# INTERVENTIONS AS ADDRESSES: CAUSAL QUERIES AS ABSTAINING MEMORY READS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Estimators of causal structure whose parameters are shared across experimental conditions suffer from two structural vulnerabilities: catastrophic forgetting when experimental contexts arrive sequentially, and uncalibrated extrapolation when queried out of distribution or across counterfactual layers. We study an alternative, a memory of every episode stored under the experimental context that produced it, observational or which variable was set, with one rule added: it answers only where the stored evidence settles the query, and otherwise abstains. Held in a vector-symbolic memory with no trainable parameters, interventional distributions, conditioning and backdoor adjustment become reads and arithmetic. The abstention threshold is computed from the memory's own interference noise, so it needs no held-out data, and it is calibrated within a band of dimension. When experimental contexts arrive one after another, a gradient estimator forgets the earliest and the memory does not, since a sum does not depend on the order of its terms. An abstention comes for one of three reasons. A key, the address an episode is stored under, seen with two outcomes omits the variable that decides between them, and a search over the observables finds it. On Baba Is AI the search returns the rule text, and the memory then answers 0.977 of the held-out wall collisions it commits to while the rule is broken. On Alchemy it returns the displayed reward, and an agent given only the description language and the observation format reaches 0.91 of the ideal observer on held-out seeds. A key whose evidence several hypotheses still fit calls for an experiment, and expected free energy over a graded belief held beside the memory chooses it. A key naming two contexts at once can never be written, and the gate abstains from 60/60 such cross-world reads unprompted. Storing experience under its context thus gives an agent an interventional model that survives the order of its experiments, needs no retraining when a new variable is recorded, and says which of three things to do next when it cannot answer. This suggests that for an agent that learns by acting the useful form of causal knowledge may be a memory rather than a fit.

## 1 INTRODUCTION

Continuous parametric estimators of causal structure (Zheng et al., 2018; Lippe et al., 2021) have two vulnerabilities that a laboratory exposes and a shuffled dataset hides. A model fitted to a stream of experiments forgets the contexts it saw first, which is catastrophic forgetting (McClelland et al., 1995; Kirkpatrick et al., 2017) under a shift of context, and a model asked a question its data do not settle answers anyway. An agent has an alternative, which is to keep what happened and look it up. Suppose it toggles a switch $X$ and notes whether a door $Y$ opens. If it files every episode under the condition that produced it, "$X$ was set to $x$" or "nothing was set", then the fraction of episodes filed under $\mathrm{do}(X{=}x)$ in which the door opened is the interventional distribution $p(y \mid \mathrm{do}(x))$, a count of filed episodes. The observational conditional $p(y \mid x)$ is a ratio of two counts among the episodes filed under "nothing was set", and backdoor adjustment (Pearl, 2009) is a sum of such ratios. In other words, once an episode is stored under its *context*, the experimental condition it was recorded under (Mooij et al., 2020; Dawid, 2021), the dataset of experience is itself an interventional model, and the context is the address.

We hold the memory as one vector, in the vector-symbolic form (Kanerva, 2009; Plate, 1995; Komer et al., 2019) with zero trainable parameters (Fig. 1a). Every value and every context is a random



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.

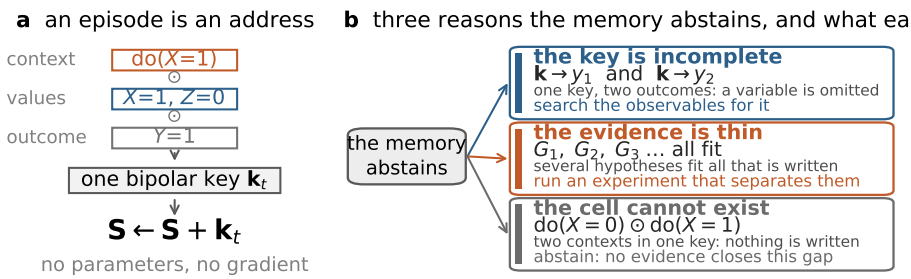


Figure 1: **A memory that abstains, and the three things an abstention can mean. a**, An episode is stored under the context that produced it, so the context is part of the address and nothing is fitted. **b**, The three reasons, taken up in turn by Sections 3 to 5.

bipolar code, an episode is the elementwise product of its codes, and a count is an inner product with the sum of all episodes. Nothing is fitted, and the only error is interference between stored episodes, which follows a law the memory computes from its own counters (Section 2). A memory answers nothing it was not shown, and instead of letting interference fill the gap it answers only where a read clears four times its own noise and otherwise *abstains*, the reject option of selective classification (Chow, 1970). What this buys is visible when experiments arrive in order. The same episodes cost a gradient estimator 2.2× on its earliest context when the contexts arrive blocked rather than shuffled, while the memory does not move (Section 2).

An abstention carries information of its own. The memory abstains for one of three reasons, and each names what the memory has to be given from outside (Fig. 1b). The first is that the key, the address an episode is stored under, is incomplete: one key has been seen with two outcomes, so a variable the key omits decides the answer, and a search over the observables can look for it. On Baba Is AI, where the rules are movable text on the grid, the search returns the rule text. The memory then answers 0.977 of the held-out wall collisions it commits to while the rule is broken, where a transformer trained on the same episodes, able to see the text and scored on the same collisions, answers 0.33 (Section 3). On Alchemy the search returns the displayed reward, and an agent given only the description language and the observation format, with its hypothesis family built from its own stored random play, reaches 0.91 of the ideal observer's return on held-out seeds. The second reason is that the evidence is thin: several hypotheses still fit what has been written, and an experiment can separate them. Expected free energy (Friston et al., 2015) over the surviving graphs chooses an intervention worth 16% of the range between the best and the worst experiment a designer who knew the truth could choose (Section 4). The third is that the cell cannot exist: a counterfactual coupling asks for a count under two contexts at once, no episode is ever written there, and the gate abstains from every such read without being told to (60/60, Section 5).

The paper is organised by that question, with related work in Appendix A. Section 2 states the memory, its laws and which causal queries are reads; Sections 3 to 5 take the three reasons in turn. Section 6 prices the memory against an exact count table, an abstaining network and a rule beside it. Every experiment but one was registered before it ran, three of them with bands taken from a pilot, and Appendix F reports the predictions that failed.

## 2 The memory, and what it answers by itself

A memory is a table from keys to what followed them, filled by observation and never fitted. For a causal query the key is a context with an assignment of variables and the value is a count; for a grid world the key is the patch around a cell and the action, and the value is the cell's next state. The table can be kept literally, which we do on grid worlds, or held as one vector, which we do everywhere else and whose costs can be written as laws. In the vector memory every *atom*, meaning every variable value, context, token or outcome that can appear in an episode, is drawn once as a random bipolar code $\mathbf{h}_a \in \{-1, +1\}^D$ (Kanerva, 2009; Plate, 1995). Binding is the elementwise product $\odot$, so an episode at step $t$ over atoms $a_1, \ldots, a_m$ has the key $\mathbf{k}_t = \mathbf{h}_{a_1} \odot \cdots \odot \mathbf{h}_{a_m}$. Writing an episode adds to the memory $\mathbf{S}$ its context-bound terms at every order up to $k_{\max}$, a unit term that counts the context, and the outcome-bound key $\mathbf{k}_t \odot \mathbf{h}_{y_t}$, where $y_t$ is the outcome that followed, and

reading a probe $\mathbf{q}$ is the inner product $\widehat{n}(\mathbf{q}) = \langle \mathbf{S}, \mathbf{q} \rangle / D$, which estimates how often that key was written. The memory is a count sketch (Charikar et al., 2002) whose keys are bindings.

What goes into a key is a choice, and the choice decides what the memory can ever tell apart. Each thing a key is written through we call a *channel*, a map $c$ from the state of the world $w$ to a value $c(w)$. The context of an episode is a channel, and so are the value of a variable, the token in a grid cell and the reward an environment displays. Writing through a set of channels $\mathcal{C}$ makes the key of $w$ the product $\mathbf{k}(w) = \bigodot_{c \in \mathcal{C}} \mathbf{h}_{c(w)}$. Two states of the world share a key exactly when they agree on every written channel, and that relation partitions the hypotheses about the world into classes the memory can never separate, however long the agent looks and however large $D$ is. A quantity the agent needs for a decision is decidable from the memory only if it is constant on every such class, and a validation metric computed through a written channel is constant on the class by construction, so it cannot detect that the class exists. In other words, when a decision varies inside a class the missing thing is a channel, and no amount of data or dimension substitutes for it. Section 3 turns this into a search.

A key seen with one outcome returns that outcome, and the answer is exact by construction. A key never seen, or seen with two outcomes, is abstained from, and the memory never guesses; the rule is a reject option (Chow, 1970) whose threshold is computed rather than fitted. In the memory every read also carries interference from the other keys, so the rule becomes a gate on the read against its own noise, and the question is how large that noise is. A key written $n$ times adds the same vector $n$ times, so it contributes $n$ times one fixed random overlap to an unrelated probe rather than $\sqrt{n}$ times. Summed over the distinct codes $c$, each written $n_c$ times, the read of a probe that was never written has mean zero and standard deviation

$$\sigma \;=\; \sqrt{\textstyle\sum_c n_c^2 \,/\, D}, \tag{1}$$

where a model of independent items would have given $\sqrt{N/D}$ for $N$ episodes (Proposition 1, Appendix B). The law can be checked exactly, since everything on its right-hand side is a counter the memory keeps. Fig. 2a shows the check on a binary confounder triangle, $Z \to X$, $Z \to Y$, $X \to Y$, written under the observational context and both values of $X$ at orders up to three: halving the dimension multiplies the measured noise by 1.413 across a 16× range, against the $\sqrt{2}$ the law predicts. Two consequences run through the paper. The memory knows its own $\sigma$, so it commits a read only where the read clears $4\sigma$, and no held-out data are needed to set that threshold; a probability read whose denominator is a context count $n$ carries the band $4\sigma/n$, and a contrast the band of its two reads. The threshold is calibrated only in a band of dimension, however, since $\sigma$ carries the interference half of the noise alone. As $D$ grows that half falls below the sampling noise and the gate over-commits: the precision of the pairs it commits (defined below) falls from 1.00 at $2^{16}$ to 0.79 at $2^{20}$ (Appendix C). At too small a $D$ almost nothing clears the gate, so $D$ is chosen from the planned counts to sit between. More episodes do not sharpen a probability read at fixed dimension either, since doubling every $n_c$ doubles both the count and $\sigma$. A key written a fixed number of times is silenced once the memory's energy grows past it, so the dimension a problem needs is chosen in advance from the counts one expects.

An episode stored under its context, observational or which variable was set to what, makes the memory an interventional model. A probability is a ratio of counts, so with $\mathbf{S}$, $\mathbf{h}_a$ and $\odot$ as above the three operations a causal query needs are

$$\widehat{p}(y \mid \mathrm{do}(x)) \;=\; \frac{\langle \mathbf{S},\ \mathbf{h}_{\mathrm{do}(x)} \odot \mathbf{h}_y \rangle}{\langle \mathbf{S},\ \mathbf{h}_{\mathrm{do}(x)} \rangle}, \tag{2}$$

$$\widehat{p}(y \mid x) \;=\; \frac{\langle \mathbf{S},\ \mathbf{h}_{\mathrm{obs}} \odot \mathbf{h}_x \odot \mathbf{h}_y \rangle}{\langle \mathbf{S},\ \mathbf{h}_{\mathrm{obs}} \odot \mathbf{h}_x \rangle}, \tag{3}$$

$$\widehat{p}(y \mid \mathrm{do}(x)) \;=\; \sum_z \widehat{p}_{\mathrm{obs}}(y \mid x, z)\, \widehat{p}_{\mathrm{obs}}(z), \tag{4}$$

the interventional distribution as a read under the context that produced it divided by that context's count, conditioning as a ratio of reads at adjacent orders, and backdoor adjustment as unbinding plus summation, every term of equation 4 being an instance of equation 3. Conditioning on $m$ variables needs stored order $m{+}1$, so a memory written to order $k_{\max}$ admits exactly the queries whose conditioning sets fit, a capacity statement rather than an expressiveness one (Proposition 3).

