TIMING OF MEMORY INTEGRATION SHAPES REPRESENTATIONS OF CAUSAL UNCERTAINTY

Anonymous authors
Paper under double-blind review

ABSTRACT

Two questions have largely been studied separately in hippocampal research: how uncertainty over hidden structure is represented and when episodic memories are integrated relative to a query. We bring them together by extending the Episodic Spatial World Model (ESWM) framework from spatial mapping to inference over latent causal structure. We compare matched architectural instantiations of *integrative encoding*, which compresses ambiguous evidence before a query is known, and *retrieval-based inference*, which combines stored episodes after the query identifies the relevant causal relation. DeepMind's Symbolic Alchemy provides an enumerable space of causal worlds, allowing us to compute the exact posterior over structures and the posterior predictive for every query. This provides a normative reference for comparing the beliefs expressed by models with nearly indistinguishable behavioral accuracy. At sufficient capacity, the two modes achieve nearly indistinguishable behavioral performance while relying on different internal computations and uncertainty representations: retrieval-based-inference models preserve more separable memories, preferentially retrieve query-relevant evidence, and represent uncertainty more faithfully, whereas integrative-encoding models form representations that generalize better across queries. As a single representation is required to support more queries, models smoothly shift from query-specific toward reusable, integrative-encoding-like representations. These results characterize integrative encoding and retrieval-based inference as endpoints of a trade-off between query-specific inference and representational reuse.

1 INTRODUCTION

Biological agents routinely infer relations they have never directly experienced by combining memories acquired at different times. The hippocampal formation is thought to support this ability (Dusek & Eichenbaum, 1997; Buckmaster et al., 2004; Eichenbaum & Cohen, 2014). Such inferences can arise in at least two ways. An agent can integrate observations into a reusable representation that supports many future queries, or it can store memories separately and combine them only when a particular inference is required. Both have received considerable attention in the hippocampus literature. **Integrative encoding (IE)** proposes that overlapping memories are merged *at encoding* into a structure that supports later inference (Shohamy & Wagner, 2008; Schlichting & Preston, 2015; Zeithamova et al., 2012b). **Retrieval-based inference (RBI)** instead proposes that memories remain separate during encoding and are recombined *on demand*, a form of hippocampal indexing theory (Teyler & DiScenna, 1986; Kumaran & McClelland, 2012; Zeithamova et al., 2012b; Koster et al., 2018). As evidence for both has grown, they have increasingly been viewed as complementary (Zeithamova & Preston, 2010; Barron et al., 2013; Schlichting et al., 2015; Zeithamova & Bowman, 2020; Molitor et al., 2021), yet their computational relationship is not fully understood: which computational properties they share, how the representations they build differ, and under what conditions each mode should be favored.

Artificial neural networks provide a complementary way to study this question. In this study, we operationalize the two computational modes by manipulating *when the query becomes available* to the network. *Query-conditioned* models access memories along with the query. They can selectively attend to query-relevant events, enabling retrieval-based inference. *Query-blind* models process memories before seeing the query and must therefore construct a reusable representation that can support unknown future queries, imposing the core computational constraint of IE. We compare

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the two model types on inference over latent causal structure. A memory is then not a fragment of a map, but evidence about an indirectly observed rule system, and sparse evidence can leave many causal structures consistent with the observations. Inference therefore requires recovering hidden structure rather than stitching together observed fragments, and representing uncertainty over possible structures rather than accumulating certainties. We study this problem in DeepMind Symbolic Alchemy (Wang et al., 2021), whose finite space of causal worlds allows us to compute the exact posterior over structures and posterior predictive for each query.

We find that capacity-matched query-conditioned and query-blind models can achieve nearly indistinguishable behavioral performance while implementing different computations: query-conditioned models preserve more separable memories, preferentially retrieve query-relevant evidence, and represent uncertainty more faithfully relative to the exact Bayesian observer. Across tasks, these differences reveal a trade-off between query-specific inference and representational reuse, with query conditioning favoring accurate inference for individual queries and query-blind integration becoming increasingly useful as a shared representation must support many possible queries. Overall, the results expose both commonalities and differences between integrative encoding and retrieval-based inference that are difficult to identify from behavior alone.

2 Related work

Integrative encoding versus retrieval-based inference. Integrative-encoding accounts propose that overlapping experiences are linked during learning, such that indirectly experienced relations are already represented when later queried (Shohamy & Wagner, 2008; Zeithamova & Preston, 2010; Schlichting et al., 2014; Schlichting & Preston, 2015). This view tends to emphasize the hippocampus function as organizing relational memories instead of storing individual items (Cohen & Eichenbaum, 1993; Eichenbaum et al., 1999; Schlichting & Preston, 2017). Retrieval-based inference instead preserves distinct episodic traces and recombines them on demand when an inference is requested (Kumaran & McClelland, 2012; Banino et al., 2016; Koster et al., 2018), an idea closely related to hippocampal indexing accounts (Teyler & DiScenna, 1986; Teyler & Rudy, 2007). This view emphasizes rapid storage and retrieval of individual episodes, such as content-addressable autoassociative memory system (Hopfield, 1982; Marr et al., 1991; McNaughton & Morris, 1987; O'Reilly & McClelland, 1994). Previous work has contrasted RBI and IE through encoding-time reactivation, test-time reactivation, response time, inference accuracy, and fits to behavior (Zeithamova & Preston, 2010; Zeithamova et al., 2012a; Banino et al., 2016; Koster et al., 2018). We extend these comparisons to an understudied question: whether matched computational instantiations of IE and RBI represent uncertainty about hidden causal structure differently, even when they have similar inference accuracy and fits.

Cognitive maps and Causality. The cognitive-map was initially proposed within the context of spatial navigation (Tolman, 1948; O'keefe & Nadel, 1978), but contemporary accounts characterize a map more broadly as a reusable representation of relational structure over spatial and non-spatial variables(Eichenbaum & Cohen, 2014; Behrens et al., 2018). Computational models including the Tolman–Eichenbaum Machine and clone-structured graphs show how such structure can be learned (Whittington et al., 2020; George et al., 2021). Emerging evidence extends this account from generic relational structure to causal structure. Hippocampal activity is shaped by an event's causal connections (Lee & Chen, 2022), CA1 population sequences emerge to bridge learned cause-effect relationships (Modi et al., 2014), and human hippocampal activity represents causal inference (Mahmoodi et al., 2024). More generally, it was suggested that hippocampus represents relational structure as a map of causal relations that supports inference about unobserved cause-effect relationships (Luo et al., 2026). We extend this line of work by asking how uncertainty over latent causal structures is represented, and whether that representation depends on integrating causal evidence before a query or combining it at retrieval.

Retrieval-based models. REMERGE and MEMO model relational inference through recurrent or multi-hop combination of distinct episodic memories (Kumaran & McClelland, 2012; Banino et al., 2020). ESWM (He et al., 2026) is a meta-learned model that stores observed transitions in a permutation-invariant bank and retrieves against the query. It was introduced for spatial layouts in which memories are fragments of the map. Here we extend ESWM from spatial map fragments to

uncertain causal relationships and, critically, introduce matched query-blind and query-conditioned variants to isolate the computational consequences of integration timing.

Bayesian inference. Bayesian accounts have characterized hippocampal representations as reflecting uncertainty and inference over latent states (Harrison et al., 2006; Sanders et al., 2020; Mahmoodi et al., 2024). Memory-based meta-learning can produce models whose behavior approximates the behavior of a Bayes-optimal learner (Ortega et al., 2019; Mikulik et al., 2020; Binz et al., 2024). However, behavioral agreement with a Bayesian observer does not by itself establish that a model represents the same uncertainty. Because Alchemy's causal worlds are finite and enumerable, we can compare model predictions with the exact posterior and ask whether predictions across different queries jointly imply a coherent belief about the underlying world.

