A LOCAL ASSOCIATIVE MEMORY THAT KNOWS WHERE IT HOLDS

Anonymous authors
Paper under double-blind review

ABSTRACT

An exemplar memory must both retrieve corrupted versions of stored items and reject cues absent from the store. We cast these as two outcomes of one local decision: whether a cue lies in a region where a stored memory has grounds to answer. The escape and retrieval field (ERF) assigns each memory such a region using the spherical support of a truncated Gaussian well, together with a shared outward radial drift. Its static rule accepts cues inside the union of the supports and rejects cues outside. Its dynamics provide conditional geometric guarantees for any chosen radii: every isolated support is forward invariant, while a cue whose ray from the reference point meets no support escapes geometrically. We construct the regions in two ways. Geometric radii, obtained from the store's angular spacing, yield a closed-form tolerance to missing coordinates filled with the reference point and capacity bounds for random memory directions. Fitted radii instead adapt a metric and one region per memory using only stored items and their corrupted views. Across two frozen encoders, three stores and six corruption families, the fitted metric identifies the source memory with 87 % accuracy, compared with 64 % for modern Hopfield retrieval in the encoder's original space. Because both methods select the same identity in the fitted metric, this gain comes from metric adaptation; the fitted radii determine whether to answer. Rejection performance depends on the store's separation: above an exploratory threshold, ERF has the highest balanced accuracy in 83 % of conditions and on average in every corruption family; below it, ERF trails the strongest comparator on average. With the threshold fixed, a subsequent test predicts the sign of ERF's mean advantage in all 10 new store–encoder settings, but only a fraction 0.68 of individual conditions. ERF therefore unifies retrieval and rejection through the local regions in which its memories hold.

1 INTRODUCTION

An associative memory over individual exemplars must both retrieve corrupted versions of stored items and reject cues it has never stored. We study stores whose unit of memory is one image: either one exemplar from each of several classes or a collection of distinct reference images. A familiar cue is a corrupted copy of one of these images, whereas a novel cue is absent from the store; recognising a new instance of a stored class is a different problem. Classical, dense and modern Hopfield networks retrieve stored patterns (Hopfield, 1982; Krotov & Hopfield, 2016; Ramsauer et al., 2021), but their standard updates do not specify when to reject a cue. Convergence identifies a preferred memory or mixture, not whether the input came from any stored item. Retrieval and rejection therefore reduce to one question: where does each memory have grounds to answer?

Most rejection rules compare an input score with a threshold, using quantities such as nearest-neighbour distance, Mahalanobis distance or classifier confidence (Hendrycks & Gimpel, 2017; Lee et al., 2018; Sun et al., 2022; Yang et al., 2024). Such thresholds are commonly calibrated on familiar validation inputs, distractors or a target acceptance rate. Our setting provides none of these independently collected observations for each stored item. Distances between clean exemplars reveal how crowded the store is, but not how far a corrupted copy of a particular exemplar may move. Synthetic corruptions can declare which variations should be tolerated, yet the resulting region must still cover familiar cues without creating excessive ambiguity between memories. The

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.

missing object is therefore a local validity region for each memory, determined without evaluation cues or distractors.

We introduce the *escape and retrieval field* (ERF), which uses these local regions to unify retrieval and rejection. Each stored item is surrounded by the spherical support of a truncated Gaussian well, and a shared radial term pushes the state away from a reference point c. A static rule accepts a cue inside the union of the supports, retrieves a memory, and rejects a cue outside. The support radii thus specify where the memory may answer. Once the radii are given, the field supplies corresponding conditional dynamics: the width and depth of each well are chosen so that its attraction at the boundary balances the same outward drift; every isolated support is forward invariant; and a cue whose ray from c meets no support escapes geometrically. These guarantees are deliberately local. Supports may overlap, and a cue starting outside all supports may enter one along its ray, so the paper distinguishes the static rule, a dynamical run and the identity readout.

The framework separates the field from the data-dependent choice of its radii, and we study two ways to specify the same local object. *Geometric radii* use the angular spacing of the available sample about c. For a stored image whose missing coordinates are filled with c, they give an exact retention criterion and a tolerance–capacity relation under random memory directions. *Fitted radii* use the stored items' own synthetically corrupted views. We adapt a metric on top of a frozen encoder and place one radius per memory between the spread of its own views and the nearest views of other memories. No evaluation cue or distractor enters the metric, radius or threshold. The geometric and fitted regimes therefore answer the same question—where each memory holds—under different assumptions about the variation to tolerate.

Experiments separate identity retrieval from the decision to answer. Across two frozen encoders, three stores and six corruption families, the fitted metric identifies the source memory for 0.870 of familiar cues, compared with 0.638 for modern Hopfield retrieval in the encoder's original space. ERF and modern Hopfield select the same identity once they share the fitted metric, so the retrieval gain comes from metric adaptation; the fitted radii determine acceptance and rejection. Their performance depends on a store-level *separation* statistic computed from the fitting views. Above the exploratory threshold 0.9, ERF reaches 0.876 mean balanced accuracy against 0.801 for the strongest comparator and leads on average in every corruption family; below the threshold, it trails the strongest comparator on average. After the threshold was fixed, a subsequent test predicted the sign of the mean advantage in 10 of 10 new store–encoder settings, but only a fraction 0.68 of individual conditions. Separation is therefore an empirical diagnostic of stores, not a condition-level guarantee. Moreover, the fitted supports overlap, so the isolation guarantees do not apply to them directly. This distinction keeps the paper's two claims aligned: the field gives conditional geometric guarantees once local regions are specified, while the data determine whether those regions are useful in practice.

Our contributions are:

- **A local field that can retrieve or decline.** We construct an energy field with one radius per memory, prove forward invariance of isolated supports, and establish geometric escape when a cue's ray meets no support. Convergence within supports is examined numerically.
- **A geometric tolerance–capacity relation.** We derive the exact tolerance to missing coordinates filled with c, identify the smallest memory-centred balls covering the prescribed tolerance region, and bound the capacity of the angular-separation certificate under independent uniform memory directions.
- **Local regions fitted from the store itself.** We fit a metric and individual radii from synthetic views, compare acceptance rules under matched information, and test on additional stores and encoders whether store separation predicts the method's mean advantage.

2 THE ESCAPE AND RETRIEVAL FIELD

Having framed retrieval and rejection as membership in per-memory local regions, we now construct a field whose dynamics respect those regions. Let $M = \{m_i\}_{i=1}^{K} \subset \mathbb{R}^d$ be the stored memories, x_0 a cue and x_t the state. The construction takes a reference point c, a positive radius R_i for each memory and, for runs outside all supports, an angular scale $\theta \in (0, \pi/2)$. We write $\nu_i = \|m_i - c\|$, $d_i(x) = \|x - m_i\|$ and $u(x) = (x - c)/\|x - c\|$. The radii define where the memory may answer;

the field defines motion within and outside the resulting regions. We first state the local decision, then construct the field, establish boundary balance and the retention–escape guarantees, describe the two choices of radius, and finally give the finite-run protocol.

2.1 Local supports and decisions

The support of memory i is the open ball $B(m_i, R_i)$, and the active set at x is $S(x) = \{i : d_i(x) < R_i\}$. The ERF accepts a cue exactly when it starts in the union of the supports. With the geometric radii used for the dynamical runs, it retrieves the memory whose support contains the cue most deeply in that memory's own units:

$$\begin{aligned} \text{recognise } x_0 \iff \min_i \frac{d_i(x_0)}{R_i} < 1, \\ \hat{\imath}(x_0) = \arg\min_i \frac{d_i(x_0)}{R_i}. \end{aligned} \tag{1}$$

The first line is the static acceptance rule for either choice of radius. In the fitted regime, the identity readout is instead $\arg\min_i d_i(x_0)$ in the fitted metric: a fitted radius calibrates whether memory i may answer, rather than ranking identities against one another.

Three decisions must therefore be kept separate. The static rule reads the cue at its starting point; a run of the field can carry a cue that starts outside every support into the first support met along its ray; and the identity readout names a memory after acceptance. With geometric radii, a ray can meet memory i only inside the cone

$$\angle\big(x_0 - c,\ m_i - c\big) < \theta. \tag{2}$$

Except for a narrow grazing band, a cue between c and such a support enters it under the update below (Appendix A.2). Thus the static rule and a dynamical verdict need not coincide, even though both are organised by the same local regions.

2.2 Field and update

For widths σ_i, depths A_i and a slope $\lambda > 0$, the energy is

$$\begin{aligned} E(x) &= -\sum_{i=1}^{K} A_i\, \phi_i\big(d_i(x)\big) - \lambda\|x - c\|, \\ \phi_i(r) &= \big[e^{-r^2/2\sigma_i^2} - \kappa\big]\, \mathbf{1}[r < R_i], \end{aligned} \tag{3}$$

where $\kappa = e^{-R_i^2/2\sigma_i^2}$ for every i. Each memory contributes a Gaussian well truncated at R_i and shifted so that the energy is continuous at the support boundary. The term $-\lambda\|x - c\|$ supplies an outward drift of constant magnitude. Off the support spheres and c, gradient descent with step size $\eta_t = \tau_t \kappa \ell_t / \lambda$ gives

$$\begin{aligned} x_{t+1} &= x_t + \tau_t \Big[\kappa \ell_t\, u(x_t) - \sum_{i \in S(x_t)} \frac{\ell_t}{R_i}\, w_i\, (x_t - m_i)\Big], \\ w_i &= e^{-d_i(x_t)^2/2\sigma_i^2}, \\ \ell_t &= \begin{cases} \dfrac{\sum_{i \in S(x_t)} w_i R_i}{\sum_{i \in S(x_t)} w_i}, & \text{if } S(x_t) \neq \emptyset, \\ \|x_t - c\| \sin\theta, & \text{if } S(x_t) = \emptyset, \end{cases} \end{aligned} \tag{4}$$

with the schedule

$$\tau_t = \frac{1}{\max(1, W_t)}, \qquad W_t = \sum_{i \in S(x_t)} \frac{\ell_t}{R_i}\, w_i. \tag{5}$$

Inside active supports, ℓ_t is their kernel-weighted mean radius. Outside every support, the state moves straight away from c by the factor $1 + \kappa \sin\theta$ per step. When $W_t > 1$, the schedule turns the active-well contribution into a kernel-weighted mean and prevents the large steps observed under crowding (Section 3.2). Appendix A.1 derives the gradient form and scaling properties.