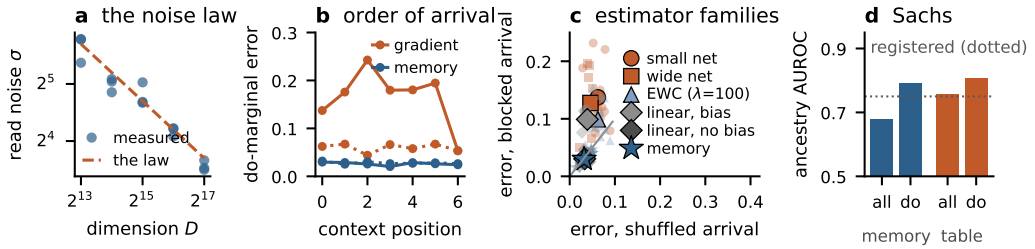


Figure 2: **The law the reads obey, and what a memory keeps that a fitted model loses. a**, Read noise $\sigma$ against dimension $D$ on a binary confounder triangle: the spread of the read error, three codebooks (points), against equation 1 from the memory's own counters (dashed). **b**, The isolation stream: error of the interventional marginals by the context's position in the stream, for a gradient estimator and the memory under blocked (solid) and shuffled (dotted) arrival, ten seeds. **c**, Earliest-context error under blocked against shuffled arrival by estimator family, ten seeds (faint) and means; the diagonal is no forgetting. **d**, Ancestry AUROC on the Sachs data for the memory and an exact count table, over all pairs and over pairs with a perturbation, against the registered criterion.

Seen this way, the identification criteria of Pearl (2009) are statements about which channels were written: a query is answered when the contexts and orders it needs are in the memory, and the memory abstains otherwise. Because an episode is an address, time does not enter a read. An interventional distribution read after up to 1000 interleaved unrelated episodes errs by at most 0.054, inside the 0.165 band the memory computes for itself, where an eligibility trace (Sutton, 1988) decays (Appendix D; the trace in Table 3).

Let us consider the stream a laboratory produces, one context after another. The isolation stream of Fig. 2b samples a random sparse graph over six binary variables under seven contexts, the observational one and both values of three intervened variables, 300 episodes each. The quantity read is the interventional marginal of every variable under every context, and the gradient estimator is a two-layer network from the one-hot context to the six marginals, trained online in one pass (Appendix D). The only manipulation is whether the contexts arrive blocked or shuffled. That manipulation costs the estimator 2.2× on the context it saw first (0.137 against 0.062, Table 1), and a replay buffer recovers the difference only as its size grows toward that of the stream (Table 2), which is a memory by another name. The memory's error on the same context moves by at most 0.028 at any position between the two orders, against a 0.21 band (Fig. 2b). One might suspect that the estimator forgets because its parameters are shared across contexts, and Fig. 2c tests that with the numbers in Table 1. A wider estimator forgets as much, elastic weight consolidation at its best reduces the gap to 1.6×, and a linear estimator forgets through its shared bias alone. The one estimator that does not forget, the same linear map without a bias, has one weight per context and marginal, which is the exact count table written as parameters. The unshared case is ENCO (Lippe et al., 2021), one parameter per ordered pair of variables, and under blocked arrival it still loses 38% of the early-context edges it recovers when shuffled ($p = 0.003$, 44 seeds), where the memory recovers 0.92 of them in either order (Appendix D). Sharing sets how much is lost, and what is lost is a parameter the objective decays while a blocked stream stops renewing it. The claim is therefore narrower than parametric against non-parametric: an estimator forgets to the extent that its parameters are shared across contexts, and the memory has no such parameter. Any context-indexed count shares that, an exact table included; what the memory adds is priced in Section 6.

Structure follows from contrasts of reads. A do-contrast, the difference between the interventional and the observational read of a second variable, identifies ancestry, and a contrast that survives conditioning on the other ancestors upgrades it to parenthood (Table 4, Appendix D). The memory commits a pair only where its source carries a context and the contrast clears $4\sigma$, and on the isolation stream the pairs it commits are true at precision 1.00 against 0.22 among the pairs it abstains from. On the one real dataset, 5,400 flow-cytometry samples over eleven phosphoproteins (Sachs et al., 2005), the registered criterion of 0.75 ancestry AUROC was missed over all pairs (0.677) and met on the 50 pairs for which a perturbation exists (0.792, Fig. 2d). Two things are in the miss: a pair with no interventional contrast falls back to a correlational score, which separates ancestry from confounding for neither estimator, so both score lower over all pairs; and interference costs the memory what the exact count table on the same channels keeps, since the table clears the criterion (0.757, Fig. 2d). The first is channel sufficiency read on real data, and the key the memory holds for

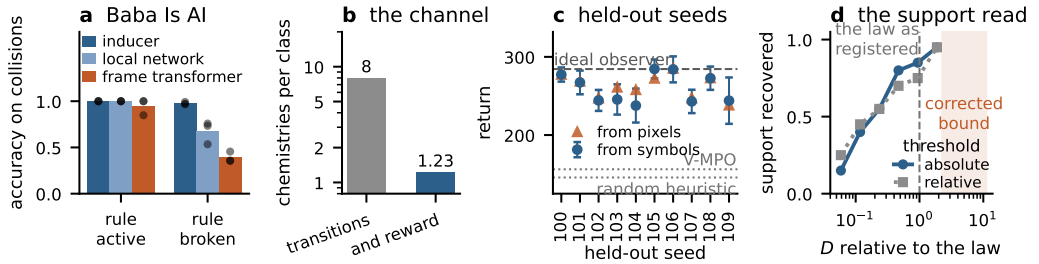


Figure 3: **A conflict in the memory directs a search for the channel the key is missing.** **a**, Baba Is AI: accuracy on held-out wall collisions while WALL IS STOP is assembled and while it is broken, random-play corpus, three levels (dots): the inducer on the collisions it commits to, and two networks trained on the same episodes scored over all collisions. **b**, Alchemy: chemistries per class, the chemistries that share one key, when only transitions are written and when the displayed reward is written too. **c**, Return on the 10 held-out seeds from symbols and from rendered stones against the published ideal observer and agents (Wang et al., 2021); bars on the symbol runs, standard error over ten episodes. **d**, The fraction of the true support a superposition read recovers against dimension relative to the registered law, with an absolute threshold and with one relative to the read's maximum; shaded, the corrected bound of Appendix B.

such a pair is incomplete. That is the first of the three reasons the memory abstains, and the next section asks what an agent can do about it.

## 3 THE KEY IS INCOMPLETE: A CONFLICT FINDS THE MISSING CHANNEL

Suppose one key has led to two outcomes. By Section 2 something the key omits decides the answer, and instead of guessing between the two, the memory reports the conflict. A search beside the memory can use that report. Its candidates are the cells of a frame, at a fixed position or at a fixed offset from a token that occurs once, or the observables of an environment, and it keeps the candidate whose value splits the conflicting outcomes and adds it to the key, with nothing retrained. The rule assumes that a complete key determines the outcome, as it does in every environment here; under aleatoric noise it would search for a separator that does not exist. On grid worlds we keep the memory as a literal table and run this search beside it, and we call the pair the *inducer*, since what the search returns is a rule.

The rules of Baba Is AI are pushable text on the grid (Cloos et al., 2024). Whether a wall stops the player depends on whether WALL IS STOP is assembled somewhere on the level, so a key made of the patch around the player and the action conflicts at every wall collision. The inducer resolves the conflict first by widening the patch, and where that fails by searching for a separator, one further cell whose value splits the two outcomes, located relative to the text that moves. On 150 episodes per level under the frontier policy the separators it validates include a WALL or STOP cell in 0.94 of the resolved keys (Appendix D), so the search returns the rule text. On held-out collisions while the rule is broken the inducer is right on 0.977 of the 0.74 of collisions it commits to, where a local network trained on the same episodes and scored over all collisions is right on 0.68 (Fig. 3a). On the collisions the inducer commits to the same network is right on 0.894. A frame transformer that can see the text reads 0.33 on the same collisions, worse where the rule is broken than the model that cannot see the text, though it reproduced its instrument on only 2 of 3 levels (Appendix D). Over every cell of the frame the networks tie the inducer, so the advantage lives on the collisions and nowhere else. The rule transfers, however: scored on five layouts it never explored the inducer reads 0.993 where it commits, and a plan through it wins every level solvable by breaking a rule on the explored layout (5 of 5, control 0, Table 5) and none elsewhere (Appendix D).

On Alchemy (Wang et al., 2021) the search can be run on the hypothesis space before any data. The environment resamples a latent chemistry every episode from a space of 1,536, and an agent earns by applying potions to stones whose value the chemistry sets. Under the transition channel alone, which every potion application writes, the space falls into classes of exactly 8 that no read separates (Fig. 3b). The value of a stone varies by 5.3 points inside a class, so the class mean is all an agent that records transitions can earn. The search selects the displayed reward first, which leaves classes

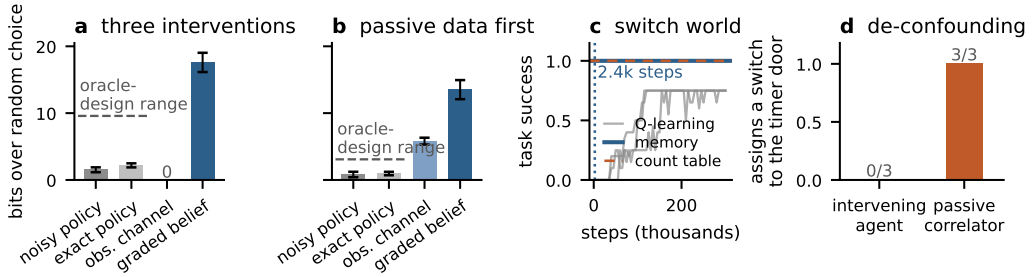


Figure 4: **Choosing the next experiment from what the memory abstains from. a**, Thirty worlds from the generator of Fig. 2b, three interventions of 300 episodes per value: bits of the true graph's surprisal removed over random choice by the expected-free-energy policy with that noise and with noise-free reads, by an observational channel, and by holding the belief as a graded posterior; dashed, the oracle-design range for the policy term; bars, paired confidence intervals over worlds. **b**, The same after 3,000 observational episodes are written first. **c**, The switch world: success on the prompted tasks against environment steps for Q-learning without a motor layer (grey, three seeds) and for the memory (solid) and an exact count table (dashed) on the shared motor layer; dotted, the steps the memory needed. **d**, Whether the door driven by a hidden timer is assigned to a switch, by the intervening agent and by a correlator fed a passive trace at matched exposure.

of about one, and with that channel written the same agent's return rises 5.5× (Appendix D). The family of hypotheses need not be given either: from 640 episodes of stored random play the agent accumulates which states co-occur and which pairs are interchanged. Closing that skeleton under the symmetry group of the description language, which is given, gives the true 1,536 chemistries with 0 spurious members (Appendix D). Given only the description language, the observation format and the support-read law below, the agent returns 260.0 ± 17.9 on 10 held-out seeds, 0.91 of the published ideal observer. From rendered stones, with the perceptual channels induced the same way, it returns 0.92, within a seed standard deviation of the symbolic result (Fig. 3c).