3 METHODS

3.1 TASK

Environments. Each causal world (termed chemistry) is a cube graph in a latent abstract space spanning three perceptual axes: shape, color, and size. Each latent cube vertex, i.e. latent state, is assigned an observable state (termed stone) whose appearance is determined by its shape, color, and size. Each observable state also carries a reward value that changes across worlds. There are six unique observable transformation cues (termed potions) that each transform the observable state along one of the latent cube edges, which induces a change in the state's appearance along one or more perceptual axes. Across worlds, the mapping between latent state and observable state, the mapping between transformation cue and latent transition, and the connectivity of the causal graph can change, contributing to a total of 167,424 possible causal worlds. We divide the worlds into train/val/test splits. The *memory* of a transition m is an observable state transitioning to new observable state after applying a transformation cue to the starting state, or, in other word, a potion-induced transition between two stones in a chemistry. A series of events collected in a world is called the *evidence set* M.

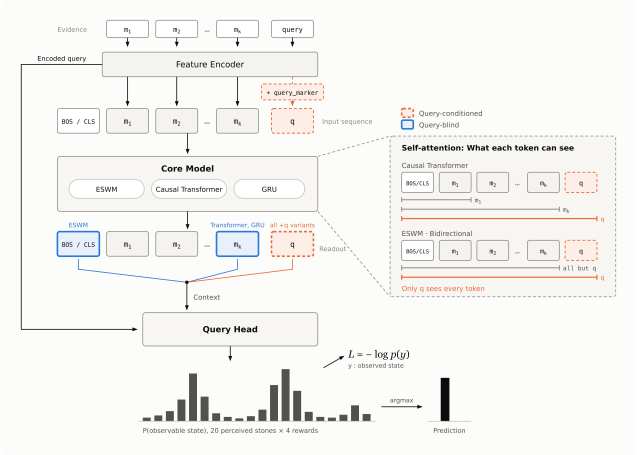

Figure 1: **Query-conditioned and query-blind models.** The sequence of events $\{m_i\}_{i=1}^{k} \cup \{q\}$ is embedded into tokens. Query-blind models receive only the evidence, whereas query-conditioned models additionally receive q.

Task. In each sample, the model receives M collected from a world and needs to infer the outcome of a transition q (e.g, predict how a transformation cue would transform a given state). We adopt a meta-learning training where each input is from a different causal world and the model need to learn to reason in-context to predict an outcome in constantly changing causal structures, 133k worlds in the train set. We then evaluate on test worlds held-out to ensure the model learned a generalizable approach to recover the hidden causal structures through relating the observed set of evidence.

We keep the reinforcement learning structure of the task where each causal world consists of 10 trials, here performed under a random policy (see Appendix A.2 for detail). Each trial contains 12 transitions, and thus 12 memories. To construct one model input, we uniformly sample transitions over $T \in \{0, \ldots, 8\}$ trials, use the $T \times 12$ memories as M which serves as context enabling the model to reason about the current world dynamics, and randomly select a transition from the final two trials as the query event. We mask the query event's outcome, the end observable state, to construct a q.

3.2 MODELS

We manipulate whether the q is accessible to the core network θ that processes the M. Prior to the θ, we first convert M into a sequence of tokens by embedding each memories m into a vector through a feature encoder. We then prepend a CLS token (i.e, a learnable token) to M. For *query-blind* models, only M is provided to θ. The network is forced to encode M without knowing which inference it needs to later support, a hallmark of IE. For *query-conditioned* model (or simply **+q**) the concatenation of evidence set and the query (M, q) is provided (Fig. 1). Such a design enables θ to selectively attend to the events that are query-relevant. Although the possibility that *query-conditioned* networks converge to the same algorithm as those learned by *query-blind* models is not eliminated, we show in various ways that this is not the case (see Sec4.1).

For θ, we consider unidirectional **GRU**, Transformer with causal mask (**Transformer**), and bidirectional Transformer (termed **ESWM**, due to the similarity of the training framework $\times$ architecture to prior work Episodic Spatial World Model (He et al., 2026)). For each architectural choice we derive its *query-conditioned* (**+q**) and *query-blind* variant. The output of θ is a context vector $c = \theta(M, q)$, where $q = \emptyset$ for query-blind variants, which we select as the final layer output of CLS token for ESWM; the last token in the sequence M for Transformer and GRU; and q token for every +q models (see readout Fig. 1). The context vector from the θ neural network and query from feature encoder are passed through a MLP prediction head, query head, to produce 80-way classification over 80 possible stone $\times$reward outcomes.

For each model, we derive variants with different number of parameters: 0.33M (1$\times$), 1.7M (5$\times$), 3.3M (10$\times$), and 6.6M (20$\times$), by systematically increasing different hyperparameters (see Appendix A.1 for model details). In addition, we train each model with three random seeds (see Appendix A.3 for training details) and report aggregated results whenever possible. We then freeze the models and compare query-conditioned and query-blind variants of the otherwise identical model.

3.3 EXACT ORACLE POSTERIOR

Because DSA has a finite, enumerable set of causal worlds W, we can compute the exact Bayesian posterior. Let $M = \{m_i\}_{i=1}^{k}$ denote the observed transitions and q a query consisting of a start state and transformation cue. Under the uniform prior $p(w) = 1/|W|$, the observations rule out incompatible worlds, leaving the support $W(M, q)$. Because transitions are deterministic and drawn independently, the posterior is uniform over this support: $p^*(w \mid M, q) = 1/|W(M, q)|$ for $w \in W(M, q)$, and 0 otherwise.

The posterior predictive probability of outcome $y \in Y$ is therefore the fraction of compatible worlds that produce y: $p^*(y \mid M, q) = \frac{1}{|W(M,q)|} \sum_{w \in W(M,q)} \mathbb{I}[y = f_w(q)]$, where $f_w(q)$ is the outcome of q in world w. We compare the exact Bayesian observer's distribution with the model prediction $p_\theta(y \mid M, q)$ using $\mathrm{KL}(p^* \parallel p_\theta)$.

We also ask whether the model's predictions jointly imply a coherent posterior over worlds. Let Q be the set of all possible queries. For each $w \in W$, we define w compatibility with the model's predictions as $L_\theta(w \mid M) = \prod_{q' \in Q} p_\theta(f_w(q') \mid M, q')$, and normalize these scores to obtain the model-implied posterior over worlds: $p_\theta(w \mid M) = L_\theta(w \mid M) \,/\, \sum_{\tilde{w} \in W} L_\theta(\tilde{w} \mid M)$. We compare this distribution with the exact oracle posterior $p^*(w \mid M)$ using KL divergence. Note that, because L_θ multiplies marginal predictions across queries, even the exact Bayesian observer has $KL(p^*(w \mid M) \parallel \hat{p}^*(w \mid M)) > 0$ where $\hat{p}^*(w \mid M) = L^*(w \mid M) \,/\, \sum_{\tilde{w} \in W} L^*(\tilde{w} \mid M)$. A smaller world-level KL indicates that the model's predictions jointly imply a distribution over causal worlds closer to the exact posterior.

4 RESULTS

4.1 SIMILAR BEHAVIORAL ACCURACY CONCEALS DISTINCT COMPUTATIONAL STRATEGIES

We first validate that all models learned to do in-context reasoning. By 5x capacity, every architecture exceeds 92% accuracy on held-out worlds, close to the exact observer's 93.9% (Table. 1). At 20x, adding the query changes accuracy by just 0.01–0.27 percentage points across matched pairs, failing to distinguish query-conditioned from query-blind computations and raising an important

Table 1: Top-1 accuracy (%) on the test set. Each column is a different model size. Last column is the accuracy gap at 20x, query-conditioned model $(+q)$ minus query-blind model. Exact Bayesian observer is an oracle ceiling. SD over three seeds.