2.3 BOUNDARY BALANCE AND LOCAL RETENTION

The radii $R_i > 0$ are the construction's free parameters. Given them, the remaining field parameters are fixed up to the kernel cutoff κ. Let R_{ref} be the median radius and set

$$\begin{aligned} \rho^* &:= \sqrt{2\ln(1/\kappa)}, \qquad & \sigma_i &:= \frac{R_i}{\rho^*}, \\ A_i &:= \frac{R_i}{R_{\mathrm{ref}}}, & \lambda &:= \frac{\kappa\rho^{*2}}{R_{\mathrm{ref}}}. \end{aligned} \tag{6}$$

Then every kernel equals κ at its support edge and every well's boundary pull, $(A_i/\sigma_i^2)R_i\kappa = \lambda$, equals the shared outward push. One slope therefore balances every well, regardless of its radius. On the outward ray $x = m_i + s\sigma_i\, u(m_i)$ of a well met by no other support, a state is stationary iff

$$s\, e^{-s^2/2} = \kappa\rho^*. \tag{7}$$

For $\kappa < e^{-1/2}$ this equation has roots $s_1 < 1 < \rho^*$. The inner root is the stable fixed point, $s_1\sigma_i = 0.1024\, R_i$ outward of the memory, and the outer root is the support edge. Both are identical in every memory's own radius units. A linear outer potential makes this balance independent of the memory's distance from c; a quadratic outer potential does not (Appendix A.4).

Proposition 1 (an isolated support holds its trajectories)**.** *Let $\|m_i - m_j\| \geq R_i + R_j$ for every $j \neq i$. Under Eqs.* (4) *and* (5)*, every trajectory started in $B(m_i, R_i)$ stays in it, and the only stationary point in $B(m_i, R_i)$ on the line through c and m_i is $x_i^* = m_i + s_1\sigma_i\, u(m_i)$. If every pair of supports is disjoint, a start on the outward ray of m_i at or beyond its support edge never enters $B(m_i, R_i)$.*

The proof in Appendix A.2 bounds the next-step distance to the memory by an increasing map whose fixed point is R_i. Under a common dilation about c, the radii, local scales and updates dilate together, so the rule and trajectories depend on direction and on distance in each memory's own units.

2.4 ESCAPE AND FIXED-POINT GUARANTEES

The outward term provides a certified rejection outcome once a run reaches a region beyond every support.

Theorem 2 (escape is geometric)**.** *Let $U = \max_i(\nu_i + R_i)$ and let $U_c = \{x : \|x - c\| > U\}$ be the exit set. Under Eqs.* (4) *and* (5)*: (i) if $x_0 \neq c$ and $S(x_s) = \emptyset$ for all $s < t$, then*

$$x_t - c = (1 + \kappa\sin\theta)^t (x_0 - c);$$

(ii) U_c is forward invariant and every trajectory entering it diverges geometrically.

A cue whose ray meets no support therefore reaches U_c at step

$$\left\lfloor \frac{\ln(U/\rho_0)}{\ln(1 + \kappa\sin\theta)} \right\rfloor + 1, \qquad \rho_0 = \|x_0 - c\|,$$

provided the run is not stopped earlier. Reversing the outer slope confines rather than rejects:

Theorem 3 (the confining counterpart is bounded)**.** *Under Eqs.* (4) *and* (5) *with $+\lambda\|x - c\|$ in Eq.* (3)*,*

$$\|x_t - c\| \leq \max\left\{\|x_0 - c\|,\, U + \kappa \max_i R_i\right\} \qquad \textit{for all } t.$$

Proposition 4 (fixed points and rate)**.** *(i) The fixed points of Eq.* (4) *off the support spheres and c are the zeros of ∇E; none lies outside every support. (ii) With geometric radii, the update linearised at the isolated fixed point x_i^* contracts along the ray and transverse to it at rates that are the same for every memory (Appendix A.3).*

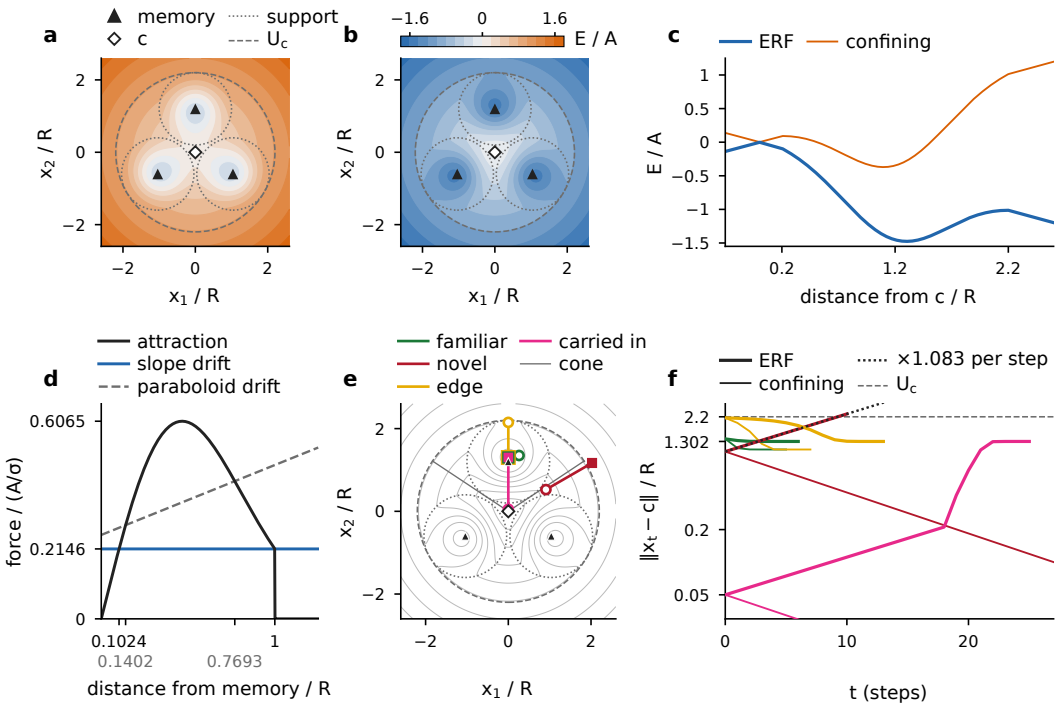


Figure 1: **One slope balances every well at its edge, so a trajectory that starts in an isolated support stays there.** Three memories $1.2\,R$ from c, $120°$ apart, with $R = \|m_i - c\| \sin\theta$ at $\theta = 56.44°$ and $\kappa = 0.1$; coordinates in units of R. **(a, b)** The confining energy ($+\lambda\|x-c\|$) and the ERF energy on one colour scale; dotted circles: supports; dashed circle: boundary of the exit set U_c (theorem 2). **(c)** Both energies on the line through c and memory 1. **(d)** Forces on the outward ray: the well's attraction equals the shared drift at the fixed point and at the support edge; a paraboloid's drift (dashed) moves both roots with the memory. **(e)** Runs from open circles to squares: supported cues settle at the fixed point, a novel cue outside every cone escapes, and a cue inside the drawn cone is carried into a support that the static rule does not count as recognition. **(f)** Outside every support, the ERF run moves away from c geometrically, while the confining counterpart remains bounded (theorem 3).

2.5 Choosing the local regions

The field above accepts any positive radii; the data determine which local regions are useful. We use two choices. *Geometric radii* take

$$R_i = \nu_i \sin\theta,$$

where $c = \bar{\mathcal{S}}$ and θ is the median angular nearest-neighbour distance in a label-free learning sample $\mathcal{S}$, fixed before any cue is seen. For CIFAR-100 and Tiny-ImageNet the learning sample is the store; elsewhere it is a fixed sample of at most 2000 vectors from the same data, drawn before the store. These radii cover the same half-angle around every memory direction and yield the exact missing-coordinate criterion tested in Section 3.1 and the capacity analysis of Appendix B.6.

Fitted radii are constructed in a metric adapted on top of a frozen encoder using only the stored items and their own synthetic corrupted views. Each radius is placed between the spread of one memory's own views and the closest views of other memories, as defined in Eq. (10). The associated separation statistic measures the fraction of memories for which these two sets of fitting distances are ordered. The fitted regime evaluates the static rule and the nearest-memory identity readout in this metric; its supports overlap, so the isolation guarantees above do not apply directly. Thus geometric and fitted radii specify the same local object under different assumptions, while the field guarantees remain explicitly conditional on the resulting support geometry.