The set of consistent hypotheses is itself a superposition, the codes of every hypothesis consistent with each observation, and reading its support is a capacity question of its own. A hypothesis scores $D$ for every observation it satisfies plus interference whose variance is $D$ times the summed squared multiplicities, so separating "consistent with all $u$ observations" from "all but one" across $V$ hypotheses needs $D \geq 8 \ln(2V) \sum_k c_k^2$, with an absolute threshold at $(u - \frac{1}{2})D$ (Proposition 2). Below that dimension the read returns a confident subset that is interference: the support recovered rises from 0.15 at about a sixteenth of the dimension the registered law named to 0.95 at about twice it, the lower edge of the corrected bound (Fig. 3d). A threshold set relative to the read's own maximum has no operating point that can be computed in advance. When the key is complete and several hypotheses still fit, the memory cannot search its way out, and the next section asks which experiment to run.

## 4 The evidence is thin: a graded belief chooses the next experiment

Suppose now that the key is complete and the memory still abstains, because several hypotheses agree with everything written. The abstention names that set, the graphs consistent with every stored contrast, and the next intervention is a choice among ways of shrinking it. With perfect-information interventions the choice is easy, and the search rule of Section 3 matches the greedy design of Hauser & Bühlmann (2014) (Appendix D). With reads that carry noise, however, the set is the wrong object to hold the belief in. The memory commits an edge when a contrast clears $4\sigma$ and abstains otherwise, so a belief held as the set of graphs consistent with every committed decision gains nothing from an abstention. A single wrong commitment, which the gate allows at a small rate, removes the true graph with no way back. A policy that picks the intervention which would shrink that set most is then ranking by what the memory has not yet decided rather than by what the evidence says, and it should do no better than chance (Appendix F).

The belief therefore has to be graded, and it is held beside the memory rather than in it: the memory's contribution at this step is the reads and their noise law, and the choice is made by a posterior

enumerated outside. The world is a graph drawn from the generator of the isolation stream, and the agent's prior is that generator's distribution over the 3,781,503 labelled graphs on six nodes, enumerated beside the memory. An intervention sets one variable to each of its values for 300 episodes and writes them under that context. A read returns the true count plus interference drawn from equation 1 at the memory's own counts and $D = 2^{20}$, which reproduces an instantiated memory's noise to within $1.5\%$ on every seed (Appendix D), so the figures bound an ideal agent that carries the memory's noise and nothing else.

Expected free energy (Friston et al., 2015), the expected information gain of an intervention under the current posterior, ranks the variables one could set, with a budget of three on 30 worlds. The score is the surprisal of the true graph under the posterior after the budget, paired within world against random choice and set against the oracle-design range, the spread of that score over every intervention set a designer who knew the truth could choose. The policy is worth 1.522 bits, 16% of the 9.58-bit range and inside its band, and 0.70 of what it earns with noise-free reads survives the memory's noise (Fig. 4a). Holding the belief as a graded posterior instead of as a set is worth 17.589 bits on the same score, 11.6 times the policy. In other words, at this budget how the agent represents its uncertainty matters an order of magnitude more than which experiment it runs next. On the interventions the set-valued policy of Appendix F chose, a batch learner, IGSP, reaches a parent F1 of 0.94 against the memory's 0.80. The belief is enumerated, which is feasible to about seven variables; past that an approximate design such as that of Tigas et al. (2022) is needed, and none is tested here.

Writing 3,000 observational episodes first settles about three quarters of the problem before any intervention, taking the truth's surprisal from 20.53 to 4.70 bits, and of what remains the order is the same, policy below channel below belief (Fig. 4b). A positive gain is guaranteed by construction, since the world is drawn from the agent's prior and the score is what the policy minimises, so only its size and order carry information. And two registered predictions were missed: the policy gain after passive data was predicted to lie within $0.3$ bits of zero and came out at 0.817 (Fig. 4b). The calibration coverage of the posterior, 0.933, lies at the edge of its band, so calibration is left undetermined (Appendix F).

An observational memory, however large, never contains the interventional channel, and the one way to write it is to act. Let us consider an $11\times11$ switch world in which three switches open door groups through a random permutation, one effect is delayed, and one door follows only a hidden timer. An agent that toggles a switch, records each outcome window under the toggle's context, and reads the contrasts solves all twenty prompted tasks across four rooms (1.00 across 3 seeds) in 2400 environment steps on a shortest-path motor layer. So does an exact count table on the same motor layer (1.00), the control to read the memory against. Tabular Q-learning (Watkins & Dayan, 1992), given no motor layer, does not reach that within 300k training steps, which mostly measures the cost of learning to navigate; a learner over the same motor layer is not run here (Fig. 4c). The timer-driven door is the confounded case. In passive experience it correlates with whatever periodic behaviour aliases the timer, and a correlator fed a phase-locked passive trace at matched exposure assigns it to a switch in 3/3 seeds, where the intervening agent, whose own toggle times are its randomisation, assigns it to no switch in 3/3 (Fig. 4d). Acting writes the channel that separates seeing from doing, and the next section asks about the one kind of cell that no action can write.

## 5 THE CELL CANNOT EXIST: NO EVIDENCE CLOSES THE GAP

The third reason the memory abstains is a key that no episode can ever be written under, and neither a missing channel nor thin evidence explains it. Let us consider a binary treatment $X$ and a binary outcome $Y$, written under the observational context and under both settings of $X$. The world is the joint of a response type, one of the four ways $Y$ can depend on $X$, with the treatment actually assigned, eight cells in all, and every read the memory can make is linear in those eight cells (Fig. 5a). The map from cells to reads has rank 6 with a null space of dimension 2 (Theorem 5, Appendix B), so the memory's ignorance about the counterfactual layer is exactly two numbers. No dimension and no number of further episodes remove them, since the null space is a property of which contexts exist and not of how much was written under them. What lies in the row space is a linear combination of reads: the average treatment effect, and the effect of treatment on the treated, which conditions on an event fixed before the antecedent. What does not is the probability of necessity and sufficiency,

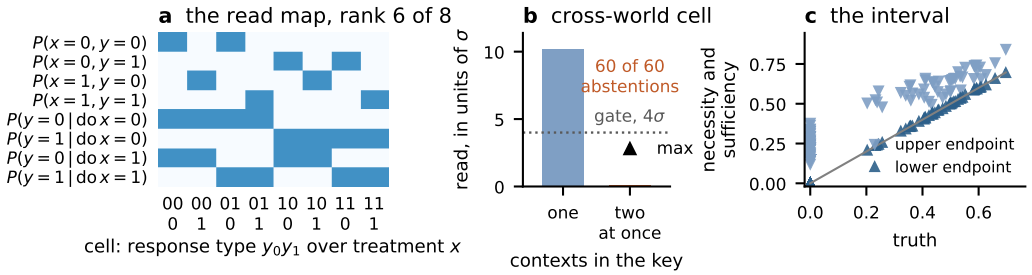


Figure 5: **Where the memory's ignorance is a property of the contexts and not of the data. a**, The read map for a binary treatment $X$ and outcome $Y$: each row a read the memory can make, each column one of the eight cells of the joint over the response type $(Y_{X=0}, Y_{X=1})$ and the treatment assigned; filled, the read counts that cell. **b**, Reads in units of the memory's own $\sigma$ on cells written under one context and on the cells a counterfactual coupling would need, against the $4\sigma$ gate; triangle, the largest cross-world read. **c**, The identified interval for the probability of necessity and sufficiency against the truth on 150 populations, isolation stream at $D = 2^{16}$; the line is equality.

which conditions on the factual outcome and is identified only to an interval, a linear program on the reads (Appendix D).

The memory abstains from the rest without any rule added for counterfactuals. A point value for necessity and sufficiency would need a count, under one context, of an event defined in another, and the key for that count carries two context codes at once. No episode is ever written under such a key, and every written key carries exactly one context code, so the probe's overlap with any written key leaves a context code unpaired. That overlap is a sum of $D$ independent signs, mean zero and sub-Gaussian by Hoeffding's inequality (Appendix B), so the gate applied to the denominator count abstains from it. Every one of the 60/60 cross-world reads sits below $4\sigma$ (Fig. 5b), which is the law's own prediction rather than a finding beyond it. Measured on the isolation stream at $D = 2^{16}$, with bands taken from a pilot on the same generator, the counterfactual layer adds nothing beyond the interventional one. The effect on the treated departs from the interventional contrast by 0.052 against a 0.834 read band (Appendix D), and the necessity-and-sufficiency interval is 0.164 wide with the truth 0.003 above its lower endpoint (Fig. 5c), on families too monotone to exercise the width; on a mixed-sign family it is 0.204 wide and the truth sits at most 0.041 above the endpoint (Appendix F). Where the memory answers a counterfactual query the answer was already an interventional read, and where it abstains from one, no evidence closes the gap. The next section asks what holding the memory costs and where, with all three answers supplied, it stops.

## 6 What the memory costs, and where it stops

Three things bound what the paper claims for the memory. The first is its size. The memory holds 1,631 distinct keys in 65,536 dimensions, 40.2 components for every key an exact count table would hold as one row (Appendix D). On the family of Fig. 6a its footprint falls below the table's only from 20 binary variables on, where its read error is 0.034 against the table's 0.0065 and above its registered 0.01, and on the Sachs data the exact count table scores above the memory in 6.1× less space (Fig. 2d, Appendix D). The comparison is generous to the memory, which is written at order one against a table of full assignments: a table of order-one marginals alone is smaller at every $N$, and a memory written at the order the queries of Section 2 need carries more energy and needs a larger $D$. An exact count table is therefore the better estimator on every problem here, and no setting in this paper has the memory both smaller and within its band. What the table lacks is the algebra: it reaches its footprint by enumeration, so it holds no binding that forms a key it has not stored and no superposition that carries a belief. The superposition carried the Alchemy belief of Section 3; the graph posterior of Section 4 was enumerated outside the memory.

The second is that exactness is not an advantage where the key covers the answer. On Conway's Life, falling sand and a walls world with actions, the next state of a cell is a function of its radius-one patch. The inducer is therefore exact wherever it commits, and the walls world's planted trap, one cell whose next state overrides the rule, costs 11 wrong cells, all before the first witnessed deviation, a count inside the bound of Proposition 4 (Appendix D). However, a local network over the same

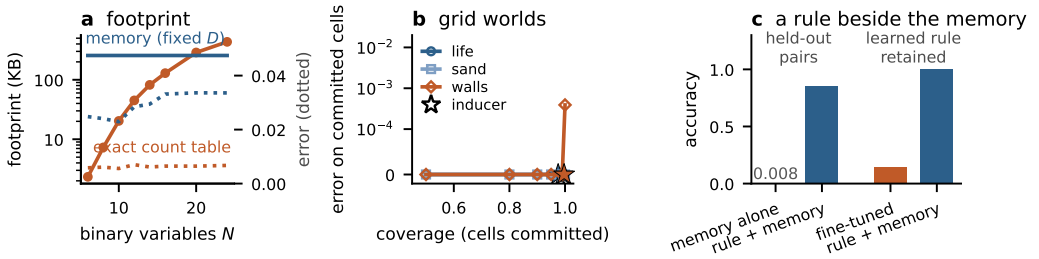


Figure 6: **Footprint, exactness where the key covers the answer, and a rule beside the memory.** **a**, Footprint of the memory at fixed $D$ and of an exact count table against the number of binary variables $N$ at 19,999 episodes; dotted, the error of the interventional reads. **b**, Conway's Life, falling sand and a walls world: error on the cells a local network commits to against its coverage; stars, the inducer's operating point; life and sand lie on zero at every coverage under the walls curve. **c**, Held-out accuracy on a walled gridworld, the memory alone against a given rule with the memory holding its exceptions; and accuracy on the regular transitions of a wormhole world after the exceptions arrive, a learned rule fine-tuned on them against the same rule kept beside a memory of them (ten seeds).

patch and action, trained on the same episodes and abstaining when its confidence is low (Chow, 1970; El-Yaniv & Wiener, 2010), matches the inducer at the inducer's own coverage on all three worlds, with every committed cell right (Fig. 6b), against a registered prediction. In other words, where the corpus covers every key, a network that knows when to abstain is as exact as a memory that abstains. The memory does better where the key is incomplete, since the conflict it reports is what the search of Section 3 used, and under blocked arrival, where the network forgets what the memory keeps.

