained VGG16 directly scores fixed next-view classes	Eleven fixed view classes; one-step classifier
SCONE / MACARONS [3], [10]	Expected surface coverage; scene-wide	Task-trained occupancy and visibility; pretrained ResNet features support MACARONS depth	Iterative candidate selection
GenNBV [7]	Coverage gain; reconstruction object	Task-trained geometric and history encoders	Continuous, history-conditioned policy
Hestia [8]	New voxel-face coverage; reconstruction object	Cumulative directional grid, shallow image features; no external pretraining	Hierarchical look-at then position; endpoint adjustment
APN [32]	Exploration and surface coverage; environment	Incremental map and perception roadmap	Non-learned, non-myopic sequence planning
FisherRF [11]	Information gain; radiance-field scene model	Per-scene model uncertainty enters an analytic score	Candidate views; greedy score
VIN-NBV [1]	Direct reconstruction improvement; evaluated geometry	Task-trained CNN over candidate-conditioned projections of current reconstruction	Sampled candidates; greedy score
Next Best Sense [12]	Color-depth information gain; radiance-field scene model	SAM2 and depth priors support 3DGS; Fisher selection remains analytic	Feasible candidates; fallback after kinematic or planning failure
Object-centric 3DGS [2]	Object-conditioned information gain; requested object	Foundation-model masks and CLIP support object state and selection; target score remains analytic	Candidate views; greedy score

Table 28: Concept-centred comparison of the literature families that bound the thesis question. APN denotes Active Perception Network. Rows distinguish utility, target scope, state evidence, action support, and decision structure; they are not ranked by reported performance across incompatible settings.

C.17 Bounded runtime attribution

A diagnostic screen completed six eight-step rollouts: three fixed scenes, each once with ordinary execution and once with capture and profiling enabled. All six

independent endpoint audits passed; the three observed pairs retained identical decisions, stored observations and raw endpoint geometry. The source was fixed at 74555a8a9, with one CPU thread and the same 60-candidate recipe. The ordinary worker times were 42.33, 526.25 and 138.49 seconds. In the corresponding instrumented runs, endpoint measurement took 16.70, 336.06 and 1.97 seconds; observed mesh-preparation host probes totaled 0.016, 0.275 and 0.087 seconds. These diagnostic scopes overlap and cannot be summed.

The registered runtime evidence preserves the source, six intentions, receipts, resource counters and independent readback. One pair per scene provides no population interval or calibration of instrumentation overhead. Candidate rendering and scoring used CUDA, whereas the independent endpoint reference measurement ran on the CPU. Full selected-view acquisition and retained depth remained part of generation. The screen selected no cache patch and launched no proposed 18-attempt optimization contrast; it establishes neither a new acceleration nor corpus admission.

C.18 Selected-camera diagnostic definitions

The diagnostics compare sampled physical camera poses. Acquisition index $t = 0$ denotes the root and $t = 1, \dots, H$ the selected observations; it is not a physical time variable. Write $\mathbf{T}_t = \mathbf{T}_{\mathbf{C}_t}^W$ for the world-from-camera transform, with rotation $\mathbf{R}_t$ and camera centre $\mathbf{c}_t$ in world coordinates. Rigid-transform inversion and frame composition give [41, December 2019 preprint, Sec. 3.3.1.1, Eq. 3.64; Sec. 3.3.1.2]:

$$\begin{aligned}\Delta \mathbf{T}_t &= \mathbf{T}_{t-1}^{-1} \mathbf{T}_t, & \Delta \mathbf{R}_t &= \mathbf{R}_{t-1}^\top \mathbf{R}_t, \\ \Delta \mathbf{c}_t &= \mathbf{R}_{t-1}^\top (\mathbf{c}_t - \mathbf{c}_{t-1})\end{aligned}\tag{96}$$

The relative translation is expressed in the preceding camera frame. Its norm equals the world displacement because the preceding rotation is orthogonal. Translation increment d_t and geodesic orientation increment α_t therefore use:

$$d_t = \|\Delta \mathbf{c}_t\|_2, \quad \alpha_t = \arccos\left(\text{clip}\left(\frac{\text{tr}(\Delta \mathbf{R}_t) - 1}{2}, -1, 1\right)\right)\tag{97}$$

The trace-angle relation follows the rotation logarithm in [41, December 2019 preprint, Sec. 3.2.3, pp. 85–86]. Clipping is our floating-point guard. Rotation includes yaw, pitch and roll; it differs from the camera centre’s orbital azimuth.