2.6 FINITE-RUN PROTOCOL

A run applies Eqs. (4) and (5) from x_0 and ends at the first step t_{end} with

$$\begin{aligned} &\text{(a)} \quad \|x_t - x_{t-1}\| < \mathrm{tol}\,\frac{\ell_t}{\rho^*}, \quad S(x_t) \neq \emptyset; \\ &\text{(b)} \quad x_t \in U_c; \\ &\text{(c)} \quad t = T \end{aligned} \tag{8}$$

(fixed point, escape, budget). The run recognises x_0 iff $x_{t_{\mathrm{end}}}$ lies in a support and retrieves $\arg\min_i d_i(x_{t_{\mathrm{end}}})/R_i$. At $K = 6$, no verdict changes when tol, T or κ is varied over the tested range, or when the schedule is replaced by a constant step. Under crowding, constant steps can end runs on the budget, whereas the schedule in Eq. (5) yields finite measured runs up to $K = 1000$ (Appendix A.4).

3 EXPERIMENTS

We first test recognition and dynamical retrieval using geometric radii (Sections 3.1 and 3.2), and then test recognition using fitted radii and examine the relation between separation and performance (Sections 3.3 and 3.4). Familiar cues are corrupted stored items; novel cues come from the distractor sets specified for each experiment. FR is the fraction of familiar cues rejected, NR the fraction of novel cues rejected, and balanced accuracy is $\frac{1}{2}[(1 - \mathrm{FR}) + \mathrm{NR}]$. Identity accuracy is computed over recognised familiar cues in the geometric experiments and over all familiar cues in the fitted experiments. The fitted experiments evaluate static acceptance and identity rules; dynamical runs are reported separately in Section 3.2.

3.1 RECOGNITION WITH GEOMETRIC RADII

Suppose that a cue retains part of a stored image m and replaces the missing part by c. Then $x_0 - c = P(m - c)$, where P is the orthogonal projection that zeroes the missing coordinates, and $q = \|P(m - c)\|^2/\|m - c\|^2$ is the fraction of the image's squared deviation from c that the cue keeps. Orthogonality gives $d(x_0, m)^2 = (1 - q)\|m - c\|^2$, so in the memory's own units

$$\frac{d(x_0, m)}{R} = \frac{\sqrt{1-q}}{\sin\theta}, \qquad \text{so } x_0 \text{ lies in the support of } m \iff q > \cos^2\theta. \tag{9}$$

Such a cue is the point on its ray closest to m, so the cone criterion of Eq. (2) and the support of m agree for this cue. Contrast cancels from the relation, while the store enters through θ. The result is exact for a cue and its own memory; a cue with $q \leq \cos^2\theta$ can still lie in another memory's support. Under random dropout of a fraction f, concentration of q near $1 - f$ predicts a rejection transition near $f = 1 - \cos^2\theta$. The transition width and its position for the full store remain empirical.

Figure 2 tests this prediction on three stores from each of five datasets, whose mean angular spacing ranges from $30.4°$ to $57.9°$. The missing fraction at which half the partial images are rejected agrees with $1 - \cos^2\theta$ within 0.0049 for each of the 15 stores, with rejection changing from none to all over a narrow range. The predicted fractions are 0.632 for CIFAR-100 and 0.718 for Tiny-ImageNet. These results test the concentration of q and the influence of other supports; they are not a second exact consequence of Eq. (9). The protocol is given in Section B.

The one-exemplar experiment tests the same geometric rule on raw pixels, without fitting. On CIFAR-100 and Tiny-ImageNet, it matches or exceeds every registered comparator in all 18 conditions per store type (Table 5). Among recognised familiar cues, the source identity is recovered at rates 0.982 and 0.988, respectively. Section B.2 gives the rule-selection history, the subsequent registered runs and the post-hoc NNDD and score-ranking comparisons. We next check the corresponding dynamical behaviour.

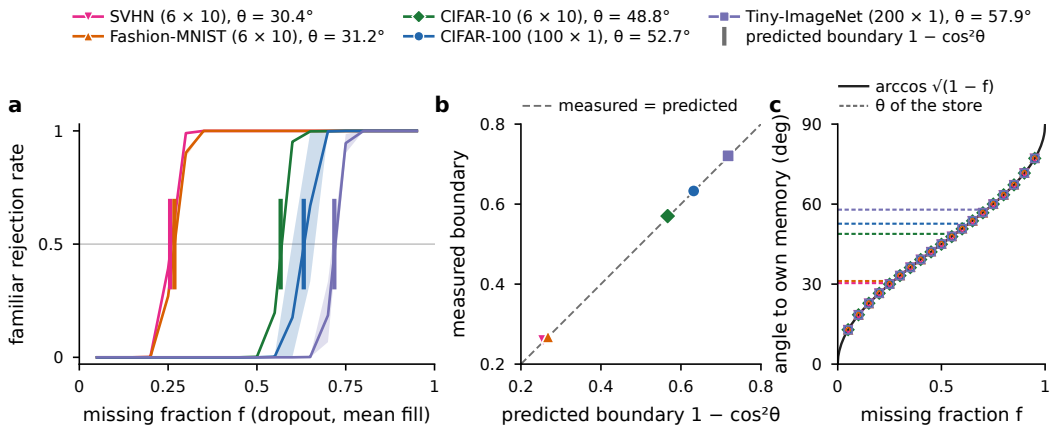


Figure 2: **Recognition under random-pixel dropout.** Stored images with a fraction f of pixels dropped at random and set to c; five datasets, three stores each, of ten images from each of six classes (SVHN, Fashion-MNIST, CIFAR-10) or one exemplar per class (CIFAR-100, Tiny-ImageNet). Curves are means and bands min–max over stores; c and θ come from a sample of the six classes (6×10 stores) or from the store (one exemplar per class, seeds 0–2). **(a)** Familiar rejection rate against f; ticks mark $1 - \cos^2\theta$. **(b)** Measured half-rejection boundary against the prediction. **(c)** Angle at c between a cue and its memory, with $\arccos\sqrt{1-f}$ and each store's θ: the cue leaves its own support when its angle reaches θ.

3.2 Retrieval and Escape

Runs of Eq. (4) test whether the geometric supports describe the observed retention and escape. On the three representation spaces of Section B, with $K = 6$, run and static rule agree on every tested cue, every rejected cue reaches U_c, and no run reaches its budget (Table 7). This agreement concerns the sampled cues; the carried-in cues of Section A.2 distinguish the two rules in general.

With ten MNIST digits stored, one per class, the ERF recognises the 583 among 1440 corrupted cues that start inside a support and retrieves every recognised cue correctly. Their end states follow the fixed-point displacement in Eq. (7); recovery results by corruption family are given in Fig. 5 and Table 8. One step of modern Hopfield retrieval (Ramsauer et al., 2021) returns an identity for every cue. It is wrong on 317 of the cues the ERF rejects and on 44 of those it recognises. These two subsets show how rejection relates to retrieval errors under the stated protocol.

With overlapping supports, up to $K = 1000$ feature memories, the scheduled runs still terminate within budget and agree with the static rule on every tested cue (Section B.5). Identity retrieval nevertheless degrades: at the largest ResNet store, only a fraction 0.74 of recognised familiar cues end nearest their source memory. Stable acceptance verdicts therefore do not imply preserved identity accuracy under crowding.

3.3 Recognition with Fitted Radii

Geometric radii specify tolerance through angular spacing. We also test radii and a metric fitted to corrupted views of the stored items. Masks remove 10–60 % of an image and fill the missing part with c; Gaussian noise supplies a second fitting family. A linear map W into 128 dimensions is fitted by InfoNCE, with a view as query, the K stored items as candidates and its source memory as target. The pretrained encoder remains frozen: the store determines the adaptation, but not the underlying representation. Write $z(x)$ for the normalised output of this map. Memory i receives

$$\begin{aligned} r_i &= \sqrt{a_i b_i}, \\ a_i &= Q_{0.9}\big\{\|z(v) - z(m_i)\| : v \in V_i\big\}, \\ b_i &= \min_{j \neq i} \min_{v \in V_j} \|z(v) - z(m_i)\|, \end{aligned} \tag{10}$$

Table 1: **Recognition and identity accuracy with fitted radii.** Two encoders, three stores and twenty corruption conditions, with five seeds (three at $K = 5000$). Sep. is the fraction of memories with $a_i < b_i$. The best comparator is selected after averaging each method over the setting's conditions. Identity uses all familiar cues; ERF uses the fitted metric and mHN the original encoder space. Section D distinguishes this aggregation from the diagnostic comparison.

Encoder	Store	Sep.	Detection (balanced accuracy)			Identity	
			ERF	best comparator	rule	ERF	mHN
DINOv2	CIFAR-100, $K = 100$	1.00	**0.922**	0.819	NNDD, fitted	0.962	0.736
DINOv2	DISC21, $K = 1000$	0.71	**0.830**	0.828	NNDD, fitted	0.869	0.647
DINOv2	DISC21, $K = 5000$	0.13	0.513	0.723	NNDD, fitted	0.769	0.592
ResNet-50	CIFAR-100, $K = 100$	1.00	**0.852**	0.720	NNDD, fitted	0.873	0.640
ResNet-50	DISC21, $K = 1000$	1.00	**0.813**	0.732	NNDD, fitted	0.820	0.536
ResNet-50	DISC21, $K = 5000$	0.37	0.699	0.750	NNDD, fitted	0.742	0.480

where V_i contains the fitting views of m_i. The ERF accepts if and only if $\min_i \|z(x) - z(m_i)\|/r_i < 1$ and reads out $\arg\min_i \|z(x) - z(m_i)\|$. The radii determine acceptance, whereas the fitted metric determines identity. Neither distractors nor evaluation cues enter either fit. We call the fraction of memories satisfying $a_i < b_i$ the store's *separation*. For $0 < a_i < b_i$, the radius lies strictly between the two distances; this is a property of the sampled fitting views, not a guarantee for unseen cues (Section C).