The third is that a memory answers nothing it was not shown, and what answers the rest is a rule that composes beside it. Write $f$ for a rule over the same keys, given or learned, which proposes an outcome $f(\mathbf{k})$ for any key including keys never written. With $\mathbf{k}_t$, $y_t$, $\odot$ and $\sigma$ as in Section 2, the whole arrangement is two lines,

$$\mathbf{S} \leftarrow \mathbf{S} + \mathbf{k}_t \odot \mathbf{h}_{y_t} \qquad \text{only if } y_t \neq f(\mathbf{k}_t), \tag{5}$$

$$\widehat{y}(\mathbf{k}) = \arg\max_{y} \langle \mathbf{S}, \mathbf{k} \odot \mathbf{h}_y \rangle / D \ \text{ if it clears } 4\sigma, \qquad \text{else } f(\mathbf{k}), \tag{6}$$

so the memory holds only what $f$ gets wrong, every key it was never written under is answered by $f$, and no gradient passes between the two. The rule assumes deterministic transitions; under stochastic ones the exceptions fill with the minority outcome. On a walled gridworld the memory alone answers held-out state–action pairs at chance (0.008 against 0.008), and with a displacement rule given as $f$ the pair reaches 0.852 (Fig. 6c). When the rule is learned rather than given, a network that learns the regular transitions and then meets the exceptions in sequence loses the rule (0.143 of regular transitions still right), while the same network kept beside a memory of the exceptions retains it (0.999). What no arrangement here supplies is a goal never witnessed: a planner over a memory cannot reach a state it has not seen, since no episode carries the goal. The reason is not that rollouts through witnessed keys abstain. The levels of Baba Is AI that require a rule to be made rather than broken are therefore lost (0 of 29, Table 5) under every policy tried.

## 7 What remains open

What the paper offers is a set of laws and a frame rather than an estimator. On every problem here an exact count table is the better estimator, and the memory is the object to reach for above the footprint crossover, or where a binding must form a key never stored or a superposition must carry a belief; its gate is calibrated only in a band of dimension (Appendix C), and the search and the rule beside it assume deterministic transitions. What the memory does not supply at all is a goal it has never witnessed: the make-a-rule levels and the ARC-AGI-3 games are lost under every policy tried, including one that read the memory's own abstentions, since an abstention names an unseen key rather than a rule (Appendices E and F). Whether abstention can direct exploration once the goal itself is a written channel is the next experiment, and a memory makes it cheap, since nothing in it is retrained.

### Reproducibility statement

An anonymised repository accompanies the submission, and every result can be regenerated from it at three levels, each one command from its root. The levels are the manuscript from the shipped evidence directory and its hash manifest, the laws of Section 2 as exact tests on the algebra, and each experiment as a module with its registration. Appendix F names the files, the commands and the integrity suites, and reports every registered prediction that failed; Appendix D gives the settings of every experiment; Appendix B states and proves the laws.

### AI use statement

In this work, we used generative AI tools to help develop the conceptual framework, to formulate mathematical claims and assist in the writing of proofs, to propose and refine hypotheses, to design and give feedback on experiments, to implement methods, and to interpret results. No generative AI tool produced any data set: the synthetic worlds are sampled by procedural generators, whose code was written with AI assistance and is shipped, and every other data set is taken from its published source. Translation, cleaning and reformatting of data sets and qualitative or thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to create and modify scientific figures, to create and edit software code, to draft and edit parts of the paper, to identify references, and to suggest the paper's structure and title. We have reviewed all AI-assisted work. Every experiment but the least-squares control was registered before it ran, with its prediction, effect size and refutation criterion fixed by the authors, and adjudicated against that registration by the authors. AI-written code is covered by the test suite named in the reproducibility statement, and the laws of Section 2 are enforced as exact tests on the algebra. Every number in the manuscript enters through a generated macro from a hash-manifested evidence directory, and every entry of the reference list is fetched from its Crossref, OpenAlex or arXiv record. The authors read and edited every section. We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

## References


Anastasios N. Angelopoulos and Stephen Bates. *Conformal Prediction: A Gentle Introduction*. Now Publishers, 2023. doi: 10.1561/9781638281597.

Burton H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7):422–426, 1970. doi: 10.1145/362686.362692.

Charles Blundell, Benigno Uria, Alexander Pritzel, Yazhe Li, Avraham Ruderman, Joel Z Leibo, Jack Rae, Daan Wierstra, and Demis Hassabis. Model-Free Episodic Control, 2016. URL https://arxiv.org/abs/1606.04460.

Philippe Brouillard, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin. Differentiable Causal Discovery from Interventional Data, 2020. URL https://arxiv.org/abs/2007.01754.

Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding Frequent Items in Data Streams. In *Lecture Notes in Computer Science*, pp. 693–703. Springer Berlin Heidelberg, 2002. doi: 10.1007/3-540-45465-9_59.

C. Chow. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, 16(1):41–46, 1970. doi: 10.1109/tit.1970.1054406.

Nathan Cloos, Meagan Jens, Michelangelo Naim, Yen-Ling Kuo, Ignacio Cases, Andrei Barbu, and Christopher J. Cueva. Baba Is AI: Break the Rules to Beat the Benchmark, 2024. URL https://arxiv.org/abs/2407.13729.

Ria Das, Joshua B. Tenenbaum, Armando Solar-Lezama, and Zenna Tavares. Combining Functional and Automata Synthesis to Discover Causal Reactive Programs. *Proceedings of the ACM on Programming Languages*, 7(POPL):1628–1658, 2023. doi: 10.1145/3571249.

Philip Dawid. Decision-theoretic foundations for statistical causality. *Journal of Causal Inference*, 9(1):39–77, 2021. doi: 10.1515/jci-2020-0008.

Yan Duan, John Schulman, Xi Chen, Peter L. Bartlett, Ilya Sutskever, and Pieter Abbeel. $RL^2$: Fast Reinforcement Learning via Slow Reinforcement Learning, 2016. URL https://arxiv.org/abs/1611.02779.

Miroslav Dudik, John Langford, and Lihong Li. Doubly Robust Policy Evaluation and Learning, 2011. URL https://arxiv.org/abs/1103.4601.

Ran El-Yaniv and Yair Wiener. On the Foundations of Noise-free Selective Classification. *Journal of Machine Learning Research*, 11(53):1605–1641, 2010.

Kevin Ellis, Catherine Wong, Maxwell Nye, Mathias Sablé-Meyer, Lucas Morales, Luke Hewitt, Luc Cary, Armando Solar-Lezama, and Joshua B. Tenenbaum. DreamCoder: bootstrapping inductive program synthesis with wake-sleep library learning. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, pp. 835–850. ACM, 2021. doi: 10.1145/3453483.3454080.

Richard Evans and Edward Grefenstette. Learning Explanatory Rules from Noisy Data. *Journal of Artificial Intelligence Research*, 61:1–64, 2018. doi: 10.1613/jair.5714.

ARC Prize Foundation. ARC-AGI-3: A New Challenge for Frontier Agentic Intelligence, 2026. URL https://arxiv.org/abs/2603.24621.

Karl Friston, Francesco Rigoli, Dimitri Ognibene, Christoph Mathys, Thomas Fitzgerald, and Giovanni Pezzulo. Active inference and epistemic value. *Cognitive Neuroscience*, 6(4):187–214, 2015. doi: 10.1080/17588928.2015.1020053.

William Gilpin. Cellular automata as convolutional neural networks. *Physical Review E*, 100(3), 2019. doi: 10.1103/physreve.100.032402.

Alain Hauser and Peter Bühlmann. Characterization and Greedy Learning of Interventional Markov Equivalence Classes of Directed Acyclic Graphs. *Journal of Machine Learning Research*, 13(1): 2409–2464, 2011.

Alain Hauser and Peter Bühlmann. Two optimal strategies for active learning of causal models from interventional data. *International Journal of Approximate Reasoning*, 55(4):926–939, 2014. doi: 10.1016/j.ijar.2013.11.007.

Pentti Kanerva. Hyperdimensional Computing: An Introduction to Computing in Distributed Representation with High-Dimensional Random Vectors. *Cognitive Computation*, 1(2):139–159, 2009. doi: 10.1007/s12559-009-9009-8.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. doi: 10.1073/pnas.1611835114.

Denis Kleyko, Dmitri Rachkovskij, Evgeny Osipov, and Abbas Rahimi. A Survey on Hyperdimensional Computing aka Vector Symbolic Architectures, Part II: Applications, Cognitive Models, and Challenges. *ACM Computing Surveys*, 55(9):1–52, 2023a. doi: 10.1145/3558000.

Denis Kleyko, Dmitri A. Rachkovskij, Evgeny Osipov, and Abbas Rahimi. A Survey on Hyperdimensional Computing aka Vector Symbolic Architectures, Part I: Models and Data Transformations. *ACM Computing Surveys*, 55(6):1–40, 2023b. doi: 10.1145/3538531.

Brent Komer, Terrence C. Stewart, Aaron R. Voelker, and Chris Eliasmith. A neural representation of continuous space using fractional binding. *Cognitive Science*, pp. 2038–2043, 2019.

Máté Lengyel and Peter Dayan. Hippocampal Contributions to Control: The Third Way. In *Neural Information Processing Systems*, volume 20, pp. 889–896, 2007.

Phillip Lippe, Taco Cohen, and Efstratios Gavves. Efficient Neural Causal Discovery without Acyclicity Constraints, 2021. URL https://arxiv.org/abs/2107.10483.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient Episodic Memory for Continual Learning. In *Neural Information Processing Systems*, volume 30, pp. 6467–6476, 2017.

James L. McClelland, Bruce L. McNaughton, and Randall C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995. doi: 10.1037/0033-295x.102.3.419.

Joris M. Mooij, Sara Magliacane, and Tom Claassen. Joint Causal Inference from Multiple Contexts. *Journal of Machine Learning Research*, 21(99):1–108, 2020.

Steven Morad, Ryan Kortvelesy, Matteo Bettini, Stephan Liwicki, and Amanda Prorok. POPGym: Benchmarking Partially Observable Reinforcement Learning, 2023. URL https://arxiv.org/abs/2303.01859.

Stephen Muggleton. Inductive logic programming. *New Generation Computing*, 8(4):295–318, 1991. doi: 10.1007/bf03037089.

Alexander Nikulin, Vladislav Kurenkov, Ilya Zisman, Artem Agarkov, Viacheslav Sinii, and Sergey Kolesnikov. XLand-MiniGrid: Scalable Meta-Reinforcement Learning Environments in JAX. In *Advances in Neural Information Processing Systems 37*, pp. 43809–43835. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024. doi: 10.52202/079017-1390.