Let $\mathbf{e}_z$ be world up. A horizontal movement bend compares two adjacent nondegenerate translation directions:

$$\begin{aligned}\mathbf{v}_t &= (\mathbf{I} - \mathbf{e}_z \mathbf{e}_z^\top)(\mathbf{c}_t - \mathbf{c}_{t-1}), \\ \beta_t &= \arccos\left(\text{clip}\left(\frac{\mathbf{v}_{t-1}^\top \mathbf{v}_t}{\|\mathbf{v}_{t-1}\|_2 \|\mathbf{v}_t\|_2}, -1, 1\right)\right)\end{aligned}\tag{98}$$

This is the ordinary angle-between-vectors construction applied to horizontal displacements. We call $\beta_t > \frac{\pi}{2}$ backtracking; that threshold is a descriptive convention, not a human-motion criterion. The first bend is undefined. Horizontal displacements at most 10^{-4} metres also leave the bend unavailable rather than contributing zero.

C.18.1 Target-centred positional occupancy

Let $\mathbf{o}$ be the target centre. The orthonormal basis $\mathbf{B}$ has a fixed horizontal first axis pointing from the root toward the target, a second axis given by world up crossed with the first, and a world-up third axis. The basis is fixed for the entire chain. Root-relative azimuth and sine elevation are:

$$\begin{aligned}\mathbf{x}_t &= \mathbf{B}^\top(\mathbf{c}_t - \mathbf{o}), \quad \psi_t = \text{atan2}(x_{t,y}, x_{t,x}), \\ \varphi_t &= W(\psi_t - \psi_0), \quad \zeta_t = \frac{x_{t,z}}{\|\mathbf{x}_t\|_2}, \quad \mathrm{d}\Omega = \mathrm{d}\varphi \mathrm{d}\zeta\end{aligned}\tag{99}$$

The root’s horizontal projection lies on the negative first axis, so subtracting ψ_0 is essential to put it at zero azimuth. For polar angle θ , the spherical area element is $\mathrm{d}\Omega = \sin(\theta) \mathrm{d}\theta \mathrm{d}\varphi$ [50, Sec. 4.2.2, Eq. 4.8]. Substituting elevation $\frac{\pi}{2} - \theta$ gives the positive measure $\mathrm{d}\varphi \mathrm{d}\zeta$. Uniform rectangular bins in azimuth and sine elevation consequently have equal solid angle, as also used by uniform sphere sampling [50, Appendix A.5.2]. This conversion is our application of the source’s Jacobian; it does not turn camera-position occupancy into observed-surface coverage.

At each acquisition, the histogram averages cell membership over available scenes. The whole-chain histogram first divides each chain’s bin counts by its number of available selected positions, then averages chains within scene and scenes equally. Roots anchor azimuth but are excluded from occupancy. The present cohort has one complete chain per scene and 97 available positions at every acquisition, so averaging its sixteen per-acquisition maps reproduces the whole-chain map. The complete atlas is [Figure 45](#).