The confirmation experiment covers two frozen encoders and three stores: one CIFAR-100 image per class and either 1000 or 5000 DISC21 reference images. Six corruption families give twenty corruption conditions per store; with the distractor sets there are 200 conditions in all. Four families are absent from fitting. Table 1 reports detection and identity accuracy by store and encoder. Detection exceeds the setting mean of the strongest comparator in both CIFAR-100 settings and both 1000-image DISC21 settings, although the difference is small for DINOv2 on DISC21. It falls below that comparator at 5000 images under both encoders. The protocol and calibration are given in Section C.

Over all 200 conditions, the fitted readout identifies the source of 0.870 of familiar cues, against 0.638 for modern Hopfield in the encoder's original space, and is at least as accurate in all 200 conditions. Since z is normalised, maximum cosine and minimum Euclidean distance select the same memory: modern Hopfield given W has the same identity readout. The identity gain in this comparison therefore comes from the fitted metric. For acceptance, per-memory radii match or exceed one radius fitted on the same views in 60 of 60 ablation conditions, by 0.084 on average. In an earlier pixel-space comparison, geometric radii without the fitted metric reach 0.660 against 0.902 for the full method; that comparison changes both the metric and the radius construction.

3.4 Separation and recognition performance

The confirmation results suggested separation as a diagnostic. The threshold 0.9 was selected after those results were observed. For this analysis, the reference is the strongest comparator in each condition, after averaging over seeds; setting-level leads average these condition-level differences. This differs from selecting one comparator by its setting mean in Table 1.

The ERF matches or exceeds every comparator in 83 % of the 140 conditions with separation above 0.9, with mean balanced accuracy 0.876 against 0.801 for the condition-wise reference. At or below the threshold, it does so in 13 % of 60 conditions; its mean deficit is 0.091 and its largest deficit is 0.324. Within the high-separation subset it also leads each family's strongest mean comparator, including all four families absent from fitting: shift gives 0.992 against 0.961 and rotation 0.970 against 0.902 (Table 9). These family results concern that subset.

Separation depends on the representation and the fitting family as well as the store size. The 1000-image DISC21 store has separation one under ResNet-50 and 0.71 under DINOv2; its mean condition-wise lead is positive only in the former setting (Fig. 3). Widening the fitting tolerance on one store reduces separation from 0.820 to 0.366, while detection varies non-monotonically (Sec-

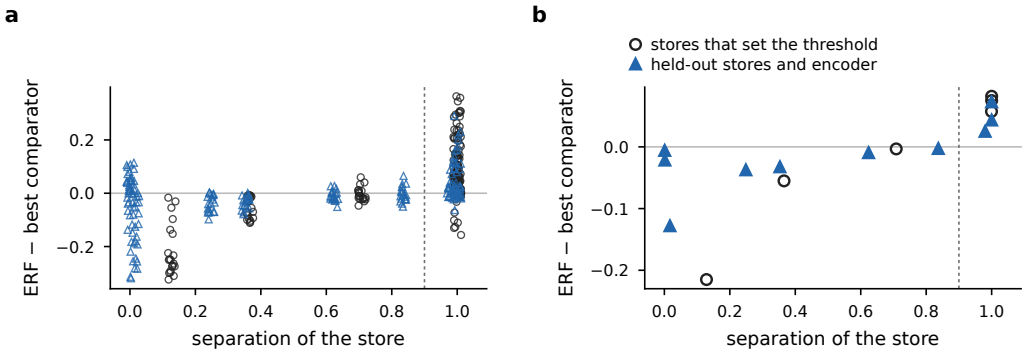


Figure 3: **Separation and relative recognition accuracy.** **(a)** ERF balanced accuracy minus the strongest comparator in each condition, for the six confirmation settings (circles) and 10 validation settings (triangles). **(b)** The same differences averaged over each store–encoder setting. The sign of each setting mean agrees with the threshold 0.9 in both experiments, although individual conditions can disagree. The comparator can differ between conditions; this aggregation differs from Table 1.

tion C). This is an empirical relation in the fitted regime; the isolated-support capacity guarantee does not apply to these overlapping supports.

The subsequent validation fixed the threshold at 0.9 and tested the prediction that the ERF matches or exceeds every comparator if and only if separation exceeds it. Five additional store sources under DINOv2 and ResNet-18, the latter absent from the fitted confirmation experiment, give 10 settings (Section D). The sign of the mean condition-wise lead agrees with the threshold in 10 of them, with mean leads of +0.047 above and -0.034 below the threshold. Both condition-level targets fail: prediction accuracy is 0.68, and the ERF matches or exceeds every comparator in 0.67 of conditions above the threshold, each below the registered target of 0.75. The evidence supports the diagnostic at the setting level in these experiments but does not establish a reliable prediction for individual corruption conditions.

4 Discussion

What the construction establishes. A memory that must know when not to answer usually borrows that decision from a detector calibrated on familiar data, which a store of exemplars lacks. The ERF puts the decision in the energy: each item acts only inside its own support, and one slope balances every well whatever the radii are. For any radii, an isolated support is forward invariant, so a cue starting in one is retrieved as its memory, and a cue whose ray from c meets no support leaves geometrically; the radii decide which cues start inside. Read from the store's angular spacing they give a tolerance in closed form and, for random directions, a capacity that falls as that tolerance rises. Fitted to the store's own views they give the accuracy: 0.870 of cues retrieved against 0.638 for modern Hopfield in the encoder's space, and the most accurate rejection wherever the store is separated.

What separates the guarantees from the measurements. The fitted supports overlap, so the isolation results do not apply to them, and the separation we measure is a statistic of the fitting views, not the pairwise disjointness the theory uses. Its threshold was read off the confirmation run; carried over unchanged it predicted the leading rule on all 10 further store–encoder settings but only 0.68 of their individual conditions, so it is a statement about stores.

Where it fails. A store must be separated by its own views. Crowding costs that—at $K = 5000$ separation falls to 0.13 and the ERF drops 0.324 below the best comparator—and so does the representation: the same 1000 images are separated under ResNet-50 and not under DINOv2. New images of a stored class, rather than corrupted copies of the stored one, are the next problem.

Reproducibility statement

Every number in the paper is written by a notebook from the result files of the experiment scripts, and each notebook ends with a verification cell that recomputes its claims; the code, the notebooks, the protocols of Sections B.2 and C and a one-command build are in the supplementary material.

AI use statement

In this work, we used generative AI tools for editing prose and table formatting. We have reviewed all AI-assisted work: AI-assisted text was revised by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

References

Keith Ball. An elementary introduction to modern convex geometry. In Silvio Levy (ed.), *Flavors of Geometry*, volume 31 of *MSRI Publications*, pp. 1–58. Cambridge University Press, 1997. URL https://www.cambridge.org/core/books/abs/flavors-of-geometry/an-elementary-introduction-to-modern-convex-geometry/09EA8B107A9ECA496914D95520C9EAFB.

Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations (ICLR)*, 2017. URL https://openreview.net/forum?id=Hkg4TI9xl.

Claus Hofmann, Simon Schmid, Bernhard Lehner, Daniel Klotz, and Sepp Hochreiter. Energy-based Hopfield boosting for out-of-distribution detection. In *Advances in Neural Information Processing Systems*, volume 37, pp. 131859–131919. Curran Associates, Inc., 2024. doi: 10.52202/079017-4192.

Benjamin Hoover, Zhaoyang Shi, Krishnakumar Balasubramanian, Dmitry Krotov, and Parikshit Ram. Dense associative memory with Epanechnikov energy. In *Advances in Neural Information Processing Systems*, volume 38, pp. 118328–118355. Curran Associates, Inc., 2025. doi: 10.52202/085713-3947.

John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982. doi: 10.1073/pnas.79.8.2554.

Georgios Iatropoulos, Johanni Brea, and Wulfram Gerstner. Kernel memory networks: A unifying framework for memory modeling. In *Advances in Neural Information Processing Systems*, volume 35, pp. 35326–35338. Curran Associates, Inc., 2022. doi: 10.52202/068431-2560.

Dmitry Krotov and John J. Hopfield. Dense associative memory for pattern recognition. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/hash/eaae339c4d89fc102edd9dbdb6a28915-Abstract.html.

Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *Advances in Neural Information Processing Systems*, volume 31, pp. 7167–7177. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/hash/abdeb6f575ac5c6676b747bca8d09cc2-Abstract.html.

Bo Liu, Hao Kang, Haoxiang Li, Gang Hua, and Nuno Vasconcelos. Few-shot open-set recognition using meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8798–8807, 2020a. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Liu_Few-Shot_Open-Set_Recognition_Using_Meta-Learning_CVPR_2020_paper.html.

Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21464–21475. Curran Associates, Inc., 2020b. URL https://proceedings.neurips.cc/paper/2020/hash/f5496252609c43eb8a3d147ab9b9c006-Abstract.html.

Pedro R. Mendes Júnior, Roberto M. de Souza, Rafael de O. Werneck, Bernardo V. Stein, Daniel V. Pazinato, Waldir R. de Almeida, Otávio A. B. Penatti, Ricardo da S. Torres, and Anderson Rocha. Nearest neighbors distance ratio open-set classifier. *Machine Learning*, 106(3):359–386, 2017. doi: 10.1007/s10994-016-5610-8.