Architecture	1x (0.33M)	5x (1.7M)	10x (3.3M)	20x (6.6M)	Δ_{+q} at 20x
GRU	88.37 ± 0.26	92.58 ± 0.08	92.76 ± 0.04	92.73 ± 0.02	
GRU + q	84.71 ± 3.39	92.68 ± 0.02	92.77 ± 0.05	92.74 ± 0.09	0.01
Causal transformer	67.00 ± 0.00	92.42 ± 0.49	92.98 ± 0.18	93.31 ± 0.05	
Causal transformer + q	92.27 ± 0.19	93.38 ± 0.01	93.50 ± 0.01	93.44 ± 0.04	0.14
ESWM	88.32 ± 0.35	93.00 ± 0.09	93.21 ± 0.03	93.25 ± 0.02	
ESWM + q	92.58 ± 0.02	93.48 ± 0.03	93.52 ± 0.01	93.52 ± 0.01	0.27
Exact Bayesian observer				93.90	

question. Did query-conditioned and query-blind models converge on the same algorithm, or did they learn divergent algorithms that are each consistent with IE and RBI?

We test the models against a hallmark of IE, whether the model's encoded evidence representation, i.e., the context vector $c_\theta(M, q)$, supports arbitrary inference. We extract $c_\theta(M, q_{enc})$ and use it to predict a different query q_{pred} and measure the performance degradation: $KL\big(p^*(y \mid M, q_{pred}) \parallel p_\theta(y \mid c_\theta(M, q_{enc}), q_{pred})\big)$ which in a normalized form give us the context transfer retention score R (see Appendix A.8). For query-blind models, c_θ does not depend on q_{enc} (i.e. $c_\theta(M, \emptyset)$), so $R = 1$ by construction. We observe that query-conditioned's context vector poorly supports novel query inference (Fig. 2a), suggesting that they learned a different algorithm from their query-blind variant (except for GRU + q; see Fig. A.8).

We next seek evidence that query-conditioned models exhibits a core computational characteristic proposed by RBI: query-relevant memories are preferentially activated during inference ((Kumaran & McClelland, 2012)). We define $m_i \in M$ as query-relevant ($m_i \in M_q$) if removing it changes the oracle's posterior predictive distribution, i.e., if $KL\big(p^*(y \mid M, q) \parallel p^*(y \mid M \setminus \{m_i\}, q)\big) > 0$. We then measure how selectively the context token c_θ attends to these memories: the probability that a randomly chosen query-relevant memory ($m_i \in M_q$) receives more attention than a randomly chosen query-irrelevant one, i.e. $P\big(a(m_i) > a(m_j) \mid m_i \in M_q,\ m_j \in M \setminus M_q\big)$, where $a(m_i)$ is the attention weight from c_θ to memory m_i. 0.5 is chance and 1 is perfect separation. As a control, we also consider world-relevant memories (M_w) which reduce world support, i.e $P\big(a(m_i) > a(m_j) \mid m_i \in M_w,\ m_j \in M \setminus M_w\big)$, where $M_w = \{m_i \in M : |W(M \setminus \{m_i\})| > |W(M)|\}$. We found that 1) query-conditioned models preferentially attends to query-relevant evidence over world-relevant memories; Fig. 2b), 2) query-conditioned and query-blind variants exhibit drastically different attention patterns for query-relevant evidence: +q attention peaks at early layers (layer 2 in particular) and later subsides; while query-blind is more uniform across layers, 3) attention pattern for +q models is strikingly similar across architecture and across capacities, suggesting that they discovered shared computational principles despite architectural and capacity differences.

Moreover, we found that the pairwise cosine similarity between $m_i, m_j \in M, i \neq j$ for query-blind is much higher compared with query-conditioned variants (Fig. 2c). This means that, during query-blind computations, the representation of individual events becomes more directionally aligned while query-conditioned representations remain more directionally separable. Interestingly, this is consistent with the view that, during IE, distinct memories are encoded in relation to one another into an integrated representation, potentially reducing episodic separability. Meanwhile, maintaining more separated representations is consistent with RBI accounts in which distinct episodic memories are preserved for selective retrieval and recombination at inference (Schlichting et al., 2015; Kumaran & McClelland, 2012).

Taken together, these results suggest that, matched query-conditioned and query-blind models can achieve nearly identical choice accuracy while converging to qualitatively different computational strategies that mirror key predictions of integrative encoding and retrieval-based inference.

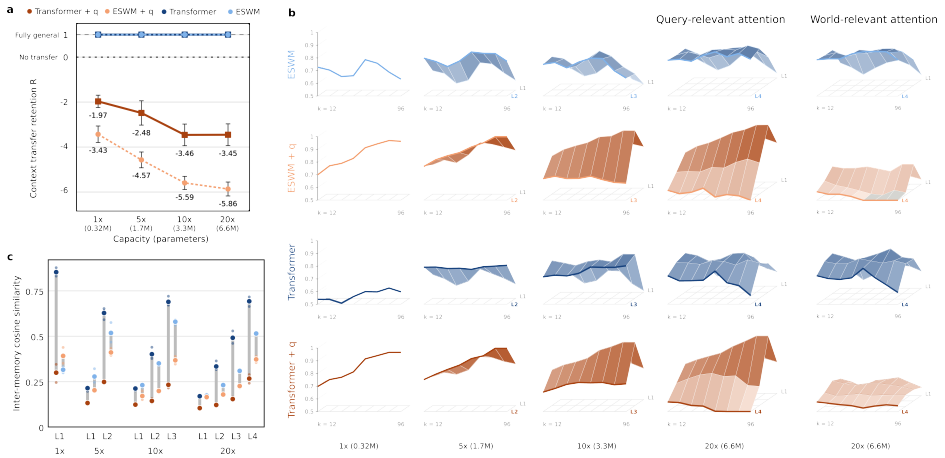

Figure 2: **Distinct learned algorithms among Retrieval-based inference (RBI) and Integrative encoding (IE) models a.** Retention score R for different models across capacities. R = 1 means c transfers perfectly to a different query; R = 0 means c is worth no more than $M = \emptyset$; R < 0 means c is worse than $M = \emptyset$. **b.** Attention AUC (see Appendix A.9) from c to query-relevant and world-relevant memories for different models, across layers and evidence set sizes. **c.** Pairwise cosine similarity between memories in M across layers for different models.

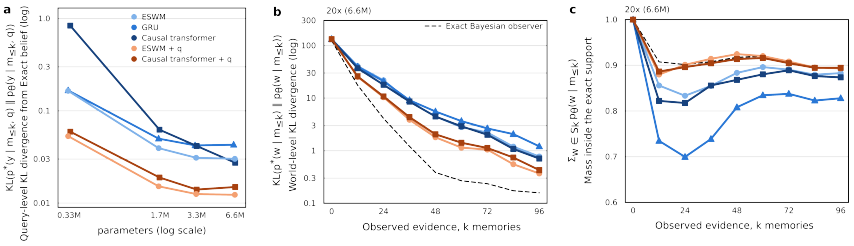

Figure 3: **Uncertainty representation at query- and world-level. a.** Model's posterior predictive against oracle for models of different capacities. **b.** Model's world-level posterior against oracle for 20× capacity models across different evidence set sizes. **c.** Mass inside support, or the sum of the model's world-level posterior probability assigned to the worlds compatible with M, for different models across different evidence set sizes.

4.2 How well do they capture uncertainty?

Similar accuracy can hide different beliefs about an uncertain world. Hippocampal theories propose that neural representations track uncertainty about hidden states (Harrison et al., 2006; Sanders et al., 2020; Mahmoodi et al., 2024), but leave open whether IE and RBI capture it in the same way. We compare query-conditioned and query-blind models with the exact oracle, measuring divergence in predictive distribution and in the world posterior implied by predictions across queries.