Judea Pearl. *Causality*. Cambridge University Press, 2009. doi: 10.1017/cbo9780511803161.

T.A. Plate. Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6(3): 623–641, 1995. doi: 10.1109/72.377968.

Doina Precup, Richard S. Sutton, and Satinder Singh. Eligibility Traces for Off-Policy Policy Evaluation. In *International Conference on Machine Learning*, pp. 759–766, 2000.

Alexander Pritzel, Benigno Uria, Sriram Srinivasan, Adrià Puigdomènech, Oriol Vinyals, Demis Hassabis, Daan Wierstra, and Charles Blundell. Neural Episodic Control. In *International Conference on Machine Learning*, pp. 2827–2836, 2017.

David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy P. Lillicrap, and Greg Wayne. Experience Replay for Continual Learning, 2018. URL https://arxiv.org/abs/1811.11682.

Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. *Science*, 308(5721): 523–529, 2005. doi: 10.1126/science.1105809.

Richard S. Sutton. Learning to Predict by the Methods of Temporal Differences. *Machine Learning*, 3(1):9–44, 1988. doi: 10.1023/a:1022633531479.

Panagiotis Tigas, Yashas Annadani, Andrew Jesson, Bernhard Schölkopf, Yarin Gal, and Stefan Bauer. Interventions, Where and How? Experimental Design for Causal Models at Scale. In *Advances in Neural Information Processing Systems 35*, pp. 24130–24143. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2022. doi: 10.52202/068431-1752.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. 2005. doi: 10.1007/b106715.

Jane X. Wang, Michael King, Nicolas Porcel, Zeb Kurth-Nelson, Tina Zhu, Charlie Deck, Peter Choy, Mary Cassin, Malcolm Reynolds, Francis Song, Gavin Buttimore, David P. Reichert, Neil Rabinowitz, Loic Matthey, Demis Hassabis, Alexander Lerchner, and Matthew Botvinick. Alchemy: A benchmark and analysis toolkit for meta-reinforcement learning agents. In *Neural Information Processing Systems*, 2021.

Yuhao Wang, Liam Solus, Karren Dai Yang, and Caroline Uhler. Permutation-based Causal Inference Algorithms with Interventions, 2017. URL https://arxiv.org/abs/1705.10220.

Archana Warrier, Dat Nguyen, Michelangelo Naim, Moksh Jain, Yichao Liang, Karen Schroeder, Cambridge Yang, Joshua B. Tenenbaum, Sebastian Vollmer, Kevin Ellis, and Zenna Tavares. Benchmarking World-Model Learning with Environment-Level Queries, 2025. URL https://arxiv.org/abs/2510.19788.

Christopher J. C. H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, 1992. doi: 10.1007/bf00992698.

Kilian Weinberger, Anirban Dasgupta, John Langford, Alex Smola, and Josh Attenberg. Feature hashing for large scale multitask learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 1113–1120. ACM, 2009. doi: 10.1145/1553374.1553516.

N. H. Wulff and John Hertz. Learning Cellular Automaton Dynamics with Neural Networks. *neural information processing systems*, 5:631–638, 1992.

Xun Zheng, Bryon Aragam, Pradeep Ravikumar, and Eric P. Xing. DAGs with NO TEARS: Continuous Optimization for Structure Learning. In *Neural Information Processing Systems*, volume 31, pp. 9472–9483, 2018.

Luisa Zintgraf, Kyriacos Shiarlis, Maximilian Igl, Sebastian Schulze, Yarin Gal, Katja Hofmann, and Shimon Whiteson. VariBAD: A Very Good Method for Bayes-Adaptive Deep RL via Meta-Learning, 2019. URL https://arxiv.org/abs/1910.08348.

## A Related work

The memory is a random-projection superposition of the kind the sketching literature studies. It is closest to a count sketch (Charikar et al., 2002), with the difference that identical episodes add the same vector and so interfere coherently, where a Bloom filter (Bloom, 1970) answers membership with one-sided error and the hashing trick (Weinberger et al., 2009) projects sparse features into a fixed dimension. Hyperdimensional computing built a cognitive architecture on the same binding and bundling operations (Kanerva, 2009; Plate, 1995; Komer et al., 2019; Kleyko et al., 2023b;a), and episodic control (Lengyel & Dayan, 2007; Blundell et al., 2016; Pritzel et al., 2017) replaced value learning with lookup. This paper prices that line, giving the coherent form of the read noise, what a stored order admits, what a set of written channels can identify, and what a support read costs, each as a law with an exact test (`tests/test_causal_memory.py`). It adds the observation that Pearl's interventional layer (Pearl, 2009) is reads and arithmetic on such a memory once the context is part of the key.

Abstention is the reject option of Chow (1970), and selective classification (El-Yaniv & Wiener, 2010) trades coverage for error in the same way. The memory's gate thresholds a read against interference it computes from its own counters rather than against a fitted confidence, which is why it needs no held-out data and why it abstains from the cross-world cell without a rule for counterfactuals. Conformal prediction (Vovk et al., 2005; Angelopoulos & Bates, 2023) obtains its guarantee from exchangeable held-out scores instead, which is the calibration the gate does without and the reason it is calibrated only in a band of dimension. Storing an episode under the context that produced it is the logged-propensity setting of off-policy evaluation (Precup et al., 2000; Dudik et al., 2011), with the context code in the place of the logging policy, except that the memory reweights nothing and reads the context's own count. Episodic control is the closest lookup system and is not run here, since it looks up states for value rather than contexts for counts. Networks have learned local rules on grids since Wulff & Hertz (1992), a cellular automaton is a small convolutional network (Gilpin, 2019), and Section 6 confirms that an abstaining network is as exact as the memory where the corpus covers the keys. Inductive logic programming (Muggleton, 1991; Evans & Grefenstette, 2018), library-learning synthesis (Ellis et al., 2021) and AutumnBench's synthesiser (Das et al., 2023) induce rules with a grammar in hand, where the inducer's search has only the cells of the frame and the conflict that directs it.

Continuous-optimisation causal discovery (Zheng et al., 2018; Brouillard et al., 2020; Lippe et al., 2021) fits a graph and a mechanism, and permutation-based methods with interventions (Wang et al., 2017), greedy interventional equivalence search (Hauser & Bühlmann, 2011) and greedy experimental design (Hauser & Bühlmann, 2014) are the batch and the noise-free references used here;

Bayesian experimental design at scale (Tigas et al., 2022) is the route past the enumeration Section 4 relies on. In batch the memory sits below the best of them, and its contribution is what survives the order in which experiments arrive. Meta-reinforcement learning on Alchemy (Wang et al., 2021; Duan et al., 2016; Zintgraf et al., 2019) amortises inference into weights, where the agent of Section 3 trains nothing and induces its hypothesis family from stored play, so the comparison is with the ideal observer rather than with the trained agents. Catastrophic forgetting and its remedies (McClelland et al., 1995; Kirkpatrick et al., 2017; Rolnick et al., 2018; Lopez-Paz & Ranzato, 2017) are the frame for Section 2: replay recovers the blocked stream by becoming a memory, and consolidation narrows the gap without closing it, because what forgets is a parameter that the stream stops renewing.

## B THE LAWS, STATED AND PROVED

**The memory.** The memory's costs are laws rather than fitted curves, and this appendix states each one and proves it. Write $\mathcal{A}$ for the alphabet of atoms, one per variable value and one per context. Each atom $a$ is drawn once as a code $\mathbf{h}_a \in \{-1,+1\}^D$ with independent Rademacher entries, so that $\langle \mathbf{h}_a, \mathbf{h}_a \rangle = D$ and $\mathbb{E}\langle \mathbf{h}_a, \mathbf{h}_b \rangle = 0$ for $a \neq b$. A key is the elementwise product of its atoms, which is again bipolar, and an episode adds its key to one vector,

$$\mathbf{S} \;=\; \sum_{t=1}^{N} \mathbf{k}_t \;=\; \sum_{c \in \mathcal{C}} n_c \, \mathbf{c}, \qquad \sum_c n_c = N, \tag{7}$$

where $\mathcal{C}$ is the set of distinct keys and $n_c$ the number of times key $c$ was written. A count is an inner product, $\widehat{n}(\mathbf{q}) = \langle \mathbf{S}, \mathbf{q} \rangle / D$, and a probability is a ratio of counts. Nothing is fitted, so the only error is interference between keys.

**Proposition 1** (Coherent interference)**.** *For a probe* $\mathbf{q}$ *drawn independently of the stored keys, the read* $\widehat{n}(\mathbf{q})$ *has mean zero and standard deviation*

$$\sigma \;=\; \sqrt{\frac{\sum_{c \in \mathcal{C}} n_c^2}{D}}. \tag{8}$$

*For a probe equal to a stored key* $c^\star$ *the read has mean* $n_{c^\star}$ *and the same standard deviation with* $c^\star$ *omitted from the sum.*

*Proof.* Write $W_c = \langle \mathbf{c}, \mathbf{q} \rangle = \sum_{d=1}^{D} c_d q_d$. The products $c_d q_d$ are independent Rademacher variables with mean zero and unit variance, so $\mathbb{E}[W_c] = 0$ and $\mathrm{Var}(W_c) = D$. Distinct keys are built from distinct products of independent atoms, so $W_c$ and $W_{c'}$ are uncorrelated for $c \neq c'$. By equation 7 the read is $\widehat{n}(\mathbf{q}) = D^{-1} \sum_c n_c W_c$, whence $\mathrm{Var}(\widehat{n}) = D^{-2} \sum_c n_c^2 \, \mathrm{Var}(W_c) = D^{-1} \sum_c n_c^2$. The multiplicity enters squared because the $n_c$ writings of one key add the same realised vector $\mathbf{c}$, so their contributions to the noise are in phase, as they are in the signal. Replacing $\mathbf{q}$ by a stored key $c^\star$ adds the term $n_{c^\star} D$ to the numerator and removes $c^\star$ from the sum. □

Two things follow that the memory uses. The noise is computable from the memory's own counters, since $\sum_c n_c^2$ is a number it keeps, so the gate of Section 2 needs no held-out data. And the relative noise on a probability read, $\sigma$ over the count it divides by, is unchanged by writing more episodes at fixed context ratios: doubling every $n_c$ doubles both. Test power at fixed dimension therefore does not improve with data, which is the capacity clause of Section 2, and only $D$ moves it.

The law is checked against a read's band at three orders (`tests/test_causal_memory.py`), and by the dimension sweep of Section 2: $1.413\times$ per halving of $D$ against the $\sqrt{2}$ that equation 8 predicts, and nine digits of agreement when equation 8 is fed the memory's counters. The independent-items form $\sqrt{N/D}$ departs from equation 8 whenever any key repeats.