Ryo Moriai, Nakamasa Inoue, Masayuki Tanaka, Rei Kawakami, Satoshi Ikehata, and Ikuro Sato. Rectified Lagrangian for out-of-distribution detection in modern Hopfield networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(18):19554–19562, 2025. doi: 10.1609/aaai.v39i18.34153.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Lukas Gruber, Markus Holzleitner, Thomas Adler, David Kreil, Michael Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield networks is all you need. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://openreview.net/forum?id=tL89RnzIiCd.

Douglas L. Reilly, Leon N. Cooper, and Charles Elbaum. A neural model for category learning. *Biological Cybernetics*, 45(1):35–41, 1982. doi: 10.1007/BF00387211.

Jie Ren, Stanislav Fort, Jeremiah Liu, Abhijit Guha Roy, Shreyas Padhy, and Balaji Lakshminarayanan. A simple fix to Mahalanobis distance for improving near-OOD detection. *arXiv preprint arXiv:2106.09022*, 2021. URL https://arxiv.org/abs/2106.09022.

Ethan M. Rudd, Lalit P. Jain, Walter J. Scheirer, and Terrance E. Boult. The extreme value machine. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):762–768, 2018. doi: 10.1109/TPAMI.2017.2707495.

Walter J. Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E. Boult. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1757–1772, 2013. doi: 10.1109/TPAMI.2012.256.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/hash/cb8da6767461f2812ae4290eac7cbc42-Abstract.html.

Yiyou Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 20827–20840. PMLR, 2022. URL https://proceedings.mlr.press/v162/sun22d.html.

David M. J. Tax. *One-class classification: Concept-learning in the absence of counter-examples*. PhD thesis, Delft University of Technology, 2001. URL https://research.tudelft.nl/en/publications/one-class-classification-concept-learning-in-the-absence-of-count/.

Engkarat Techapanurak, Masanori Suganuma, and Takayuki Okatani. Hyperparameter-free out-of-distribution detection using cosine similarity. In *Computer Vision – ACCV 2020*, volume 12625 of *Lecture Notes in Computer Science*, pp. 53–69, Cham, 2021. Springer. doi: 10.1007/978-3-030-69538-5_4.

Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/hash/90e1357833654983612fb05e3ec9148c-Abstract.html.

Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *International Journal of Computer Vision*, 132(12):5635–5662, 2024. doi: 10.1007/s11263-024-02117-4.

A PROOFS AND CHECKS FOR SECTION 3

This appendix follows the theoretical and empirical branches of Sections 2 and 3. Appendix A gives the proofs and supplementary checks for the field construction: Appendix A.1 derives the update and its scaling properties, Appendix A.2 establishes boundary balance, isolated-support retention and ray capture, Appendix A.3 proves the escape, boundedness and fixed-point claims, and Appendix A.4 collects the finite-run and numerical checks. Appendix B documents the geometric-radius experiments of Sections 3.1 and 3.2, and Appendix B.6 analyses their capacity and tolerance. Appendix C gives the fitted-radius protocols and ablations of Section 3.3, while Appendix D reports the subsequent validation of the separation diagnostic in Section 3.4. Forward invariance of an isolated support is proved for the discrete update; convergence within a support is assessed numerically.

Throughout, $u(x) = (x - c)/\|x - c\|$ for $x \neq c$, $\nu_i = \|m_i - c\|$, the drift of Eq. (4) is defined to be zero at c, and the constants are those of Eq. (6). Most proofs reduce to a single-well argument expressed in memory i's own units, with length scales σ_i and $R_i = \rho^* \sigma_i$. Figure 4 then checks the corresponding geometric and finite-run predictions on data.

A.1 FIELD CONSTRUCTION, GRADIENT FORM AND SCALING

Table 2: **Default field parameters and run settings.**

Parameter	Default or definition	Role
κ	$1/10$	Kernel cutoff
θ	Angular spacing of $\mathcal{S}$	Support geometry
c	Mean of $\mathcal{S}$	Reference point
$\mathcal{S}$	$n = \min(2000, N)$, seed 7	Learning sample
τ_t	$1/\max(1, W_t)$	Step size
tol	$10^{-4}\sigma_i$	Stopping tolerance
T	400	Step budget
U	$\max_i(\|m_i - c\| + R_i)$	Escape threshold

Parameter choices. Table 2 summarises the defaults. The kernel cutoff κ is the only free scalar parameter of the field once the radii are specified; c and θ are estimated before any cue is presented. Cutting the kernel at a tenth of its peak gives $\rho^* = 2.146$, well depth $0.9A_i$, fixed-point displacement $0.1024\, R_i$ and an escape factor $1 + \kappa \sin\theta$ per step. The angular-spacing criterion makes each support cover the angle at which a typical stored item meets its angular nearest neighbour. It sets $R_i = \|m_i - c\| \sin\theta$, $\sigma_i = R_i/\rho^*$ and $A_i \propto R_i$, with one shared slope λ (Eq. (6)). The reference point c is the apex of the slope and cones; translation and dilation covariance are discussed below.

Learning sample and observed geometry. The sample $\mathcal{S}$ is drawn before the store: MNIST digits 0–5 for every MNIST store, including the ten-digit store; six classes of SVHN, Fashion-MNIST and CIFAR-10; or the 800 prototypes not held out. For CIFAR-100 and Tiny-ImageNet, the store itself is the learning sample. Angular spacing is $48.8°$ for CIFAR-10 pixels, $48.2°$ for MNIST pixels, $40.4°$ for ResNet-MNIST, $63.6°$ for prototypes and $6.6°$ for CIFAR-10 ResNet-18. With one exemplar per class, the ranges are 52.2–$58.3°$ for CIFAR-100 and 55.8–$58.6°$ for Tiny-ImageNet.

Here $\mathrm{nn}_{\mathrm{med}}$ is the median nearest-neighbour distance of $\mathcal{S}$, used only as a reporting scale. For MNIST pixels, ResNet-MNIST and prototypes, respectively, the median $R_i/\mathrm{nn}_{\mathrm{med}}$ over memories is 0.88, 0.90 and 0.88, and the median $U/\mathrm{nn}_{\mathrm{med}}$ over seeds is 2.89, 2.65 and 2.00. The budget $T = 400$ is at least 50 times the median run length of 8 steps. At $K = 6$, varying the stopping tolerance, budget or constant step changes run speed but not the verdict (Table 4); scheduled steps maintain stable verdicts among overlapping wells.

Gradient-descent form. E is differentiable off the support spheres $\{d_i(x) = R_i\}$ and off c. With $A_i/\sigma_i^2 = \rho^{*2}/(R_i R_{\mathrm{ref}})$ and $\lambda = \kappa\rho^{*2}/R_{\mathrm{ref}}$,

$$\nabla E(x) = \sum_{i \in S(x)} \frac{A_i}{\sigma_i^2}\, w_i\,(x - m_i) - \lambda\, u(x)$$

$$= \frac{\rho^{*2}}{R_{\mathrm{ref}}}\Big[\sum_{i \in S(x)} \frac{w_i}{R_i}(x - m_i) - \kappa\, u(x)\Big],$$

so $x_t - \eta_t \nabla E(x_t)$ with $\eta_t = \tau_t \kappa \ell_t/\lambda = \tau_t \ell_t R_{\mathrm{ref}}/\rho^{*2}$ is Eq. (4). The pull of well i at its support edge is $(A_i/\sigma_i^2) R_i e^{-R_i^2/2\sigma_i^2} = \rho^{*2}\kappa/R_{\mathrm{ref}} = \lambda$ for every i: one slope balances every well. $\ell_t > 0$ for $x_t \neq c$ (inside a support ℓ_t is a weighted mean of the active radii; outside every support it is $\|x_t - c\| \sin\theta$), so $\eta_t > 0$ there.

Homogeneity. Let $a > 0$ and $D_a(x) = c + a(x - c)$, and replace every memory m_i by $D_a(m_i)$, with c, θ and κ unchanged. Then R_i, σ_i and U scale by a and, at $x' = D_a(x)$, the active set, the weights w_i, W, τ and u are those at x while ℓ scales by a. Every term of the bracket of Eq. (4) scales by a, so $x'_{t+1} = D_a(x_{t+1})$: the trajectory of $D_a(x_0)$ is the dilated trajectory of x_0. Exit (a) of Eq. (8) compares $\|x_t - x_{t-1}\|$ with $\mathrm{tol}\,\ell_t/\rho^*$ and exit (b) $\|x_t - c\|$ with U; both are invariant, so the ending step and the verdict are unchanged. With R_{ref} held as the unit, A_i and the slope term scale by a and $E(D_a(x); D_a(M)) = a\, E(x; M)$: the energy is positively homogeneous of degree one about c.