Despite similar top-1 accuracy, query-conditioned models are more parameter efficient and predict outcome distributions consistently closer to the oracle (Fig. 3a). Their predictions across queries also imply world posteriors closer to the oracle (Fig. 3b; Fig A7a, similar trend at every capacity) and place more probability on worlds consistent with the evidence (Fig. 3c; Fig A7b, similar trend at every capacity), showing that they are more evidence efficient, i.e. requiring less observations to reduce their uncertainty. The results highlight a dissociation between having an explicit general representation and producing query-conditioned predictions that are jointly consistent with one.

4.3 Advantages per inference types

Memory-based inference takes many forms, from pattern completion to associative inference and schema-based generalization (Teyler & Rudy, 2007; Zeithamova et al., 2012b; Shohamy & Wagner,

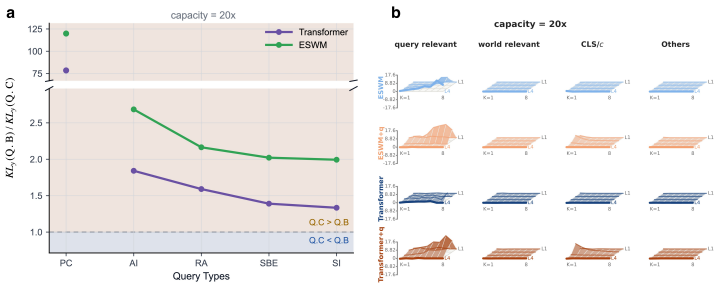

Figure 4: **Inference over query types and attention causal manipulation. a.** The ratio of KL_y between query-blind models (Q.B) and query-conditioned models (Q.C) over different query types. The orange region means Q.C is closer to the oracle compared with Q.C; the blue region the other way around. 20x capacity models are shown. **b.** Performance degradation in Top-1 accuracy over attention manipulation to different memory groups.

2008; Tse et al., 2007). We ask whether the advantage of RBI over IE depends on the inference required. We define five DSA query type analogues: pattern completion (**PC**), exact queried memory is in the evidence set (Horner et al., 2015); associative inference (**AI**), the start state and transformation cue were observed but separately (Zeithamova & Preston, 2010); rule abstraction (**RA**), the q's transformation cue has been applied to other states in memories, but not to the query's start state; structure-based exclusion (**SBE**), where q's start state is observed but the transformation cue is not; and schema-based inference (**SI**), neither q's start state nor transformation cue is observed (Tse et al., 2007; Whittington et al., 2020; Preston & Eichenbaum, 2013). SBE and SI require structural knowledge learned across worlds to be combined with in-context evidence (see AppendixA.6 for distribution).

Query-conditioned models are closer to the oracle in every class, with their largest advantage when the answer can be retrieved directly (PC; Fig. 4a). Across architectures and capacities, the gap tends to narrow in the order PC, AI, RA, SBE, SI (see A. 8). Thus, RBI benefits most from direct retrieval. Its advantage shrinks as inference depends more on reconstructing latent structure from many observations and less components of the query are directly in the evidence set.

4.4 Query-conditioned models depend on selectively retrieved memories

Query-conditioned models benefit most when an answer can be retrieved directly, and they attend strongly to query-relevant memories M_q (Fig. 2b). To test whether this attention contributes to inference, we perturb four disjoint token groups: M_q, M_w, context vector c_θ, and all other tokens. At a selected layer, we replace the retrieved content, the value vector v_i, of target token m_i, with their sample mean, removing the token m_i-specific content. We match the number of perturbed memories across memory groups.

Perturbing M_q causes the largest performance drop (Fig. 4b). Query-conditioned models are more sensitive to this perturbation than query-blind models across architectures. The effect peaks at layer 2, where attention to M_q is also highest, showing that the retrieved content contributes significantly to their predictions. This result also reveals that the query-conditioned models have disproportionately high reliance on a selective subset of memories, making it more susceptible to query-dependent memory deformation.

Attention and perturbation effects both peak early, consistent with accounts in which retrieval feeds later computation (Koster et al., 2018; Kumaran & McClelland, 2012). These results highlight a similar temporal computation pattern across query-conditioned models: retrieve query-relevant memories early and compute later.

4.5 Behavioral utility of the learned cognitive maps

Previous sections suggest that query-conditioned and query-blind models differ in their computational strategies and representations of uncertainty. We now ask whether these differences affect

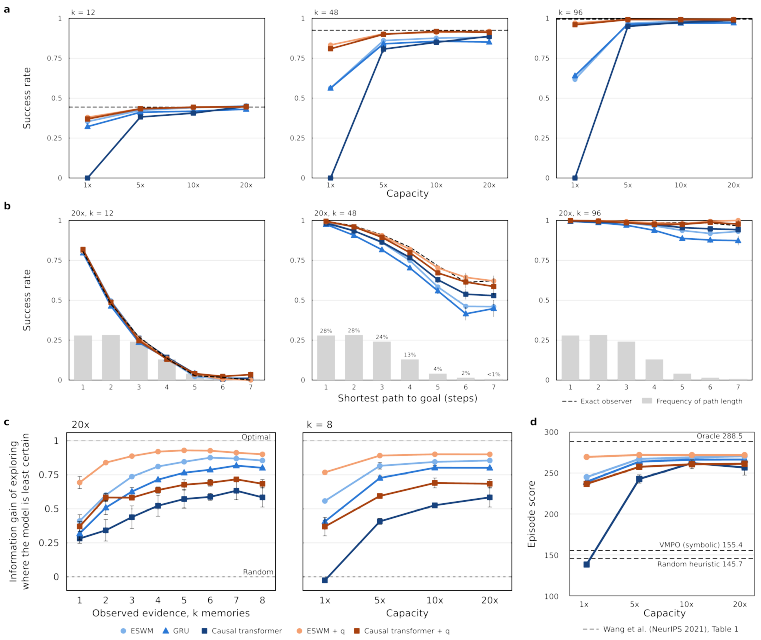

Figure 5: **Cognitive map utility a.** Planning success rate compared with exact observer (dashed line) for different models. Different columns corresponds to different evidence set sizes (12, 48, 96 memories) **b.** Success rate at 20x over the shortest path to the goal. Grey bars show distribution of paths over path lengths. Different columns are different evidence set sizes (12, 48, 96 memories). **c.** Information gained (IG) under model-selected exploration transition. IG closer to 1 means the model chooses an oracle-maximally informative transistion. Left - IG at 20× over different number of memories. Right - IG under fixed memories over different capacities. **d.** Cumulative reward in DeepMind Alchemy Benchmark reinforcement learning task for different models over model capacities.

how well their cognitive maps can support functions such as planning, uncertainty-guided exploration, and reward maximization (see Appendix A.10 for methods).

For planning, we use each model's predictions to build a transition map, search routes to the highest-reward state with breadth-first search, and execute those routes in the true world. A plan succeeds if it reaches the highest-reward state. Performance converges as model capacity increases, and the remaining gap is concentrated in rare long routes (Fig. 5a,b).

We next investigate whether the differences in uncertainty representation identified in Section 4.2 affect active exploration. Does a model that better represents uncertain transitions choose more informative ones to observe? Each model follows a max-entropy policy, choosing the transition whose outcome is most uncertain. We measure the information gained from that choice relative to the exact Bayesian observer's choice. Differences in IG across models are driven mainly by architecture rather than query conditioning (Fig. 5c).

Finally, we combine exploration and navigation and perform the DeepMind Alchemy reinforcement-learning task, which requires maximizing total reward within a limited number of steps. Both modes achieve about 94% of the oracle's episode score, compared with 54% for VMPO, the model proposed by (Wang et al., 2021) (Fig. 5d). Despite their differences in strategy and uncertainty representation, the two model types are broadly comparable, highlighting that distinct internal representations produced by query-conditioning may not produce distinct downstream behavior.

4.6 Multi-query demands favor transferable representations

The previous results establish advantages for query-conditioned computation when answering a single query. However, this comparison ignores a potential advantage of IE-like transferable representations: a query-blind representation can be computed once and reused across many future queries.