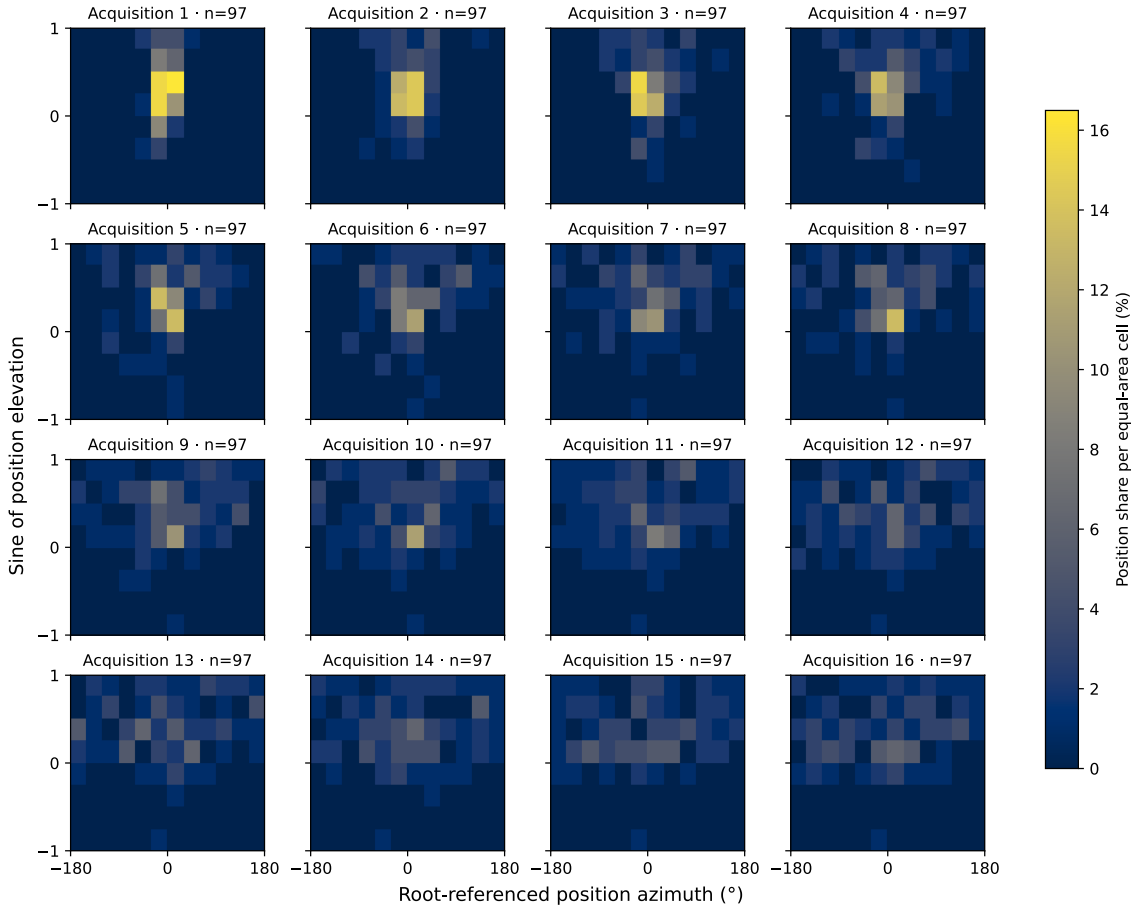


Figure 45: Positional occupancy at all sixteen acquisitions in the same 97 scenes. Every panel uses the same 12 by 8 equal-solid-angle cells and common colour scale. Percentages are available-scene shares at that acquisition, not confidence limits. Zero azimuth is the initial camera side; the vertical coordinate is sine elevation. The maps omit distance, gaze and mesh occlusion. Later cross-scene dispersion does not imply that each individual chain completes an orbit.

C.18.2 Signed angular increments and whole-chain quantities

We choose the shorter signed endpoint difference using this branch convention:

$$W(a) = \text{atan2}(\sin(a), \cos(a)), \quad \delta\varphi_t = W(\psi_t - \psi_{t-1}) \quad (100)$$

Positive $\delta\varphi_t$ is counterclockwise when viewed from world up. Azimuth is unavailable when horizontal target radius is at most 10^{-4} m; exact half-turns have ambiguous direction. Nonzero direction counts use a numerical tolerance of 10^{-6} degrees. These conventions are ours, not a claim that a cited source prescribes the collector. A counterclockwise share counts positive increments among available nonzero increments; counting scenes at a fixed acquisition and counting acquisitions within a chain are different aggregation levels. The fraction of available bends exceeding $\frac{\pi}{2}$ similarly defines the backtrack frequency.