A.2 Boundary balance, isolated supports and ray capture

Derivation of Eq. (7). Let no other support meet $B(m_i, R_i)$ and $x(s) = m_i + s\sigma_i \hat{u}$, $\hat{u} = u(m_i)$, $s > -\nu_i/\sigma_i$. For $|s| < \rho^*$ only i is active, so $\ell = R_i = \rho^*\sigma_i$, $W = w \leq 1$, $\tau = 1$, and the bracket of Eq. (4) is $\kappa\rho^*\sigma_i\hat{u} - e^{-s^2/2} s\sigma_i \hat{u}$ for $s > 0$. It vanishes iff $s\, e^{-s^2/2} = \kappa\rho^*$. The function $g(s) = s\, e^{-s^2/2}$ rises on $[0, 1]$ to $e^{-1/2}$ and falls on $[1, \infty)$; ρ^* is a root because $\rho^* e^{-\rho^{*2}/2} = \kappa\rho^*$. For $\kappa < e^{-1/2}$, $\rho^* > 1$, and the other root $s_1 \in (0, 1)$ exists iff $\kappa\rho^* < e^{-1/2}$: with $v = \ln(1/\kappa)$, $h(v) = \sqrt{2v}\, e^{-v}$ has the single maximum $h(1/2) = e^{-1/2}$ at $\kappa = e^{-1/2}$, so $\kappa\rho^* < e^{-1/2}$ for every other κ. On (s_1, ρ^*) the pull exceeds the push. For $-\nu_i/\sigma_i < s < 0$ the pull and the drift point the same way, so the inward ray holds no stationary point. No quantity in Eq. (7) involves ν_i or R_i: the fixed point is $s_1\sigma_i = (s_1/\rho^*)R_i$ outward and the edge R_i, in every memory's own units.

Proof of theorem 1: isolated-support retention (A1).

Support invariance. Let $\|m_i - m_j\| \geq R_i + R_j$ for every $j \neq i$ and $x_t \in B(m_i, R_i)$. Then $S(x_t) = \{i\}$, $\ell_t = R_i$, $W_t = w_i \leq 1$ and $\tau_t = 1$. With $d_t = \|x_t - m_i\|$ and $w_t = e^{-d_t^2/2\sigma_i^2}$, Eq. (4) gives

$$x_{t+1} - m_i = (1 - w_t)(x_t - m_i) + \kappa R_i\, u(x_t),$$
$$d_{t+1} \leq (1 - w_t) d_t + \kappa R_i =: b(d_t).$$

The function b is increasing on $[0, R_i]$, since

$$b'(d) = 1 - e^{-d^2/2\sigma_i^2}(1 - d^2/\sigma_i^2) \geq 0,$$

with equality only at $d = 0$. Moreover, $b(R_i) = (1 - \kappa)R_i + \kappa R_i = R_i$, so $d_t < R_i$ implies $d_{t+1} < R_i$. The same bound holds for a constant step $\tau \leq 1$, with $b_\tau(d) = (1 - \tau w)d + \tau\kappa R_i$ and $b_\tau(R_i) = R_i$. A run started inside the support ends by (a) or (c) of Eq. (8) with $i \in S(x_{t_{\mathrm{end}}})$: the run recognises x_0 as m_i, as Eq. (1) does for every cue with $d_i(x_0) < R_i$.

Outward-ray starts. On the outward ray of m_i a start at or beyond the support edge lies outside $B(m_i, R_i)$. While it lies in no support it moves outward along that ray (theorem 2(i)), away from m_i; when every pair of supports is disjoint it stays in any support it enters (A1) and never leaves U_c once there (theorem 2(ii)), so it never enters $B(m_i, R_i)$. This is the last claim of theorem 1. □

Convergence and numerical check. The preceding argument proves forward invariance for the discrete update but not convergence to the interior fixed point. In memory i's units, $y = (x - c)/\sigma_i$, the frozen-drift flow $\dot{y} = \kappa\rho^*\hat{u} - w(y)(y - \mu)$, $\mu = (\nu_i/\sigma_i)\hat{u}$, is the gradient flow of $-\phi(\|y - \mu\|) - \kappa\rho^*\langle y, \hat{u}\rangle$; the support is invariant under it ($\dot{d} = \kappa\rho^* \cos\alpha - d\, e^{-d^2/2} < 0$ on $s_1 < d < \rho^*$, α the angle between $y - \mu$ and $\hat{u}$) and $\mu + s_1\hat{u}$ is its only stationary point, so every trajectory of that flow inside the support converges to it.

This continuous-flow argument still does not establish convergence of the discrete update; that claim is therefore assessed numerically rather than asserted as part of theorem 1. Figure 4a sweeps starts at $\|m_0 - c\| = 1.342\, R_0$: 5019 of the 14641 starts lie inside the support, 0 of them are rejected, and every recognised start ends within $4.7 \times 10^{-6}\, R_0$ of one point $s_1\sigma_0$ outward of the memory.

Ray capture and the cone at θ (A2). While $S(x_t) = \emptyset$, the update has $\tau_t = 1$ and $\ell_t = \|x_t - c\| \sin\theta$, hence

$$x_{t+1} - c = (1 + \kappa \sin\theta)(x_t - c).$$

The state therefore remains on the ray from c through $x_0 \neq c$, and its distance from c grows by the factor $1 + \kappa \sin\theta$ per step. It stays on that ray until it enters a support or U_c (theorem 2); once in U_c it never enters a support (theorem 2(ii)), and once in an isolated support it stays (A1). So for a store of isolated wells the run recognises the cue iff the discrete ray enters a support. In angle, with $\varphi = \angle(x_0 - c,\, m_i - c)$: the line through c and x_0 passes $\nu_i \sin\varphi$ from m_i and its closest point lies at arclength $\nu_i \cos\varphi$ from c, so the forward ray meets $B(m_i, R_i) = B(m_i, \nu_i \sin\theta)$ iff $\sin\varphi < \sin\theta$, that is $\varphi < \theta$, and $\nu_i \cos\varphi \geq \|x_0 - c\|$ (the case of a closest approach behind x_0 needs $d_i(x_0) < R_i$, which is excluded outside every support). This is Eq. (2) with x_0 between c and the support. $R_i/\nu_i = \sin\theta$ for every memory, so every memory's cone has the half-angle θ; with one radius R for every memory the half-angle would be $\arcsin(R/\nu_i)$, wide for memories near c and narrow for memories far from it.

The grazing band. A step from a state at distance ρ from c has length $\kappa\rho\sin\theta$. For a memory ahead of the state, $\nu_i \geq \nu_i \cos\varphi \geq \rho$, so $R_i = \nu_i \sin\theta \geq \rho \sin\theta$ and the step is at most κR_i, a tenth of the support's radius. The ray meets $B(m_i, R_i)$ on the arclength interval $[a, b]$, $a, b = \nu_i(\cos\varphi \mp h)$, $h = \sqrt{\sin^2\theta - \sin^2\varphi}$, a chord of length $2\nu_i h$. The states on the ray are $\rho_0(1 + \kappa\sin\theta)^k$, and one step crosses the support without landing in it only if $\rho_k < a$ and $\rho_{k+1} > b$, which needs

$$\frac{\cos\varphi + h}{\cos\varphi - h} < 1 + \kappa \sin\theta,$$

that is a chord shorter than the step. The condition involves φ, θ and κ alone, so the band is the same for every memory; h vanishes at $\varphi = \theta$, and to leading order in κ the condition holds for $\theta - \varphi < (\kappa^2/16)\sin 2\theta$. Outside that band a ray that meets a support ahead lands in it.

Consequences for static and dynamical decisions. The carried-in cue of Fig. 1e lies on the segment from c to m_1 ($\varphi = 0$, chord $2R_1$), outside every support: its distance from c grows geometrically and it enters the support at step 18. Equation (1) recognises every cue inside a support and rejects every other cue. For a store of isolated wells a run recognises every cue inside a support (A1) and rejects every cue that escapes, so the carried-in cues are the only ones on which the run and Eq. (1) differ. In the plane of Fig. 4a, with $\nu_0/R_0 = 1/\sin\theta = 1.342$, they form the cone of half-angle θ from c, cut at the tangency chord: they are exactly the 264 starts outside the support with $|u_\perp| < (u_{\text{out}} + \nu_0/R_0)\tan\theta$ and $u_{\text{out}} < -\sin\theta = -0.745$; a start in the cone beyond the chord has passed its closest approach to the memory and escapes.

A.3 Escape, boundedness and fixed points

Proof of theorem 2.

(i) Motion outside all supports. If $S(x_s) = \emptyset$, then $W_s = 0$, $\tau_s = 1$ and $\ell_s = \|x_s - c\| \sin\theta$. Hence $x_{s+1} - c = (1 + \kappa\sin\theta)(x_s - c)$, so $x_{s+1} \neq c$; induction from $x_0 \neq c$ gives the claim.

(ii) Forward invariance of the exit set. For $x \in U_c$ and every i, $\|x - m_i\| \geq \|x - c\| - \nu_i > U - \nu_i \geq R_i$, so $S(x) = \emptyset$. By part (i), $\|x' - c\| = (1 + \kappa\sin\theta)\|x - c\| > U$, and therefore

$x' \in U_c$. Hence the distance from c grows by the factor $1 + \kappa \sin\theta$ at every later step, $S(x_t) = \emptyset$ for every $t \geq t_U$, and the verdict of Eq. (8) at any budget $T \geq t_U$ is rejection, the verdict returned at t_U. The exit step of Section 2.4 follows from (i) with $\rho_0 = \|x_0 - c\| \leq U$: the first t with $\rho_0(1 + \kappa \sin\theta)^t > U$ is $\lfloor \ln(U/\rho_0)/\ln(1 + \kappa \sin\theta) \rfloor + 1$. □

Proof of theorem 3. With $+\lambda\|x - c\|$ the update is

$$x_{t+1} = x_t + \tau_t \Big[-\kappa \ell_t \, u(x_t) - \sum_{i \in S(x_t)} \frac{\ell_t}{R_i} w_i (x_t - m_i) \Big].$$

If $S(x_t) = \emptyset$, then $\tau_t = 1$ and $x_{t+1} - c = (1 - \kappa \sin\theta)(x_t - c)$: the distance from c falls by the factor $1 - \kappa \sin\theta \in (0, 1)$ and the state never overshoots c (the confining runs of the novel and the carried-in cue in Fig. 1f contract towards c until the budget). If $S(x_t) \neq \emptyset$, write $x_t - m_i = (x_t - c) - (m_i - c)$:

$$\begin{aligned} x_{t+1} - c = (1 - \tau_t W_t)(x_t - c) \\ + \tau_t \sum_{i \in S(x_t)} \frac{\ell_t}{R_i} w_i \, (m_i - c) - \tau_t \kappa \ell_t \, u(x_t). \end{aligned}$$

For $W_t \leq 1$, $\tau_t = 1$ and the first two terms form a convex combination of $x_t - c$ and the $m_i - c$ (weights $1 - W_t$ and $(\ell_t/R_i)w_i$); for $W_t > 1$, $\tau_t = 1/W_t$, the first coefficient is 0 and the second term is a convex combination of the $m_i - c$. In both cases their norm is at most $\max\{\|x_t - c\|, \max_i \nu_i\} < U$, since some i is active and $\|x_t - c\| < \nu_i + R_i \leq U$. The drift adds at most $\tau_t \kappa \ell_t \leq \kappa \max_i R_i$, because ℓ_t is a weighted mean of active radii. So $\|x_{t+1} - c\| < U + \kappa \max_i R_i$, and with the first case $\|x_{t+1} - c\| \leq \max\{\|x_t - c\|, U + \kappa \max_i R_i\}$; induction gives the bound. □

Proof of theorem 4.