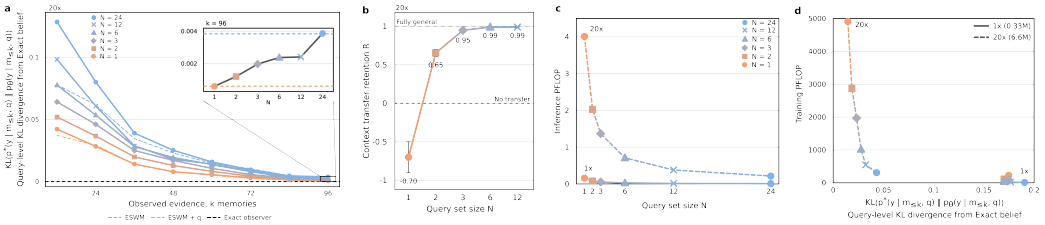

Figure 6: **Specificity versus Generality a.** Posterior predictive KL_y for N-query ESWM over different sizes of evidence set M. Smaller N achieves a better KL_y. At max $|M|$, different N occupies a continuum between ESWM's KL_y and ESWM+q's KL_y. **b.** Context retention score over different N for N-query ESWM. **c.** Total PFLOP of inference over a fixed test set for different N at 20x and 1x N-query ESWM. The y-axis is proportional to the per-query amortized inference FLOP by a constant. **d.** Total PFLOP of training on a fixed train set over the N-query model's KL_y. The y-axis is proportional to the per-query amortized training FLOP by a constant.

We therefore ask whether increasing the number of queries that a representation must support creates pressure toward IE-like computation.

We interpolate between the computational constraints of query-conditioned and query-blind models using an N-query model, where $c = \theta(M, \{q_i\}_{i=1}^N)$ must support predictions for all $\{q_i\}_{i=1}^N$ (see A.11 for schematic). When $N = 1$, the model faces the same constraint as a query-conditioned model. As N increases, c must instead support a larger set of queries. We find that the N-query model's KL smoothly interpolates between query-conditioned and query-blind models as N increases (Fig. 6a). At the same time, c becomes increasingly transferable across queries (Fig. 6b). Thus, increasing pressure to support multiple inferences naturally drives the model toward a more IE-like, reusable representation. This reuse also provides a computational advantage when many queries are made about the same hidden world. The N-query model's amortized inference cost per query decreases sharply with N (Fig. 6c), because the expensive context computation $c = \theta(M, \{q_i\}_{i=1}^N)$ is performed once and the resulting c is reused across N lightweight readouts. At $N = 24$, the amortized cost becomes comparable to that of a model $20\times$ smaller in capacity. Likewise, training cost decreases sharply with N, reaching a level comparable to the $1\times$ model while achieving substantially better uncertainty representation (Fig. 6d).

Together, these results suggest that integration timing may not correspond to two fundamentally discrete algorithms. Instead, it can emerge along a continuum determined by how narrowly or broadly a representation must be reused. Query-specific computation better preserves Bayes-optimal predictions for individual inferences, whereas pressure to support many future inferences favors increasingly general representations and amortizes their computational cost.

Conclusion. We extend the study of hippocampal inference from relational structure to uncertain causal structure, asking how integrative encoding and retrieval-based inference differ when memories provide incomplete evidence about a hidden causal world. Although the two modes can support nearly indistinguishable behavior, they rely on different internal computations and uncertainty representations: retrieval-based inference favors query-specific, selectively retrieved representations, whereas integrative encoding favors representations that generalize across queries. Our multi-query results further suggest that these are not strictly discrete strategies, but endpoints of a continuum trading query-specific accuracy against representational reuse. Together, these findings provide a computational account of how complementary memory mechanisms can support causal inference under uncertainty.

AI USE STATEMENT

In this work, we used generative AI tools to assist with editing the manuscript, and with implementing code. We have not used generative AI tools to generate experimental results, or data. We take responsibility for the final content of this work.

REFERENCES

Andrea Banino, Raphael Koster, Demis Hassabis, and Dharshan Kumaran. Retrieval-based model accounts for striking profile of episodic memory and generalization. *Scientific Reports*, 6(1): 31330, 2016.

Andrea Banino, Adrià Puigdomènech Badia, Raphael Köster, Martin J Chadwick, Vinicius Zambaldi, Demis Hassabis, Caswell Barry, Matthew Botvinick, Dharshan Kumaran, and Charles Blundell. Memo: A deep network for flexible combination of episodic memories. In *International Conference on Learning Representations*, 2020.

Helen C Barron, Raymond J Dolan, and Timothy EJ Behrens. Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature neuroscience*, 16(10):1492–1498, 2013.

Timothy EJ Behrens, Timothy H Muller, James CR Whittington, Shirley Mark, Alon B Baram, Kimberly L Stachenfeld, and Zeb Kurth-Nelson. What is a cognitive map? organizing knowledge for flexible behavior. *Neuron*, 100(2):490–509, 2018.

Marcel Binz, Ishita Dasgupta, Akshay K Jagadish, Matthew Botvinick, Jane X Wang, and Eric Schulz. Meta-learned models of cognition. *Behavioral and Brain Sciences*, 47:e147, 2024.

Cindy A Buckmaster, Howard Eichenbaum, David G Amaral, Wendy A Suzuki, and Peter R Rapp. Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. *The Journal of Neuroscience*, 24(44):9811–9825, 2004.

Neal J Cohen and Howard Eichenbaum. *Memory, amnesia, and the hippocampal system*. MIT press, 1993.

Jeffery A Dusek and Howard Eichenbaum. The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences*, 94(13):7109–7114, 1997.

Howard Eichenbaum and Neal J Cohen. Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, 83(4):764–770, 2014.

Howard Eichenbaum, Paul Dudchenko, Emma Wood, Matthew Shapiro, and Heikki Tanila. The hippocampus, memory, and place cells: is it spatial memory or a memory space? *Neuron*, 23(2): 209–226, 1999.

Dileep George, Rajeev V Rikhye, Nishad Gothoskar, J Swaroop Guntupalli, Antoine Dedieu, and Miguel Lázaro-Gredilla. Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps. *Nature communications*, 12(1):2392, 2021.

Lee M Harrison, Andrew Duggins, and Karl J Friston. Encoding uncertainty in the hippocampus. *Neural Networks*, 19(5):535–546, 2006.

Zizhan He, Maxime Daigle, and Pouya Bashivan. Building spatial world models from sparse transitional episodic memories. In *International Conference on Learning Representations*, volume 2026, pp. 21251–21286, 2026.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

Aidan J Horner, James A Bisby, Daniel Bush, Wen-Jing Lin, and Neil Burgess. Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature communications*, 6(1):7462, 2015.

Raphael Koster, Martin J Chadwick, Yi Chen, David Berron, Andrea Banino, Emrah Düzel, Demis Hassabis, and Dharshan Kumaran. Big-loop recurrence within the hippocampal system supports integration of information across episodes. *Neuron*, 99(6):1342–1354, 2018.

Dharshan Kumaran and James L McClelland. Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychological review*, 119(3):573, 2012.

Hongmi Lee and Janice Chen. Predicting memory from the network structure of naturalistic events. *Nature communications*, 13(1):4235, 2022.

Shuyi Luo, Muchen Li, Deng Pan, Naomi Kingston, Miriam Klein-Flügge, Ali Mahmoodi, and Matthew Rushworth. A frontal cortical world model shapes inference of causal relationships in hippocampus. 2026.

Ali Mahmoodi, Shuyi Luo, Caroline Harbison, Payam Piray, and Matthew FS Rushworth. Human hippocampus and dorsomedial prefrontal cortex infer and update latent causes during social interaction. *Neuron*, 112(22):3796–3809, 2024.

David Marr, David Willshaw, and Bruce McNaughton. Simple memory: a theory for archicortex. In *From the retina to the neocortex: Selected papers of David Marr*, pp. 59–128. Springer, 1991.