The sampled polyline length L_H , root-to-final displacement D_H , signed angular progress Θ_H and absolute angular travel A_H are:

$$\begin{aligned} L_H &= \sum_{t=1}^H d_t, \quad D_H = \|\mathbf{c}_H - \mathbf{c}_0\|_2, \\ \Theta_H &= \sum_{t=1}^H \delta\varphi_t, \quad A_H = \sum_{t=1}^H |\delta\varphi_t| \end{aligned} \tag{101}$$

Directional angular amounts retain magnitudes rather than just signs:

$$\begin{aligned} A_H^{\text{CCW}} &= \sum_{t=1}^H \max(\delta\varphi_t, 0), \quad A_H^{\text{CW}} = -\sum_{t=1}^H \min(\delta\varphi_t, 0), \\ A_H &= A_H^{\text{CCW}} + A_H^{\text{CW}}, \quad \Theta_H = A_H^{\text{CCW}} - A_H^{\text{CW}} \end{aligned} \tag{102}$$

All angular formulas are in radians; the figures convert them to degrees. Sums require complete available increments. Neither sampled polyline length nor summed shorter angular changes identifies the intervening physical route or continuous winding. A geometric path becomes a timed trajectory only after time parameterization [41, Sec. 9.1]. Temporal smoothness also depends on task and measurement conditions [42, Eq. 1 and Table 1]. Thus the spatial descriptors in Results do not establish jerk or natural wearer movement.

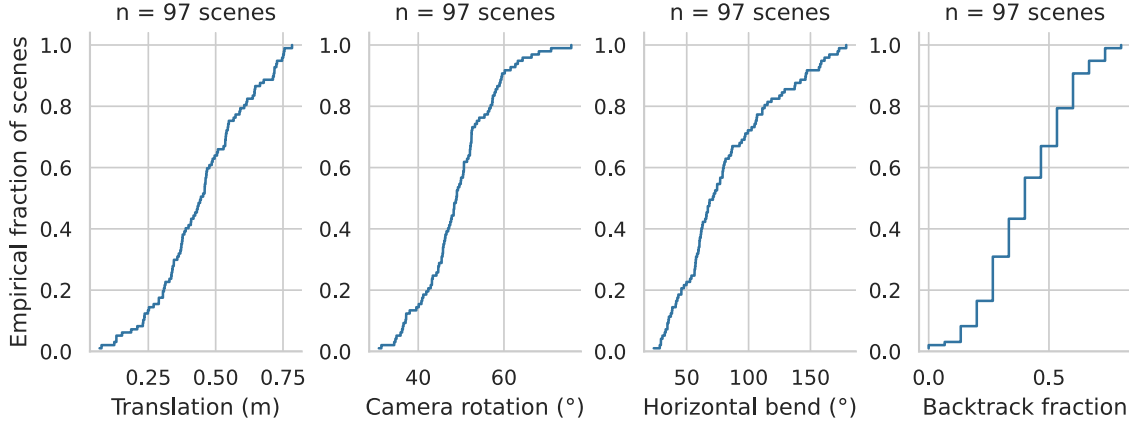


Figure 46: Within-chain continuity summaries across 97 physical scenes. Empirical distributions retain each scene’s median translation, geodesic camera rotation and horizontal bend, plus its available-bend backtrack fraction. First and degenerate bends remain unavailable. The summaries collapse acquisition position; Results presents the complementary per-acquisition view. No time normalization or human-motion claim is made.

C.19 Historical candidate-profile diagnostics

C.19.1 Radius and motion-bound alignment

A paired diagnostic tested whether aligning the maximum proposal radius with the existing motion bound preserved candidate support. It compared 1.1 m against 1 m while retaining a 1 m step bound. Across 6 fixed training-scene roots, 2 seeds, and a requested horizon of 2 under privileged ground-truth target selection, alignment reduced proposals beyond the motion bound from 209 of 1920 to 0 of 1920. Both profiles completed 10 of 12 intended chains; the remaining attempts failed root admission. No family with positive baseline support lost all support at the paired root-precheck or step-0 states. This bounded result supports retaining the aligned setting as a named diagnostic profile. It does not establish improved reconstruction quality, population-level advantage, equivalence, or non-inferiority, does not establish observed-target behavior, and leaves the principal profile’s declared 1.1 m radius unchanged.