(i) Stationary points. Write Eq. (4) as $x_{t+1} = x_t + \tau_t G(x_t)$ with

$$\begin{aligned} G(x) &= \kappa \ell(x) \, u(x) - \sum_{i \in S(x)} \frac{\ell(x)}{R_i} w_i (x - m_i) \\ &= -\ell(x) \frac{R_{\mathrm{ref}}}{\rho^{*2}} \nabla E(x) \end{aligned}$$

off the support spheres and c. With $\tau_t > 0$ and $\ell(x) > 0$ for $x \neq c$, $x_{t+1} = x_t$ iff $\nabla E(x_t) = 0$; the schedule enters only through the positive factor τ_t. If $S(x) = \emptyset$ and $x \neq c$ then $\|G(x)\| = \kappa \ell(x) > 0$, so no fixed point other than c lies outside every support; c is fixed when $S(c) = \emptyset$ ($\ell = 0$ there).

(ii) Linearisation at an isolated fixed point. For geometric radii, $R_i = \nu_i \sin\theta$. Only memory i is active near x^*, so $\ell = R_i$, $W = w < 1$, $\tau_t = 1$, and the update is the single-well update in memory i's units. With $y = (x - c)/\sigma_i$, $\nu = \nu_i/\sigma_i = \rho^*/\sin\theta$, $G(y) = \kappa \rho^* \hat{y} - w(y)(y - \mu)$, $w(y) = e^{-\|y - \mu\|^2/2}$, $\mu = \nu \hat{u}$, and $y^* = \mu + s_1 \hat{u}$ is a zero by Eq. (7), inside the support since $s_1 < \rho^*$. Every zero of G satisfies $w(y)(y - \mu) = \kappa \rho^* \hat{y}$, so y lies on the line through c and μ, and since $\nu > \rho^* > s_1$ the only zero in the open support is y^*. Using $\nabla w = -w \, (y - \mu)$ and $\partial \hat{y}/\partial y = (I - \hat{y}\hat{y}^\top)/\|y\|$,

$$\begin{aligned} DG(y^*) &= \frac{\kappa \rho^*}{\nu + s_1} \big(I - \hat{u}\hat{u}^\top\big) - w^* \big(I - s_1^2 \, \hat{u}\hat{u}^\top\big), \\ w^* &= e^{-s_1^2/2}. \end{aligned}$$

For a constant step τ, the linearised map $I + \tau DG(y^*)$ has eigenvalues

$$\begin{aligned} m_r(\tau) &= 1 - \tau(1 - s_1^2) w^*, \\ m_\perp(\tau) &= 1 - \tau w^* + \frac{\tau \kappa \rho^*}{\nu + s_1} = 1 - \tau w^* + \frac{\tau \kappa \rho^* \sin\theta}{\rho^* + s_1 \sin\theta}, \end{aligned}$$

along $\hat{u}$ and on the transverse space, respectively. The last equality uses $\nu = \rho^*/\sin\theta$, so the multipliers are the same for every memory. At $\tau = 1$ the first lies in $(0, 1)$ because $0 < (1 - s_1^2) w^* < 1$; with $\kappa \rho^* = s_1 w^*$ the second equals $1 - w^* \nu/(\nu + s_1) \in (0, 1)$. The radial multiplier vanishes at $\tau = 1/((1 - s_1^2) w^*) = 1.076$, where the familiar count of Fig. 4e has its minimum. □

A.4 FINITE-RUN AND SUPPLEMENTARY NUMERICAL CHECKS

The preceding results concern the ideal update and its local geometry. This subsection checks the alternative outer potential, finite stopping rule, step schedule and parameter sensitivity used in the reported runs.

The paraboloid. With equal radii $R = \rho^*\sigma$ the field is the single-radius field, and with the outer term $-\varepsilon\|x - c\|^2$, $\varepsilon = \kappa A/2\sigma^2$, the drift at x has magnitude $\kappa\|x - c\|/\sigma$ in units of A/σ, so the balance on the outward ray of a memory at ν_i from c reads $s\,e^{-s^2/2} = \kappa(\nu_i/\sigma + s)$: its roots move with ν_i (Fig. 1d) and exist only for $\nu_i/\sigma \le \max_s \left[s\,e^{-s^2/2}/\kappa - s\right]$, that is $\nu_i \le 2.379\,R$ at $\kappa = 1/10$. Table 3 compares the two outer terms at $R = \mathrm{nn}_{\mathrm{med}}$: on the three spaces of Section 3.2, where every memory lies below the bound, they return the same verdicts; on CIFAR-10 ResNet-18 features, where every memory lies above it (median distance $8.47\,\mathrm{nn}_{\mathrm{med}}$ from c), the paraboloid rejects displaced stored vectors that the linear slope recognises, and the linear slope recognises every memory presented as its own cue.

Table 3: **Linear and quadratic outer potentials.** Comparison at a common radius $R = \mathrm{nn}_{\mathrm{med}}$.

Space	K	Outer term	FR	NR	Agree.	Above bound	Memory as cue
pixels	6	slope	0.090	0.997	1.000	0/36	36/36
		paraboloid	0.090	0.997	1.000	0/36	36/36
ResNet	6	slope	0.000	1.000	1.000	0/36	36/36
		paraboloid	0.000	1.000	1.000	0/36	36/36
prototypes	6	slope	0.000	0.990	1.000	0/36	36/36
		paraboloid	0.000	0.990	1.000	0/36	36/36
CIFAR	10	slope	0.000	1.000	1.000	20/20	20/20
		paraboloid	1.000	1.000	0.500	20/20	0/20
CIFAR	100	slope	0.000	1.000	1.000	200/200	200/200
		paraboloid	1.000	1.000	0.500	200/200	0/200
CIFAR	1000	slope	0.000	0.998	1.000	2000/2000	2000/2000
		paraboloid	0.946	1.000	0.526	2000/2000	149/2000

Notes. FR, NR and agreement compare the run with the ball rule $d_0 < R$. The last two columns count memories beyond the paraboloid bound and memories recognised as their own cues. The three spaces of Section 3.2 use seeds 0–5 (36 memories); CIFAR-10 ResNet-18 features use seeds 0–1, with 1000 displaced familiar cues and 1000 SVHN cues per run. The two potentials agree on the three spaces; on CIFAR-10 features, the paraboloid rejects almost every displaced stored vector.

Table 4: **Verdict stability under protocol variations at $K = 6$.**

Setting	Tested values	Verdict changes
Constant step τ	0.1, 0.25, 0.5, 1, 1.5, 1.8	0 / 10800
Tolerance (σ_i)	10^{-2}, 10^{-3}, 10^{-4}, 10^{-6}	0 / 7200
Budget T	50, 100, 400, 4000	0 / 7200
No early stop (T)	400, 2000	0 / 3600
Kernel cutoff κ	0.02, 0.05, 0.2, 0.3	0 / 7200

Changes / tested cues, relative to scheduled steps, tol $10^{-4}\sigma_i$, $T = 400$ and $\kappa = 1/10$. Each setting covers the three spaces of Section 3.2, six seeds and 100 cues per run.

Termination checks. None of the 55441 early-stopping runs in Fig. 4a, c–f ends on the budget. This pool covers the verdict map, angular sweep, scheduled runs, speed settings and kernel-cutoff sweep, including constant $\tau \le 1.8$ at $K = 6$; it excludes the crowding runs of Section B.5. Under

crowding, a constant step $\tau = 1$ instead gives 20 budget-ended runs out of 1600 at $K \leq 1000$. Step counts and fixed-point displacements are shown in Fig. 4e–f.