Bruce L McNaughton and Richard GM Morris. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in neurosciences*, 10(10):408–415, 1987.

Vladimir Mikulik, Grégoire Delétang, Tom McGrath, Tim Genewein, Miljan Martic, Shane Legg, and Pedro Ortega. Meta-trained agents implement bayes-optimal agents. *Advances in neural information processing systems*, 33:18691–18703, 2020.

Mehrab N Modi, Ashesh K Dhawale, and Upinder S Bhalla. Ca1 cell activity sequences emerge after reorganization of network correlation structure during associative learning. *Elife*, 3:e01982, 2014.

Robert J Molitor, Katherine R Sherrill, Neal W Morton, Alexandra A Miller, and Alison R Preston. Memory reactivation during learning simultaneously promotes dentate gyrus/ca2, 3 pattern differentiation and ca1 memory integration. *The Journal of Neuroscience*, 41(4):726–738, 2021.

John O'keefe and Lynn Nadel. *The hippocampus as a cognitive map*. Oxford university press, 1978.

Randall C O'Reilly and James L McClelland. Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus*, 4(6):661–682, 1994.

Pedro A Ortega, Jane X Wang, Mark Rowland, Tim Genewein, Zeb Kurth-Nelson, Razvan Pascanu, Nicolas Heess, Joel Veness, Alex Pritzel, Pablo Sprechmann, et al. Meta-learning of sequential strategies. *arXiv preprint arXiv:1905.03030*, 2019.

Alison R Preston and Howard Eichenbaum. Interplay of hippocampus and prefrontal cortex in memory. *Current biology*, 23(17):R764–R773, 2013.

Honi Sanders, Matthew A Wilson, and Samuel J Gershman. Hippocampal remapping as hidden state inference. *Elife*, 9:e51140, 2020.

Margaret L Schlichting and Alison R Preston. Memory integration: neural mechanisms and implications for behavior. *Current opinion in behavioral sciences*, 1:1–8, 2015.

Margaret L Schlichting and Alison R Preston. The hippocampus and memory integration: building knowledge to navigate future decisions. In *The hippocampus from cells to systems: Structure, connectivity, and functional contributions to memory and flexible cognition*, pp. 405–437. Springer, 2017.

Margaret L Schlichting, Dagmar Zeithamova, and Alison R Preston. Ca1 subfield contributions to memory integration and inference. *Hippocampus*, 24(10):1248–1260, 2014.

Margaret L Schlichting, Jeanette A Mumford, and Alison R Preston. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature communications*, 6(1):8151, 2015.

Daphna Shohamy and Anthony D Wagner. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron*, 60(2):378–389, 2008.

Timothy J Teyler and Pascal DiScenna. The hippocampal memory indexing theory. *Behavioral neuroscience*, 100(2):147, 1986.

Timothy J Teyler and Jerry W Rudy. The hippocampal indexing theory and episodic memory: updating the index. *Hippocampus*, 17(12):1158–1169, 2007.

Edward C Tolman. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.

Dorothy Tse, Rosamund F Langston, Masaki Kakeyama, Ingrid Bethus, Patrick A Spooner, Emma R Wood, Menno P Witter, and Richard GM Morris. Schemas and memory consolidation. *Science*, 316(5821):76–82, 2007.

Jane X Wang, Michael King, Nicolas Porcel, Zeb Kurth-Nelson, Tina Zhu, Charlie Deck, Peter Choy, Mary Cassin, Malcolm Reynolds, Francis Song, et al. Alchemy: A benchmark and analysis toolkit for meta-reinforcement learning agents. *arXiv preprint arXiv:2102.02926*, 2021.

James CR Whittington, Timothy H Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy EJ Behrens. The tolman-eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183(5):1249–1263, 2020.

Dagmar Zeithamova and Caitlin R Bowman. Generalization and the hippocampus: More than one story? *Neurobiology of Learning and Memory*, 175:107317, 2020.

Dagmar Zeithamova and Alison R Preston. Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *The Journal of Neuroscience*, 30(44):14676–14684, 2010.

Dagmar Zeithamova, April L Dominick, and Alison R Preston. Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, 75(1): 168–179, 2012a.

Dagmar Zeithamova, Margaret L Schlichting, and Alison R Preston. The hippocampus and inferential reasoning: building memories to navigate future decisions. *Frontiers in human neuroscience*, 6:70, 2012b.

A Appendix

A.1 Additional Tables

Table 2: Model configurations across capacity scales. H denotes the hidden dimension, and the head dimension is H divided by the number of attention heads. Transformer and ESWM layers contain attention and feed-forward sublayers, whereas GRU layers are stacked unidirectional recurrent layers.

Capacity	Hidden dim. H	Attn. heads	Head dim.	FFN dim.	Transformer layers	ESWM layers	GRU layers
$1\times$	128	4	32	512	1	1	2
$5\times$	256	8	32	640	2	2	3
$10\times$	320	8	40	672	3	3	4
$20\times$	416	8	52	768	4	4	5

A.2 Episode Generation details

We collect evidence set in a given causal world through a structured task. In the beginning of each trial, three stone and twelve potions, randomly selected from the world, are presented. An agent then randomly applies a potion to a stone, observing an event. The potion is removed from the available potions while the transformed stone remain. The random-policy agent repeat this process until all potions are exhausted. Hence, each trial consists of 12 events, and in total, we collect 10 trials per world (10 trials is termed an *episode*). Importantly, agent acts over discrete action (or potion) space, and all elements of the task are represented symbolically using vectors, factoring out any visual and motor complexities.

A.3 Training details

All runs are trained with cross-entropy loss on the 80-way classification. We train with a batch size of 455 episodes (effective number of sample in a batch is 98,280) over at most 5,900 steps with early stopping. We use AdamW with cosine learning rate scheduling with 295 linear warm-up steps, a peak learning rate of 3×10^{-4}, and a minimum learning rate of 0. Weight decay of 10^{-4} and gradient-norm clipping at 1.0 are applied

We train each model with 3 random seeds and report aggregated results whenever possible.

A.4 Negative log-likelihood loss on test set

Table 3: Negative log-likelihood loss on the test set. Last column is the accuracy gap at 20x, query-conditioned model $(+\,q)$ minus query-blind model. SD over the three seeds. Lower is better. Exact Bayesian observer is an oracle ceiling.

Architecture	1x (0.33M)	5x (1.7M)	10x (3.3M)	20x (6.6M)	Δ_{+q} at $20\times$
GRU	0.3617 ± 0.0054	0.2457 ± 0.0017	0.2379 ± 0.0013	0.2377 ± 0.0006	
GRU + q	0.4453 ± 0.0820	0.2435 ± 0.0017	0.2393 ± 0.0012	0.2373 ± 0.0027	-0.0004
Causal transformer	1.0295 ± 0.0012	0.2564 ± 0.0127	0.2362 ± 0.0053	0.2225 ± 0.0018	
Causal transformer + q	0.2515 ± 0.0067	0.2128 ± 0.0005	0.2081 ± 0.0002	0.2090 ± 0.0010	-0.0135
ESWM	0.3591 ± 0.0077	0.2345 ± 0.0030	0.2255 ± 0.0008	0.2248 ± 0.0006	
ESWM + q	0.2462 ± 0.0038	0.2094 ± 0.0005	0.2068 ± 0.0003	0.2064 ± 0.00004	-0.0185
Exact Bayesian observer	0.1921				