Transition consistency was audited separately over 40 selected transitions from 20 completed stores. Committed successor point tensors and persisted selected labels agreed exactly. Scalar successor recomputation was bit-exact in 17 of 20 stores and agreed within $\text{atol} = 0\text{m}^2$ and $\text{rtol} = 10^{-6}$ in all stores. These parity

results test scoring-to-continuation agreement; they do not establish that the proposed views are useful. The archived machine-readable bundle retains source identities, per-run configuration hashes, failure records, observer provenance, and per-run transition-receipt summaries and hashes.⁷

C.19.2 Three-scene development confirmation

A frozen development confirmation compared the original 1.1 m sampled-family radius with a 1 m challenger. The artifact identifies these profiles as B0 and S, respectively. Both used 60 candidates per step, one 8-step chain, and temperature-softmax selection at temperature 1. Across 24 intended cells per profile, B0 completed 11 and S completed 9. Of the discordant outcomes, 5 cells completed only under B0 and 3 only under S. The conditional shared-complete analysis retained 6 pairs across 3 scenes. Its scene-level mean S-minus-B0 endpoint-gain difference was -0.0098 , the median was 0.0044 , and the 95% scene-cluster percentile interval was $[-0.079, 0.0452]$. The B0 profile had 16 admitted, 4 non-admitted, and 4 undefined cells. The S profile had 16 admitted, 4 non-admitted, and 4 undefined cells. Admission preservation remained undefined because target-current support was insufficient for two scenes. The challenger therefore failed the development gate, was not promoted to a holdout test, and does not justify changing the declared 1.1 m profile. This bounded panel does not establish a population-level advantage.⁸

C.20 Historical collection coverage

The frozen September collection records 210 attempts from 77 source roots, with 186 validated stores from 75 roots and 2,749 selected acquisitions. Ten attempts failed for insufficient proposal support, thirteen were rejected by measurement, and one failed during execution. These are recorded-attempt summaries from training-split, ground-truth-target collection under the 60-candidate recipe and its

⁷Diagnostic evidence: `candidate_scale_h2_diagnostic_20260918.json`, with the companion report and observer source retained beside it.

⁸Evidence: `rollout_profile_development_confirmation_20260920.json` and its companion report; public metric receipt: Weights & Biases.

historical evaluator. They neither replace the later exact-box cohorts nor establish a complete campaign census.

The successful records contain 164,940 proposed candidate rows and 78,430 hard-valid rows. The train-labelled total is also 78,430; equal totals alone do not prove equal masks. Twenty-one recorded states have fewer than two valid actions. Thus completing many trajectories still leaves weakly supported individual decisions.

Minimum suffix	Source roots	Stores	Starts
8 acquisitions	74	175	1,482
12 acquisitions	72	165	802
16 acquisitions	71	156	156

Table 29: Recorded historical continuation support. Roots and stores have at least the stated remaining length; starts count states with that many subsequent acquisitions. Rows are cumulative across lengths. Starts can overlap within a stored trajectory; they are neither independent samples nor complete recursive targets for every candidate.

The sum of durations over all recorded attempts is 15.25 hours, including failed attempts. Dividing the 12- and 16-acquisition start counts by that sum gives 52.60 and 10.23 starts per summed attempt-runtime hour, respectively. These descriptive rates use summed attempt time, not elapsed campaign time. Longer retained suffixes support more return queries, but do not identify an optimal planning horizon or supply successors for unselected actions.