**Proposition 2** (Support read)**.** *Let a belief superpose, for each of* $u$ *observations, the codes of every hypothesis consistent with it, so that* $\mathbf{M} = \sum_k c_k \mathbf{h}_k$ *with* $c_k$ *the number of observations hypothesis* $k$ *satisfies. A read* $s_j = \langle \mathbf{h}_j, \mathbf{M} \rangle$ *has mean* $c_j D$ *and variance* $D \sum_{k \neq j} c_k^2$*. Thresholding absolutely at* $(u - \frac{1}{2})D$ *separates the hypotheses consistent with all* $u$ *observations from the rest, with error at most one half, as soon as*

$$D \;\geq\; 8 \ln(2V) \sum_k c_k^2, \tag{9}$$

*where $V$ is the number of hypotheses.*

*Proof.* The mean and variance are Proposition 1 applied to $\mathbf{M}$. Write $\mathcal{I}_j = s_j - c_j D$ for the interference on hypothesis $j$. Conditioned on $\mathbf{h}_j$ each $\langle \mathbf{h}_j, \mathbf{h}_k \rangle$ is a sum of $D$ independent Rademacher variables, and the $k$ are independent of one another, so $\mathcal{I}_j$ is sub-Gaussian with variance proxy $D \sum_k c_k^2$ and its tail beyond $D/2$ is at most $\exp(-D/(8 \sum_k c_k^2))$. A hypothesis consistent with all $u$ observations has mean $uD$ and is lost only by falling below the threshold; one consistent with at most $u - 1$ of them has mean at most $(u - 1)D$ and is admitted only by rising above it. Each hypothesis therefore fails on one side and at distance at least $D/2$, so a union bound over the $V$ of them gives error at most $V \exp(-D/(8 \sum_k c_k^2))$, which is at most one half exactly when equation 9 holds. □

In the superposition that holds a belief (Section 3) the hypothesis codes $\mathbf{h}_k$ are fresh atoms, one independent Rademacher vector per hypothesis, and not bindings over a shared alphabet; the independence the proof uses is that of the atoms.

The sum in equation 9 is over squared multiplicities, and that is what the bound turns on. Writing $\sum_k c_k$, the number of codes superposed, in its place gives the same answer only when every hypothesis satisfies at most one observation, which a belief never does, since the hypotheses that survive satisfy all of them. Over 300 codebook draws at $V = 200$, $D = 4096$ the two forms agree where every $c_k = 1$ (905 predicted against 882 measured), and separate once they do not: with six observations and masks of thirty equation 9 predicts 2103 against 2008 measured, where the linear sum gives 859. Which sum is taken moves the dimension the belief needs by between 2.3 and 11 times, depending on how the masks overlap, although a sweep will not separate them, since equation 9 is a union bound at error one half and is loose by a comparable factor. The implementation carries the belief as a Gram matrix rather than as a $D$-dimensional vector, and that substitution is the superposition read exactly, interference included.

**Proposition 3** (Query order)**.** *A memory written at order $k_{\max}$, meaning that every subset of at most $k_{\max}$ variable assignments is bound into a key beside the context code, which does not count toward the order, answers a conditional on $m$ variables if and only if $m + 1 \leq k_{\max}$.*

*Proof.* Conditioning is a ratio of counts, $\widehat{p}(y \mid z) = \widehat{n}(y, z)/\widehat{n}(z)$, whose numerator is a key over the $m$ conditioning variables together with the target and whose denominator is a key over the $m$ conditioning variables alone. A key is readable exactly when it was written, so both exist if and only if $m + 1 \leq k_{\max}$, and the denominator alone if and only if $m \leq k_{\max}$. □

This is a capacity statement and not an expressiveness one: the adjustment formulas of Section 2 are available at the order their conditioning sets need, and the memory abstains rather than guesses when a query exceeds the order it was written at.

**Proposition 4** (Light cone)**.** *Let a rule of radius $r$ map a cell's neighbourhood to its next state, and let one patch of the initial frame be undetermined. Then after $k$ steps the undetermined region lies within the square of radius $rk$ about that patch, every cell outside it is exact, and the region fills the square whenever the rule is not constant on the patch's values.*

*Proof.* A cell at step $k$ is a function of the cells within radius $r$ at step $k - 1$, so by induction it is a function of the cells within radius $rk$ at step 0. If that disc misses the undetermined patch the cell is determined, which gives exactness outside the square. A cell whose disc meets the patch has two completions of the initial frame that differ there, and the rule is a function, so some completion changes the cell unless the rule is constant on the patch's values in that frame; a canalising rule, such as an isolated live cell in Life dying under every completion, can leave the region smaller than the square, so the square is the envelope. Either way a rollout is a survival process rather than a diffusion. □

The exactness is checked cell by cell against the closure (`tests/test_grid_rules.py`). The planner of Section 3 relies on it: unknown cells spread along the cone and never manufacture a determined value, so a plan can only be found through derivations the memory settles.

**Theorem 5** (The counterfactual boundary)**.** *Take a binary treatment $X$ and a binary outcome $Y$, and write the world as the joint of a response type $R = (Y_{X=0}, Y_{X=1}) \in \{0,1\}^2$ with the treatment actually assigned, eight cells in all. Every observational and interventional read is linear in those eight cells, and the map $\mathbf{A}$ they form has*

$$\operatorname{rank}(\mathbf{A}) = 6, \qquad \dim \ker(\mathbf{A}) = 2.$$

*The effect of treatment on the treated and the average treatment effect lie in the row space and are therefore point-identified by reads. The probability of necessity and sufficiency does not, and is identified only to an interval.*

*Proof.* An observational read $P(X = x, Y = y)$ selects the cells whose assigned treatment is $x$ and whose response type gives $y$ under $x$, and an interventional read $P(Y = y \mid \mathrm{do}(X = x))$ selects the response types giving $y$ under $x$, summed over the assignment. These eight functionals span a six-dimensional space: computing the singular values of the resulting $8 \times 8$ matrix gives six non-zero and two zero, so the rank is six and the kernel is two-dimensional. The effect on the treated is $P(Y_{X=1}{=}1 \mid X{=}1) - P(Y_{X=0}{=}1 \mid X{=}1)$, whose numerator is the difference of two cells with assignment $X = 1$; that vector is orthogonal to the kernel, as is the average effect, so both are determined by the reads. The probability of necessity and sufficiency is $P(Y_{X=1}{=}1, Y_{X=0}{=}0)$, the marginal of one response type, whose projection on the kernel is non-zero; it is therefore not determined, and its identified set is the interval the linear program over the simplex returns. □

The rank, the kernel dimension and every closed-form interval endpoint are checked against a linear program on 1,993 random structural models. The consequence the memory acts on is in Section 5: a cross-world query needs a key carrying two context codes at once, and no episode can occupy that cell. The tail claim behind the gate's abstention needs more than the pairwise uncorrelatedness Proposition 1 uses. Every written key is $\gamma_g \odot \mathbf{v}$ for one context code $\gamma_g$ and a product $\mathbf{v}$ of value atoms, and a cross-world probe is $\gamma_{g_1} \odot \gamma_{g_2} \odot \mathbf{v}'$ with $g_1 \neq g_2$. Their elementwise product contains at least one context code that appears exactly once, and that code is independent of every other factor, so each such overlap is a sum of $D$ independent Rademacher variables whatever relations the value atoms satisfy. The memory read $\langle \mathbf{S}, \mathbf{q} \rangle = \sum_d q_d S_d$ is then a sum of $D$ independent terms with mean zero and variance $\sigma^2 D^2$, so the gate at $4\sigma D$ is a four-standard-deviation event: its probability is at most $2e^{-8}$ by a sub-Gaussian bound, and about $6 \times 10^{-5}$ under the normal approximation. The gate of Proposition 1 therefore abstains from the probe without being told, and did on 60/60 cross-world reads.

## C WHAT DIMENSION FIXES, AND WHAT IT DOES NOT

Three registered experiments asked whether the memory's read noise is what limits its structure learning on the isolation stream, and each answered no. They are reported here because they bound what a change of write rule, readout or read operator could buy.

A write rule was proposed to break the clause of Section 2 that more episodes do not sharpen a probability read at fixed dimension. The derivation refuted its premise: at fixed context-size ratios the relative noise on a probability read is constant in the number of episodes. The clause therefore follows from the interference law, and a write rule controls only the stored energy, the sum of squared counts in the interference law. Measured on the same stream, with the band restated to include sampling and bands again from a pilot, a memory with exact counts and no interference commits 4.70 of 15 pairs at parent F1 0.630, against the current rule's 0.558. A write rule therefore has 0.072 of F1 and 1.00 pair left here, and then nothing.

The memory commits an edge when a contrast clears four sigmas computed from the interference half of the noise alone, and that half falls with dimension while sampling noise does not. Committed-pair ancestry precision falls with it, from 1.00 at $2^{16}$ through 0.97 to 0.79 at $2^{20}$, a defect the interference law predicted before it was measured. The likelihood the same read implies, under calibrated contrast laws for the three roles, parent, non-parent ancestor and neither, keeps committed precision at 1.00 at $2^{20}$ on the pairs that carry a context. Scoring the pairs abstained from through the observational channel gains 0.258 of parent F1 there, paired over ten seeds, at a pooled calibration error of 0.046, and moves precision by -0.112 on the isolation stream for answering where the rule

abstained. The bands were fitted to a pilot of the same experiment, so this is a reproduction, and the observational gain needs a calibration fitted to the same generator, which a tenth of error in the assumed parent contrast removes.

Read noise is the suspect in each paragraph above, so the same writes were read by least squares against the codebook instead of by the matched filter, the dot product every read above uses. The relative read noise falls from 0.0483 to $4.8 \times 10^{-7}$ on 5 seeds, parent F1 is 0.657 before and 0.657 after, and the committed sets are identical on every seed. In this setting read noise is not what binds.

## D THE MEMORY AND THE INDUCER

**The memory.** One bipolar hypervector over $\{\pm 1\}^D$ (Kanerva, 2009; Kleyko et al., 2023b); per-variable random codebooks; one random context code per context; binding is the elementwise product; an episode writes its context-bound interaction terms at orders $1..k_{\max}$ plus a unit term for counting. Counts are inner products, probabilities are ratios, adjustment sums reads over the confounder's stored values. Writes are ungated. Read noise $\sigma = \sqrt{\sum_c n_c^2 / D}$ is computed by the memory from its own counters. Per operation at $D = 65{,}536$ and $k_{\max} = 1$ a write costs 1.8 ms and a read costs 0.50 ms on one laptop core; at $k_{\max} = 3$ a write costs 16.6 ms and a backdoor adjustment costs 4.8 ms. The interference law is checked on a binary confounder triangle $Z \to X$, $Z \to Y$, $X \to Y$, written under the observational context and both values of $X$ at orders up to three with $2{,}000$ observational and $800$ episodes per do-context, over $D$ from $2^{13}$ to $2^{17}$ and three codebooks; the measured $\sigma$ is the spread of the read error over every order-one to order-three probe in every context, against the exact count from the memory's own multiplicity counters.

**The inducer.** Transitions are keyed by action and the radius-$r$ patch; a key with two witnessed successors is a conflict, resolved first by escalating the radius and then by searching for a separator, a cell at a fixed position or at a fixed offset from a token that occurs once on the grid, whose value splits the outcomes. Symmetry licences (the dihedral group on grids, the cube group on Alchemy's latent space) are validated against the data and forfeited where validation fails. Observing a $32 \times 32$ Life frame costs 1.6 ms and predicting one 0.96 ms.

**The trained baselines.** The local network is a multilayer perceptron over the one-hot radius-1 patch and action (two hidden layers of 256, 60 epochs of Adam) trained on the identical corpus per world; abstention by softmax confidence with the threshold set so the network commits on exactly the inducer's coverage. For Baba Is AI a frame transformer over all cell tokens (two layers, a positional embedding and an action embedding), trained on the same 150-episode corpus and scored on the same held-out collisions.

**Matched coverage on the collision cells.** The same three levels, corpus, networks, seeds and epochs as above; for every held-out collision cell (the agent cell and the target cell of each into-wall transition) the arm's softmax confidence and correctness beside the inducer's decision. Per rule state: accuracy on the top-confidence cells, exactly as many as the inducer commits on, the tie band at the cut committed in expectation; accuracy on exactly the cells the inducer commits on; the abstention curve at coverages $0.1$ to $1.0$; and, of the wrong committed cells, the share at confidence $0.99$ or more, split by whether the cell's patch key occurs in training. Means over the three levels, pooled figures by counts, never a joint ranking across networks. The theory arm ran first: a lookup table of the (patch, action) keys of the training set, prediction the majority token and confidence its frequency, from which the bands were set; the collision counts of the retrained arms are asserted equal to that instrument at run time, and the gate is those collision accuracies within $0.01$ (the local network reproduces them on 3 of 3 levels, largest departure 0.0000; the frame transformer on 2 of 3, largest departure 0.3483; its backward is not bit-identical across thread counts, and the first run logged none).