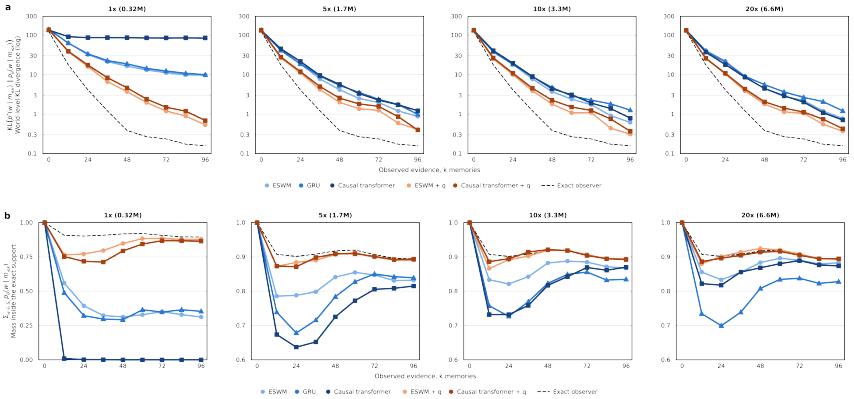

Figure 7: **Uncertainty representation at query- and world-level at every capacities a.** At world-level, RBI is more sample efficient, requiring less observed transitions to capture the same level of uncertainty. **b.** Part of the better world uncertainty representation is explained by IE models being less efficient at eliminating worlds that are not longer plausible based on the observed evidence

A.5 Uncertainty representation at every capacities

A.6 Different Query Types

Stratum	Ratio (N = 973,336)
Pattern completion (PC)	57.96%
Associative inference (AI)	29.6%
Rule abstraction (RA)	10%
Structure-based exclusion (SBE)	0.87%
Schema-based inference (SI)	0.73%

Table 4: Distribution of query types over samples ($T \geq 1$) in test splits.

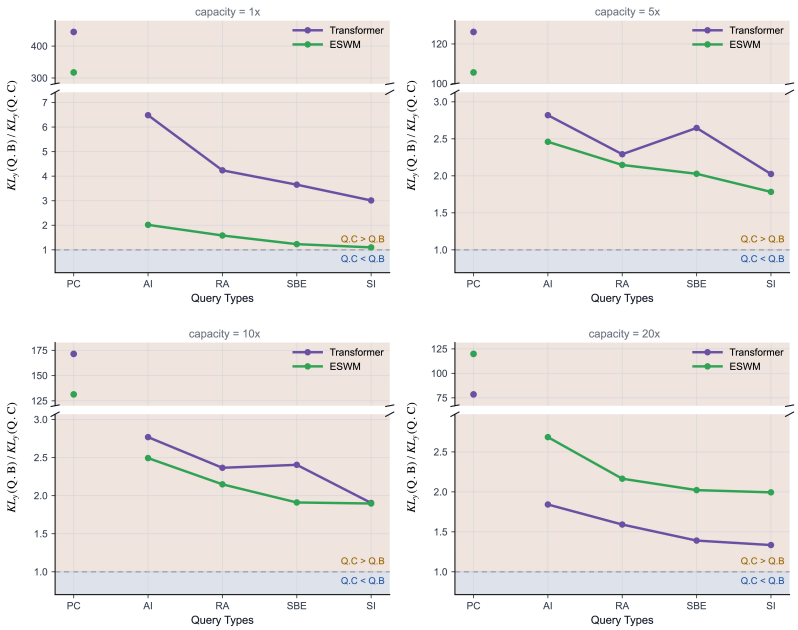

Figure 8: Additional results from Section 4.3 for models of different capacities.

A.7 Attention Value Ablation

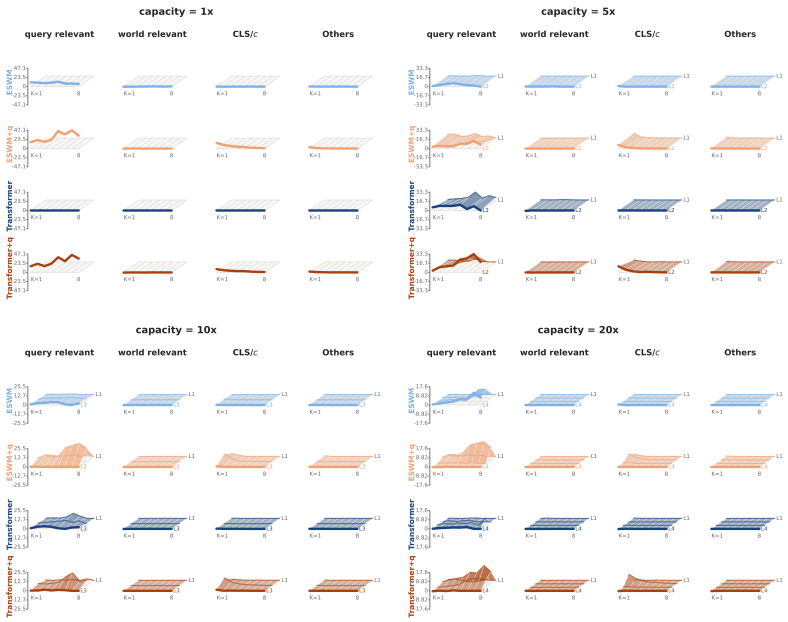

Figure 9: Additional results from Section 4.4 for models of different capacities.

A.8 Different Learned Strategies

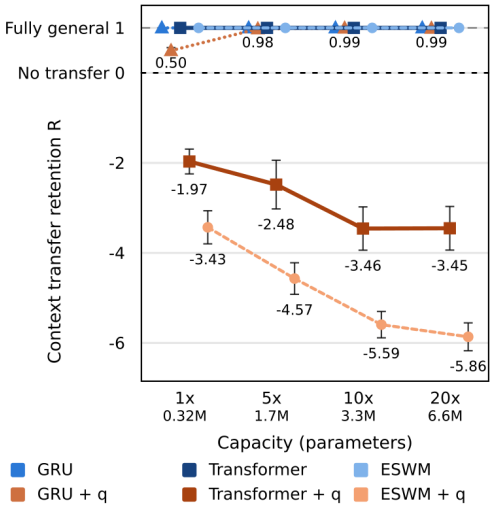

Figure 10: **Context transfer retention** R**.** All models are displayed. SD over three seeds.

Retention score R Define

$$K_{matched} = KL(p^*(y|M, q_{enc}, q_{pred})|p_\theta(y|M, q_{enc}, q_{pred})$$

when $q_{enc} = q_{pred}$ and $K_{transfer}$ be the case when $q_{enc} \neq q_{pred}$.

Let $K_{prior} = KL(p^*(y|\emptyset, q_{pred})|p_\theta(y|\emptyset, q_{pred}))$.

Then retention R is defined as:

$$R = \frac{\mathrm{K}_{\mathrm{prior}} - \mathrm{K}_{\mathrm{transfer}}}{\mathrm{K}_{\mathrm{prior}} - \mathrm{K}_{\mathrm{matched}}}$$

R = 1 means the context vector transferred perfectly. R = 0 means the transferred context vector provides no benefit over prior. R < 0 means the transferred context vector is worse than the prior, i.e, worse than no evidence.

A.9 THE ATTENTION STATISTICS

.

Attention–relevance AUC. We use the area under the receiver operating characteristic curve to measure whether the c token (q in query-conditioned models) assigns a higher attention score to memories $m_i \in M$ that are query-relevant, i.e, $m_i \in M_q$, compared with other query-irrelevant events. Importantly, this quantity is a detection AUC computed independently at each evidence budget and layer; it is not the area under a performance curve as the evidence budget varies.

Let $M_k = \{m_1, \ldots, m_k\}$ denote the k stored memory. For query q, we define the exact relevance of memory m_i by the change in the oracle posterior predictive distribution when that memory is removed:

$$r(m \mid M_k, q) = KL(p^\star(\, y \mid M_k, q) \,\|\, p^\star(\, y \mid M_k \setminus \{m\}, q))\,. \tag{1}$$

An memory is labelled relevant when $r(m \mid M_k, q) > 0$. Thus, $y_m = \mathbb{I}[r(m \mid M_k, q) > 0]$, and the attention weight $a_{\ell,m}$ at layer ℓ is treated as a score for detecting this binary label.