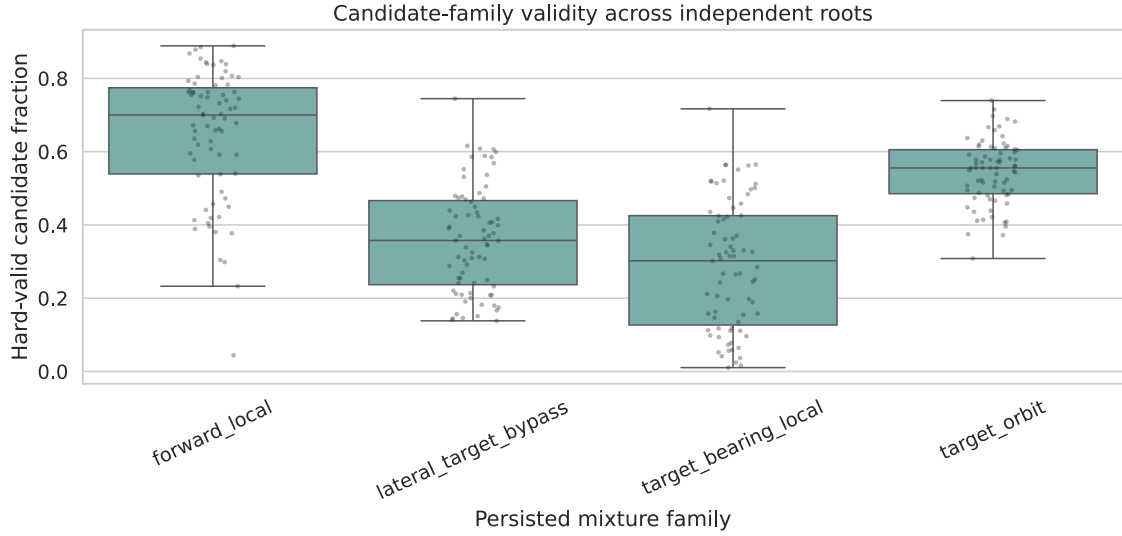


Figure 47: Historical candidate-family validity across 75 successful source roots. Each root’s fraction pools valid and proposed candidates across its recorded acquisitions. The recipe allocates 18 forward, 18 bearing, 12 bypass and 12 orbit proposals per state. Family medians are 0.70, 0.30, 0.36 and 0.56, respectively. Variation reflects quotas, filtering and scene geometry, not a causal family comparison. The archived title’s “independent roots” does not establish independent physical scenes.

The source projection, ledger, configuration, roster and figure are mapped in [Chapter A](#). Their preserved identities support this historical readout; relocated records alone do not revalidate the original child stores or queue-completion scope. No counts are pooled with the later 97-scene collections or the 21-scene learning screen. This evidence establishes recorded generation coverage, not current training admission, value accuracy or learned policy benefit.

C.21 Historical no-render proposal audit

The proposal audit covered 100 target states from 100 scenes. Of 6000 generated candidate views, 3146 remained after validity filtering. We also checked whether each applicable candidate-generation family retained any valid candidates at each state. In 44 such state–family combinations, none remained. It failed its support gate: 24 states missed the aggregate non-forward target-aware-family floor, and 8 states missed the root-support threshold. All 100 reviewed source rows, 100 scenes, and 100 target states were represented; the immutable bundle records 0 source-row exclusions. Because this no-render audit contains no reward labels, flat-gain status is unavailable with label denominator 0 and eligible-state denominator

0. The result diagnoses proposal support. It provides no evidence about RRI, candidate quality, rollout throughput, or policy performance and does not admit broad rollout generation.

Phase-A diagnostic	Bundle-derived value
Generated candidate views	6000
Candidates remaining after validity filtering	3146
State-family combinations with no valid candidates	44
States below target-family floor	24
States below root-support floor	8
Excluded source rows	0

Table 30: Bounded Phase-A proposal-support diagnostic projected from the authenticated evidence bundle. These counts describe support and exclusions; they are not policy-performance estimates.

Earlier training-source rollout attempts reached mesh rendering, target-specific oracle scoring, and selected-action replay. An out-of-memory failure during unbatched candidate rendering and subsequent memory-bounded attempts identify rendering as a scale gate; they do not establish population throughput or storage feasibility.



Figure 48: Candidate-family survival for one deterministic factual state from each persisted Phase-A audit stratum. Each cell reports the number of candidates remaining after validity filtering divided by the number proposed; the complete 100-state matrix remains in the authenticated evidence bundle. The figure, table, prose counts, and exclusion summary are generated by `make phase-a-thesis-projection`. The source artifact has digest `6d33e9e3d68737c8...`; the derived projection has digest `c1f8791c1f09f71e...`.

Ehrenwoertliche Erklaerung

Ich versichere, dass ich diese Masterarbeit selbstständig verfasst, noch nicht anderweitig für Prüfungszwecke vorgelegt, keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt sowie wörtliche oder sinngemäße Zitate als solche gekennzeichnet habe.

Muenchen, 7. Jan 2027

Jan Duchscherer