**Rules across layouts.** The same 53 levels, the same 150 random-play episodes and the same instrument as the single-layout ladder, in three arms that separate a change of layout from a change of corpus: the episodes on one layout (the single-layout ladder); on five layouts of the family at thirty episodes each, scored on held-out episodes from those same layouts; and on five layouts scored on episodes from five further layouts the inducer never explored, the fit and score cut at the episode

| estimator | shuffled | blocked | ratio |
|---|---|---|---|
| gradient estimator | 0.062 | 0.137 | 2.2× |
| wider estimator | 0.046 | 0.127 | 2.8× |
| elastic weight consolidation | 0.062 | 0.099 | 1.6× |
| linear, shared bias | 0.039 | 0.099 | 2.5× |
| linear, no bias | 0.033 | 0.030 | 0.9× |
| the memory | 0.032 | 0.030 | 0.9× |

Table 1: **Error on the earliest context, by estimator family.** The same episodes in both columns, ten seeds, on the stream of Fig. 2b; the means of Fig. 2c. ENCO is scored on edge recall rather than on this error and is reported in the text.

| replay buffer | blocked mean | blocked earliest | shuffled mean |
|---|---|---|---|
| 0 | 0.192 | 0.176 | 0.056 |
| 25 | 0.176 | 0.169 | 0.056 |
| 100 | 0.122 | 0.146 | 0.048 |
| 800 | 0.060 | 0.051 | 0.043 |
| the memory (no replay) | 0.027 | 0.030 | 0.027 |

Table 2: **Replay on the isolation stream.** Mean and earliest-context error of the gradient estimator under blocked arrival against the size of its replay buffer, with its shuffled mean and the memory without replay, three seeds.

boundary so the scored cells come from unseen layouts by construction. Before the run the layouts were checked to vary with the reset seed on every level, and the scored cells were measured to be largely fixed by the benchmark, the only pushable-wall structure being the two-room divider, while under a fifth of the keys at those cells recur across layouts. Coverage is decomposed into the four reasons the inducer abstains. The planning half is the same in all three arms: the test episode resets the level at its own seed and the planner reads a memory fitted on every interaction episode, the scored fifth included, so in the ladder that layout is the one the episodes explored and in the other two it is one the inducer never interacted with.

**The estimator families.** On the isolation stream, beside the gradient estimator of Section 2: a wider one (two layers of 256), elastic weight consolidation on the original with a Fisher penalty at each context boundary and $\lambda$ swept over $\{1, 10, 100\}$, and a linear map from the one-hot context to the six logits, with and without a bias vector; ten seeds, blocked and shuffled, no replay. On the wormhole world, beside the original network: a wider one (three layers of 512) and an embedding-table network (a 64-dimensional embedding per state from a $14\times14$ table, an action embedding and a two-layer head); ten seeds.

**The delay, chain and isolation streams.** The delay stream writes a four-variable system and reads each interventional distribution after up to $1000$ interleaved unrelated episodes, scored against the true distribution; the trace arm trains an eligibility-trace learner on the same delayed-outcome attribution; the chain stream plants a three-variable chain and reads each do-contrast with and without the middle variable held. The isolation stream draws its world from one generator, used again by the policy experiment of Section 4: six binary variables in a random order, each later variable a child of each earlier one with probability $0.45$, and each variable equal to one with probability $0.15$ plus a weight drawn from $[0.4, 0.8]$ for every parent that is one, clipped to $[0.05, 0.95]$. Seven contexts, the observational one and both values of three variables chosen at random, contribute $300$ episodes each, $2{,}100$ in all, and the true interventional marginal of every variable under every context is computed by enumeration. The same episodes arrive blocked or shuffled. The gradient estimator maps the one-hot context through two hidden layers of $32$ units to six sigmoid outputs, one per variable, trained by stochastic gradient descent in a single pass with the learning rate swept over $\{0.3, 0.1, 0.03\}$ and the best kept, with and without a reservoir replay buffer of $25$, $100$ or $800$ episodes from which one episode joins each update; the memory holds the same episodes at $D = 65{,}536$ and $k_{\max} = 1$. The metric is the mean absolute error of the six interventional marginals per context, and the memory's band is $4\sigma$ over the context count plus the binomial term at $300$ episodes.

**The world models.** The first is a $12\times12$ walled gridworld with four unit moves and states as spatial codes; the memory stores transitions and reads by unbinding and cleanup, and the hybrid

| delay | store | eligibility trace |
|---|---|---|
| 0 | 1.00 | 1.00 |
| 5 | 1.00 | 0.71 |
| 20 | 1.00 | 0.08 |
| 100 | 1.00 | 0.12 |

Table 3: **Attribution across a delay.** Accuracy of delayed-outcome attribution for the memory and for an eligibility-trace learner as the delay between action and outcome grows.

| experiment | quantity | value |
|---|---|---|
| planted chain | direct contrasts $A \to B$, $B \to C$ | 0.72, 0.72 |
| | two-step contrast $A \to C$ | 0.50 |
| | the same, $B$ held | 0.18 |
| | band | 0.24 |
| isolation stream | ancestry precision, committed pairs | 1.00 |
| | ancestry precision, pairs abstained from | 0.22 |
| Sachs | ancestry AUROC, all pairs | 0.677 |
| | pairs with a perturbation | 0.792 |
| | exact count table, all pairs | 0.757 |
| | exact count table, pairs with a perturbation | 0.808 |

Table 4: **Structure from contrasts of reads.** The planted chain (ten seeds), the pairs the memory commits and abstains from on the isolation stream, and the Sachs ancestry AUROC of Fig. 2d.

adds a displacement operator with an exception memory written only where the operator is wrong. The second is the wormhole world, a $14\times14$ grid with twelve (cell, action) pairs whose outcome is a distant cell; the rule network is trained on regular transitions first and the exceptions then arrive as a stream, fine-tuned into the network or written to a memory beside it.

**The structure learners.** On the isolation stream the memory's estimator scores an ordered pair by its do-contrast where the source has a context and by the observational conditional contrast otherwise, the latter ranking the pair for AUROC but never committing it; a pair is committed when its contrast clears four memory sigmas and the minimum effect $0.05$, and abstained from otherwise, and a committed ancestor is a parent when its contrast survives conditioning on every other committed ancestor at $k_{\max} = 3$. Blocked and shuffled runs are the same table filled in a different order. IGSP runs from `causaldag` with partial-correlation and Gaussian invariance tests at $\alpha = 10^{-3}$ and depth four, which are misspecified for binary data and may cost IGSP orientations here; DCDI-G runs from its released code in batch only, and ENCO from its released code with the sparsity weight fixed at a pilot value. ENCO's block-sequential run trains one context's episodes at a time in arrival order with warm starts and no replay, so it measures what an unbuffered sequential fit of that model forgets; the replay result of Table 2 says what a buffer would recover, at the cost of holding the stream. One arm runs IGSP with its tests read from the memory at full order; another a PC-stable skeleton at depth two on the observational context with a chi-square whose per-cell variance adds the memory's own $\sigma^2$, orientation by do-contrasts, v-structures and the Meek rules, undirected edges abstained from. On Sachs the memory commits by the same rule at $k_{\max} = 2$ and IGSP and DCDI-G take the same discretised data with the same context labels, all scored on the consensus network's ancestry closure. The design comparison enumerates the Markov equivalence class of each seed's DAG and counts single-vertex interventions, with perfect-information elimination, until one DAG remains, for the channel search's rule, uniform random, round-robin and the greedy design, twenty draws per seed.

**Expected free energy as the policy.** No memory is instantiated: the read is simulated from the interference law, which predicts the recorded run's read noise to within 1.5% on every seed. The belief is a posterior over every labelled DAG on six nodes (3,781,503) under the generator's prior (entropy 20.48 bits). The likelihood is the probability of the read given the graph under contrast laws calibrated at the measured noise, and the policy takes the intervention with the largest expected information gain, ties evaluated as exact expectations over the tie set. The primary setting is six variables, a budget of three interventions of 300 episodes per value at $D = 2^{20}$, 30 seeds; the passive setting adds 3,000 observational episodes first. Arms: information gain and its pragmatic

variant, uniform random averaged over draws, round-robin, the best fixed triple, and the set-valued belief reproduced, the graded posterior fed only binary dependence facts, the discovery policy's realised intervention sets rescored under the graded posterior, and two truth-seeing references. The score is the truth's surprisal under the arm's posterior after the budget, paired within seed against random choice and normalised by the per-seed oracle-design range; the attainability gate admits the metric only where its within-seed spread exceeds the seed-to-seed spread, evaluated on the noisy channel. The bands came from a five-variable pilot (ten seeds, budget two) through the ceiling script, committed before any six-variable setting ran.

**The set-valued policy.** The isolation-stream generator at six variables, 10 seeds, an observational block of $3{,}000$ episodes and then interventions on every variable at both values, $300$ episodes each, at $D = 2^{20}$ and at the registered $65{,}536$. The belief is the class of graphs consistent with every decision the memory's estimator commits, with the observational dependence facts read at the same gate; the policy picks the intervention expected to shrink that class most, against uniform random, round-robin and a designer who knows the truth, scored by parent recall and by the bits of class left at the budget. IGSP is run from `causaldag` on the samples the same loop collected.

**The readout, the write rule and the counterfactual boundary.** All three run on the isolation stream in the setting above, $D = 65{,}536$ with seven contexts of 300 episodes, ten seeds in blocked order, with the memory, episodes and reads unchanged. The bands of all three were fitted to a pilot of the same computation. The readout replaces the four-sigma decision with the posterior over three roles (parent, non-parent ancestor, neither) that the same read implies under contrast laws calibrated on this generator. Two variants are scored: interventional, scoring only pairs whose source carries a do context, and observational, scoring the rest through the observational channel at its own read noise, four times the interventional one. It is repeated at $2^{18}$ and $2^{20}$, and the comparison against IGSP is paired on shared graphs. The write rule derives the relative noise on a probability read at fixed context-size ratios, $\sqrt{\sum_g c_g / D}$ with $c_g$ the energy coefficient of context $g$, one plus the summed squared probabilities of its stored assignments, constant in the number of episodes, and measures three write rules (additive, per-context gain, centred basis) against a noise-free memory with exact counts under a band restated to include the sampling term. Its imbalanced setting adds 3,000 observational episodes. The counterfactual arm writes a binary treatment and outcome under the observational context and both do-contexts, reads the eight-cell joint over response types, checks every closed-form interval endpoint against a linear program on 1,993 random structural models. It scores the effect of treatment on the treated and the necessity-and-sufficiency bounds on 150 populations, with truth from a canonical coupling over $20{,}000$ units per cell; cross-world cells are read through the existing denominator gate. The read noise on those cells is computed from the covariance form of the interference law rather than from a diagonal approximation: over 300 codebooks at $D = 4{,}096$ the predicted read standard deviations, 15.26 to 15.34, match the measured 14.40 to 16.01 (ratios 0.96 to 1.07, and the law reproduces the memory's own noise from empirical counts to 0.000); the diagonal rule overstates the standard deviation of the effect of treatment on the treated by 14%, because the two reads a contrast subtracts share codes and their errors are correlated. The control reads the same writes by ridge-regularised least squares against the codebook on 5 seeds and re-runs the estimator unchanged.