For the sets of query-relevant and query-irrelevant memories $\mathcal{P} = \{m : y_e = 1\}$ and $\mathcal{N} = \{m : y_e = 0\}$, respectively, the AUC is

$$\mathrm{AUC}_{\ell,k,q} = \frac{1}{|\mathcal{P}|\,|\mathcal{N}|} \sum_{m^+ \in \mathcal{P}} \sum_{m^- \in \mathcal{N}} \left[\mathbb{I}\big[a_{\ell,m^+} > a_{\ell,m^-}\big] + \frac{1}{2}\mathbb{I}\big[a_{\ell,m^+} = a_{\ell,m^-}\big] \right]. \tag{2}$$

It is the probability that a uniformly sampled query-relevant event receives more attention than a uniformly sampled query-irrelevant event, with tied scores contributing one half. Consequently, an AUC of 0.5 denotes chance ranking, 1 denotes a uniformly sampled query-relevant memory always a receive higher attention score than a uniformly sampled query-irrelevant memory, and values below 0.5 indicate a systematically reversed ranking. The chance value remains 0.5 even when relevant events are rare, which is important here because the relevance labels become increasingly sparse as the evidence budget grows.

The height-axis in Fig. 2b are the means of these values over three independent seeds. The figure places 0.5 on the base plane, so surface height represents retrieval performance above chance. AUC measures the ordering induced by attention, not the absolute attention mass assigned to relevant events.

A.10 PLANNING, EXPLORATION AND EPISODE SCORE

All three analyses of Section 4.5 are inference-only: the trained models are frozen, and each analysis reads the model's predictions through a fixed procedure.

Reading out a cognitive map. In a given world, the eight latent states are displayed as a set of eight observable states S, and there are six transformation cues P. The set of all possible queries is $Q = S \times P$, $|Q| = 48$ (start state, transformation cue) pairs. For each $q = (s, j) \in Q$, we query the model with the same evidence set M and obtain its posterior predictive $p_\theta(y \mid M, q)$ over the 80 outcomes. The model-inferred transition map is

$$\hat{T}_\theta(s, j) = \arg\max_y \, p_\theta(y \mid M, (s, j)),$$

The exact Bayesian observer's map $\hat{T}^*$ is read out the same way from $p^*(y \mid M, (s, j))$.

Planning. We use 512 held-out test episodes and evidence sets of $k \in \{12, 48, 96\}$ memories (1, 4 and 8 trials). In every world, exactly one latent state carries the highest reward. This reward is visible, so the goal state g needs no inference. The route to it is the inferred part. From each of the seven other states $s_0 \neq g$, we run breadth-first search on $\hat{T}_\theta$ for the shortest sequence from s_0 to g, with a longest path possible on the cube of seven steps. We then execute the sequence in the ground-truth world T^*, stopping as soon as g is reached. A start counts as a success if the executed sequence reaches g. It counts as a failure if the route misses g, or if $\hat{T}_\theta$ proposes no route at all. The goal is reachable from every start in every episode, with ground-truth shortest distances $d^* \in \{1, \dots, 7\}$ (mean 2.45). The success rate is averaged over the 7×512 starts, then over seeds. Fig. 5b stratifies the same starts by d^*. The exact Bayesian observer (dashed line) is scored with the identical procedure applied to $\hat{T}^*$. It is the best achievable *per-query* read-out, not a Bayes-optimal planner. A model can occasionally exceed it.

Exploration. Here the evidence set grows one memory at a time, $k \in \{1, \dots, 8\}$. Beyond one trial the exact posterior is nearly pinned (median support of 4 worlds at $k = 24$), and every exploration rule scores alike. The model explores the query it is least certain about, $q_m = \arg\max_{q \in Q} H\big(p_\theta(y \mid M, q)\big)$.

The model's choice is worth $\mathrm{IG}^*(q_m)$, where $\mathrm{IG}^*(q)$ is the exact Bayesian observer's expected information gain for query q. We normalize it between a uniformly random choice ($\bar{H}$) and the oracle's choice (q^*):

$$\mathrm{IG} = \frac{\sum_{e \in \mathcal{E}} \left[\mathrm{IG}^*_e(q_m) - \bar{H}_e\right]}{\sum_{e \in \mathcal{E}} \left[\mathrm{IG}^*_e(q^*) - \bar{H}_e\right]}, \qquad \mathcal{E} = \big\{e : \ \mathrm{IG}^*_e(q^*) > \bar{H}_e\big\},$$

where e indexes episodes and $\mathcal{E}$ keeps only the episodes in which the most informative query beats chance. $\mathrm{IG} = 1$ means the model always selects a maximally informative query, $\mathrm{IG} = 0$ means its selections are no more informative than random, and $\mathrm{IG} < 0$ means they are less informative than random.

Episode score. We replay Symbolic Alchemy as specified in Appendix A of Wang et al. (2021). An episode consists of 10 trials in one fixed world. Each trial deals 3 stones and 12 potions and allows 20 actions. The actions are either use a potion on a stone, or places a stone in the cauldron. The cauldron pays -3, -1, $+1$ or $+15$ according to the stone's latent state. The episode score is the total over the 10 trials. Worlds are sampled with the benchmark's own generative process. We score on 1000 episodes. The evidence received by the model are from a heuristic agent acting based on the model's cognitive map rather than the random-policy of the training set. The model's evidence set M i truncated to the first 96 memories (the training context length). From M, the agent reads the transition map $\hat{T}_\theta$ and the uncertainty of every query $U(s, j) = H\big(p_\theta(y \mid M, (s, j))\big)$. The heuristic policy is:

1. **Exploit.** Under $\hat{T}_\theta$, compute a plan from the currently available stones and potions that maximizes the points banked in this trial, and execute its next pouring. The plan is recomputed by dynamic programming whenever the map changes or a pouring produces an outcome other than the predicted one.
2. **Explore.** If the plan can't increases the value of the bench, consider every pair of a stone and potion available, excluding stones already worth $+15$. Perform the pair with the highest $U(s, j)$, provided $U(s, j) > 0.05$.
3. **Bank.** Otherwise, place every stone of positive value in the cauldron.

A.11 MULTI-QUERY ESWM

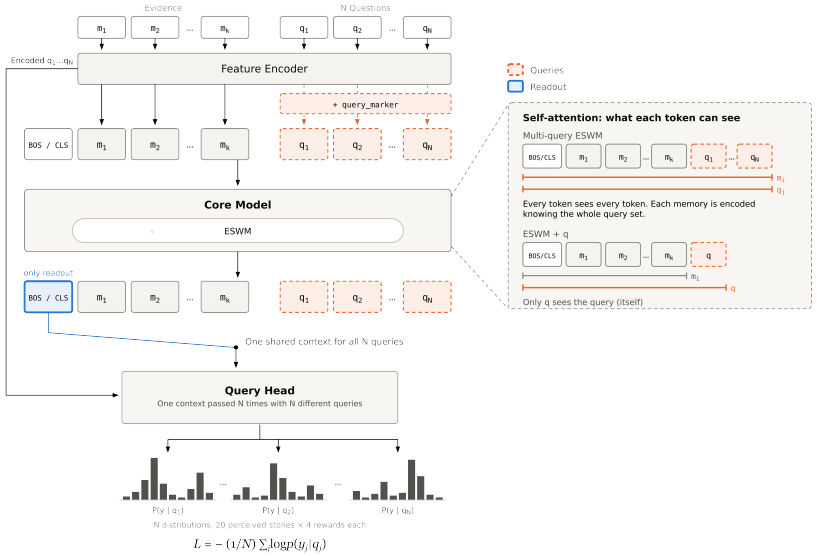

Figure 11: **Multi-query ESWM** This variant receives a fixed number of queries (N) at every forward pass. A single context readout is used to answer those N queries by passing it N times through the query head, one time with each of the N queries.