**The switch world.** An $11 \times 11$ layout in which three switches open door groups through a random permutation, one effect delayed and one door driven by a hidden timer; a 2400-step footprint, a $1200$-step discovery phase and twenty prompted tasks capped at $60$ steps, the cap amended from forty before adjudication. The memory's agent and the exact count table share one motor layer, shortest path over known free space, and differ only in where the contrasts are read; Q-learning is tabular $Q$ over (position, doors) trained from reward, and is given no motor layer; a learner over the same motor layer was not run.

**The Alchemy agent and the induced family.** From stored random play the agent accumulates a skeleton (which states co-occur, which pairs are ever interchanged, which side of each axis is which) and closes it under the group orbit to construct the hypothesis family. The channel search runs over the environment's observables, as in Section 3. Channel sufficiency is an exact enumeration: every one of the 1,536 chemistries is hashed by its signature on the transition channel, the class sizes and the spread of stone value inside a class are read off the classes, and the displayed reward is added

| Benchmark | Interface given | Memory (no training) | Comparison |
|---|---|---|---|
| POPGym (Morad et al., 2023) | read-out rule per family | at or above the best published model on 9 of 12 settings | published table |
| AutumnBench change detection (Warrier et al., 2025) | symbolic grid, unmodified program | 0.412 | Gemini 2.5 Flash 0.323; human 0.908 |
| AutumnBench counterfactual | symbolic grid, exact truth | 0.84 on the changed cells it determines | copy-current: 0.47 |
| XLand-MiniGrid (Nikulin et al., 2024) | rule grammar | 40 rules, 0 wrong | own random control |
| ARC-AGI-3 (Foundation, 2026) | raw frames | 0 levels | random control 1 |
| Baba Is AI (Cloos et al., 2024), break-a-rule | symbolic grid | 5 of 5 | control 0 |
| Baba Is AI, make-a-rule | symbolic grid | 0 of 29 | every policy tried |
| Alchemy (Wang et al., 2021), held-out seeds | description language, families induced | 0.91 of the ideal observer | ideal observer 284.4 |
| Alchemy from pixels | rendered stones, families induced | 0.92 of the ideal observer | symbolic result above |

Table 5: **External benchmarks, with the interface each result carries.** Each row is a contrast across categories, licensed by its interface column.

as a channel with 3 readings per episode; the gain is the same agent's return with and without that channel written. The belief is a superposition of the codes consistent with each observation, sized by the support-read law of Appendix B, and thresholded absolutely at $(u - \frac{1}{2})D$ for $u$ observations, never relative to the read's own maximum; the planner acts on the surviving support and abstains when it is empty. The support read is swept on that belief at 11 observations per episode, $D$ from $8{,}192$ to $262{,}144$ against the 139,634 the registered law asks for, twenty trials per point, scored by the fraction of the true support recovered with the absolute threshold and with one relative to the read's own maximum. From pixels the potion and stone channels are induced the same way before the family is built.

**The exact count table.** Exact counts keyed by (context, full assignment) with a tight encoding of one byte per variable, one for the context and four for the count; the memory's footprint is $4D$ bytes of float components. The footprint family sweeps the number of binary variables from six to twenty-four at 19,999 episodes over seven contexts, three seeds per size, with the memory at $D = 65{,}536$; the crossover is the first size at which the table's bytes exceed the memory's, which the encoding puts where the distinct keys exceed $4D/(N + 5)$. Do-marginal error is the isolation-stream metric.

**Benchmarks.** POPGym at the published difficulty tiers with a given read-out rule per family and no training; the comparison column is the benchmark's own published table, re-verified against two later methods before any claim. FFM raises the aggregate MMER to 0.400 from 0.349 and publishes no per-task table. SHM reports mean return at the end of training rather than the maximum over epochs, reaching -0.194 on RepeatPrevious hard and -0.439 on Autoencode hard, so neither overturns a win here and Concentration stays below every published method. Alchemy at the perceptual-mapping-with-rotation-and-random-bottleneck level, three development seeds throughout the eight designs plus 10 held-out seeds run once; the same published table gives 156.2 for V-MPO, 140.2 for IMPALA and 145.7 for a random heuristic and 155.4 for V-MPO on symbolic observations, quoted here as scale rather than as comparison. XLand-MiniGrid `R1-9x9` on the `small-1m` family. AutumnBench on its 21 public environments, with the published per-environment table stored as evidence and the comparison computed on the shared 16; the counterfactual score is accuracy on the changed cells the memory determines, which are 0.59 of them at one step, where the exactness figure is accuracy on every determined cell. Baba Is AI on all 53 public levels, split by whether a twenty-episode random control ever wins: on the 23 where it does the memory wins 22 and the control 11, and on the 30 where it never does the memory wins 1 (the control 0 by construction), with the break-a-rule and make-a-rule levels reported separately in Table 5; 150 same-layout interaction episodes then one test episode, under two registered interaction policies, random and frontier, whose separator rates are 0.51 (at its registered 50% line) and 0.94 and whose break-a-rule wins are 5 and 2 of 5. ARC-AGI-3 on the three public games at five times their published per-level human baselines. The environment's rule set or program source, where available, is read for scoring only.

## E STRESS TESTS ON BENCHMARKS WE DID NOT BUILD

The memory and the search beside it were taken unchanged to benchmarks we did not build, with the interface stated in advance (Table 5). POPGym asks whether a positional register, an ordered replay and a counting memory hold against thirteen sequence architectures trained end to end, each behind a read-out rule the memory is given rather than learns. The three settings it loses are the Concentration variants, where the score depends on which card to turn over rather than on what is remembered. AutumnBench asks whether a memory that abstains can tell a changed program from an unchanged one, which it does with 1 false alarm on 21 environments where the registration allowed none. It also asks whether its answer to a counterfactual query, the frame after a swapped action, matches the interpreter's exact answer on the cells the swap changes (Table 5); over every determined cell it misses the registered exactness band (0.975 against $0.98$ at one step). XLand-MiniGrid asks whether production rules are identified from play, which the observable channels license only up to a per-trial rotation the memory cannot see, so the claim is that no identified rule is wrong. ARC-AGI-3 is a recorded miss, no level completed against one for the random control, and it fails at the exploration boundary Section 6 names. Each row is handed something its comparison was not, which the interface column states, so the rows measure what the memory's representation can carry under a given read-out interface, not parity with agents that learn the interface from raw interaction. The registered bands and the criteria that fired ship with the registrations.

## F REPRODUCTION, REGISTRATIONS, AND WHAT FAILED

An anonymised repository accompanies the submission, and everything the manuscript reports can be regenerated from it at three levels, each one command from its root. At the first level the manuscript is rebuilt from the shipped evidence: every result file the paper uses is in `paper_iclr/evidence/` with a sha256 manifest, no measured number is typed into the text, and every one enters through `generated/macros.tex`, which `scripts/make_tables.py` emits from the evidence and `scripts/make_figures_iclr.py` draws the figures from; an integrity suite (`tests/test_causal_paper_macros.py`, `tests/test_iclr_paper_macros.py`) checks that the macros are a fixpoint of the evidence, that the manifest hashes match, and that the prose contains no number except through a macro. At the second level the laws of Section 2 are unit tests on the algebra itself (`tests/test_causal_memory.py`), which hold exactly or fail. At the third level each experiment is a module under `experiments/`, run as `python -m`, that writes its result file to `results/`, from which `scripts/collect_evidence.py` copies the registered files into the evidence directory and rewrites the manifest; the repository's README maps every figure panel and the main runs to the modules that produced them. Every result file records the host, platform and library versions it was produced with, and the one comparison sensitive to the machine, ENCO under blocked against shuffled arrival, was run on a single node with requeueing off. The registrations that fixed each prediction, band and kill criterion before its run are shipped as `docs/PREREG_*.md` with their amendments dated in place, and the theory the predictions were computed from as `docs/causal_memory_theory.md`.

Every experiment but the least-squares control was registered with its prediction, effect size and refutation criterion before it ran. The policy experiment's bands preceded its runs. Three others, the graded readout, the write rule and the counterfactual boundary, took their bands from a pilot that had already run the same computation, so their numbers are reported as measurements rather than forecasts. The frame transformer's Baba number carries a caveat of its own: its reproduction gate passed on 2 of 3 levels.

Registrations that failed bound the claims above. The interference law's first registered form, $\sqrt{N/D}$ for independent items, was refuted by its first exact test and replaced by the coherent form of Proposition 1. ARC-AGI-3 was closed after three registered attempts, and AutumnBench's planning arm was killed on a metric that is constant across the class its channels leave. Registered bars were missed rather than met on the memory's error past the footprint crossover (0.034 against a registered $\leq 0.01$), on the Alchemy support read at its registered dimension, which was not re-run to clear it, on the Sachs ancestry AUROC over all pairs (0.677 against a registered $\geq 0.75$, met on the pairs with a perturbation), and on two readings of the policy experiment, the gain after passive data (0.817 against a band within $0.3$ bits of zero) and the posterior's calibration coverage (0.933, at the edge of its band, so calibration is left undetermined). A further bar was missed on AutumnBench's

counterfactual exactness (0.975 at one step against a registered $\geq 0.98$), and its change detection raised 1 false alarm on 21 environments where the registration allowed none (Appendix E). The premise of abstention-directed exploration, that pushing a rule block leaves more cells abstained from than a floor step does, was refuted by direct measurement and closed at its pilot. Across Baba layouts the rule transferred where the plan did not, so the break-a-rule kill fired on 0 levels won of 5. A registered prediction that a trained abstaining network would be less exact than the memory on grid worlds had its criterion fire, the network matching the memory at matched coverage. Verifying the counterfactual setting found two registered readings missed, a seed band and a maximum on the mixed-sign family (0.041 against a registered $0.016$), with the numbers unchanged, and on families too monotone to exercise the interval's width. One proposed write rule had its committed-pair figures withdrawn when an independent check found its deciding read computed rather than read from the memory. The acceptance rule registered before the causal-discovery runs, three clauses on what the memory had to reach (`docs/PREREG_causal_discovery.md`), failed on two of them, which is why those results are one section here rather than the paper's organising claim.

Four registered attempts chose the next intervention from a signal read off the memory, and none separated from round-robin or random choice. Three ranked interventions by the familiarity of the context, by the novelty of the content, and by both again at a small budget. The fourth is the set-valued policy of Section 4: it held the belief as the class of graphs consistent with every committed decision and picked the intervention expected to shrink that class most. Over 10 seeds at $D = 2^{20}$ it recovered 0.43 of the true parents against 0.46 for random choice, 0.41 for round-robin and 0.68 for a designer who knew the truth, so its kill criterion, a gain of at least $0.10$ over random choice, fired. Its ceiling had been computed on the assumption that every intervention yields the true ancestry of the intervened variable, which gave 0.73 recall and 2.5 bits of class left at the budget; the memory yields the ancestry its estimator commits, which is that row times its own recall, and at the budget the class still held 14.3 bits and the truth remained inside it on 0.20 of seeds, against 0.40 for round-robin. A class that is mostly prior ignorance, anchored by the occasional wrong fact, ranks nothing. At the originally registered $D = 65{,}536$ the read noise after the observational block let almost nothing commit (0.01 recall) and every arm kept the truth in its class, since a memory that commits nothing eliminates nothing. The same data handed to IGSP reached a parent F1 of 0.94 against the memory's 0.80, so the policy's data is richer than the batch stream's and a batch learner converts it better. The claim that the memory's recall gap to IGSP was an artefact of who chose the interventions is refuted by that reading.