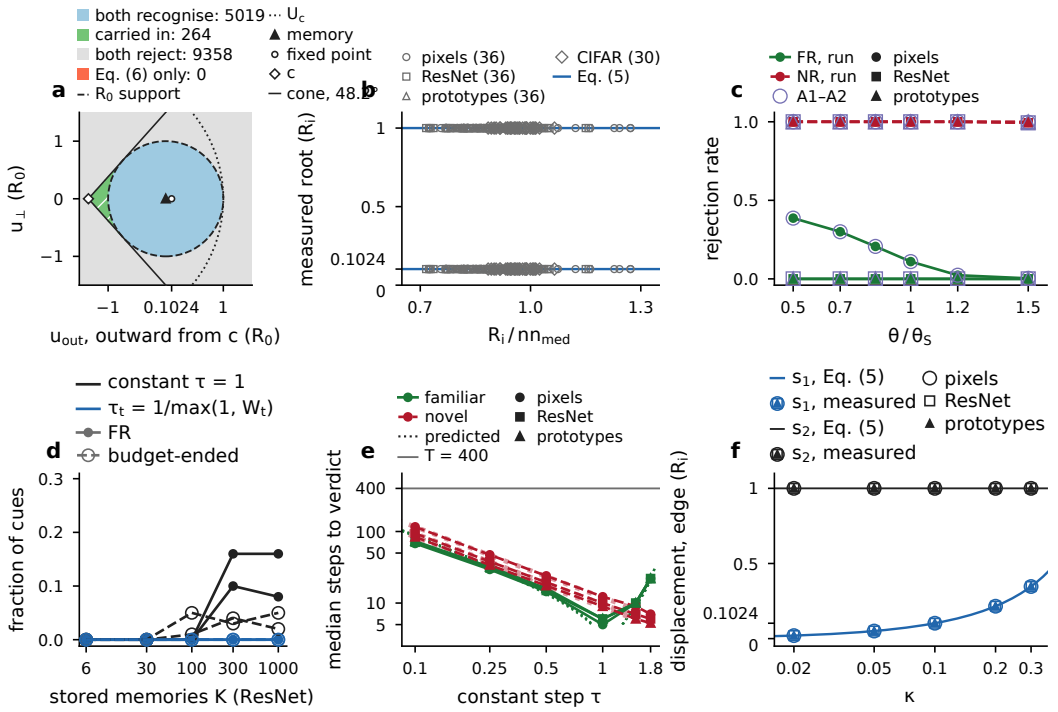


Figure 4: **Numerical checks of the field geometry and run protocol.** **(a)** Run verdicts around a stored pixel memory; the support and capture cone are overlaid. **(b)** Fixed-point displacement and support edge versus support radius, compared with Eq. (7). **(c)** FR and NR versus the relative half-angle θ/θ_S; lines show runs and rings show A1–A2. **(d)** FR and budget-ended fractions under crowding, with constant and scheduled steps. **(e)** Median steps to the verdict versus constant step size; dotted curves give the reference predictions. **(f)** Displacement and support edge versus κ; lines show Eq. (7) and markers show measured medians. Marker shapes identify the representation spaces. Settings and numerical comparisons are given in the accompanying text.

Protocol settings for Fig. 4. Unless varied in the relevant panel, the experiments use $\kappa = 0.1$, $K = 6$, seeds 0–5, and 50 familiar plus 50 novel cues per run. The three representation spaces are MNIST pixels, ResNet-18 features of MNIST and ImageNet prototypes. Panel b additionally includes CIFAR-10 ResNet-18 features; panel d varies K and uses seeds 0 and 1.

Basin geometry and fixed points (a–b). Panel a classifies 121×121 starting points on $[-1.5, 1.5]^2 R_0$ by the run and by Eq. (1), with counts in the legend. The plane passes through pixel memory m_0 (seed 0), with axes along its outward direction u_{out} from c and a perpendicular $u_\perp$; distances are in units of R_0. Here $\|m_0 - c\| = 1.342\, R_0 = R_0/\sin\theta$ and $\theta = 48.2°$. No run from inside the support rejects its cue (A1). The 264 starts carried into the support fill the cone of half-angle θ, cut at the tangency chord $u_{\text{out}} = -\sin\theta\, R_0 = -0.745\, R_0$ (A2). Together, the support and cone predict the verdict for 14641 of the 14641 starts. Every recognised start ends at the fixed point $s_1\sigma_0 = 0.1024\, R_0$ outward of m_0, marked by a tick.

Panel b measures s_1 and the outward edge s_2 for 138 memories: 108 from the three spaces and 30 from CIFAR-10 ResNet-18 (seeds 0–2). The radii span $R_i/\text{nn}_{\text{med}} \in [0.72, 1.27]$. In each memory's own units, the measurements agree with the roots of Eq. (7), $s_1\sigma_i = 0.1024\, R_i$ and $s_2\sigma_i = R_i$, within $1.8 \times 10^{-6}\, \sigma_i$ and $9.6 \times 10^{-8}\, \sigma_i$, respectively. Here s_2 is the largest recognised start on the outward ray. All 138 of the 138 memories are recognised when presented as their own cues, and all return their own identity.

Angular tolerance and crowding (c–d). Panel c varies the half-angle relative to the measured angular spacing θ_S and reports mean FR and NR over six seeds. The runs agree with A1–A2 in all 108 runs; pixel FR is 0.39/0.30/0.21/0.11/0.02/0.00 at $\theta/\theta_S = 0.5/0.7/0.85/1/1.2/1.5$. Panel d varies the number of stored ResNet memories, with one curve per seed. At $K \geq 300$, constant steps $\tau_t = 1$ give FR 0.08–0.16 ($\max_t W_t$ up to 16.4). The schedule in Eq. (5) keeps FR at 0.00 with 0 budget-ended runs.

Step size and kernel cutoff (e–f). Panel e reports median steps to a fixed point for familiar cues and to U_c for novel cues. The dotted reference curves are

$$\frac{\ln(U/\rho_0)}{\ln(1+\tau\kappa\sin\theta)} \quad \text{(novel)}, \qquad \frac{\ln(\delta_0/\mathrm{tol})}{-\ln|m_r(\tau)|} \quad \text{(familiar)},$$

where m_r is the radial multiplier in theorem 4 and δ_0 is the median first-step displacement. Each of the 6300 novel runs ends at the floor of the novel-cue prediction plus one; the horizontal line marks the budget $T = 400$. Across the 16 speed settings in Table 4 (constant τ, tol, T and no early stop), 0 of 28800 verdicts change.

Panel f varies $\kappa \in \{0.02, 0.05, 0.1, 0.2, 0.3\}$ and compares measured medians in each space with Eq. (7), using nested markers for the measurements. The displacement changes with κ, while the support edge remains at R_i. Relative to $\kappa = 0.1$, 0 of 7200 verdicts change.

B EXPERIMENTS WITH GEOMETRIC RADII

This section gives the protocols and supplementary results for Sections 3.1 and 3.2. The first two subsections evaluate the static geometric rule, and the remaining three examine dynamical retention, retrieval and escape. This order mirrors the geometric branch of Section 3: recognition is tested first, followed by the behaviour of the field.

B.1 RANDOM-PIXEL DROPOUT

Figure 2 uses three stores from each of five datasets. SVHN, Fashion-MNIST and CIFAR-10 stores contain ten images from each of six classes; CIFAR-100 and Tiny-ImageNet stores contain one exemplar per class. For the six-class experiments, c and θ are obtained from a sample of those classes; for the one-exemplar experiments they are obtained from the store (seeds 0–2). Missing pixels are chosen at random and set to c.

For each missing fraction f, the static rule gives the familiar rejection rate. The measured boundary is the value of f at which this rate crosses one half, compared with $1 - \cos^2\theta$. Means and min–max bands summarise the stores of each dataset. The cue–memory angle is also compared with $\arccos\sqrt{1-f}$. These aggregate measurements supplement the exact own-support relation in Eq. (9); acceptance by another support is included in the measured store-level verdict.

B.2 ONE-EXEMPLAR RECOGNITION

This experiment evaluates Eq. (1) on raw pixels, with radii read from the store's angular spacing and comparator thresholds calibrated on the clean stored items. Neither the representation nor the radii are fitted to corrupted views. These results are reported separately from the fitted experiments of Section 3.3.

The recognition rule in Eq. (1) was selected instead of the runs of Eq. (4) after examining seeds 0–4, on which the runs were at or below the static rule in all 36 conditions. The rule, protocol and hypotheses were then recorded in a document included in the supplementary material with its SHA-256 hash, before seeds 10–14 were run. Table 5 reports only seeds 10–14, using new random draws of the stored images, out-of-distribution images, and block and dropout masks; the rows mask is fixed. The test sets are finite, so images recur between the two seed sets: of the images drawn for seeds 10–14, 79 % of the DTD images, 21 % of the CIFAR-10 images, 10 % of the SVHN images and 2 % of the stored CIFAR-100 images were also drawn for seeds 0–4.

Learning sets and parameters. L1: CIFAR-100 training split, one image per class, $K = 100$. L2: Tiny-ImageNet training split, one image per class, $K = 200$, resized to 32×32 with a box filter. Pixels are divided by 255. The ERF uses c, the mean of the K exemplars, and θ, the median over the store of the leave-one-out angular nearest neighbour at c ($\theta = 52.2$–$58.3°$ and 55.8–$58.6°$ over seeds), and recognises a cue iff $\min_i \|x - m_i\|/R_i < 1$ with $R_i = \|m_i - c\| \sin\theta$. Each comparator's threshold is the median over the store of the leave-one-out nearest score of an exemplar against the other $K - 1$: the nearest Euclidean distance; the normalised distance $\min_j \|x - m_j\|/\|m_j - c\|$; the centred angle at c (whose threshold is θ); the uncentred cosine angle; the Mahalanobis distance to the nearest exemplar under the store covariance with a ridge of 0.05 times its mean eigenvalue; and kNN, the distance to the third nearest exemplar on L_2-normalised vectors. No partial cue and no out-of-distribution image enters any parameter or threshold. With $K \leq 200$ images in 3072 dimensions the store covariance has rank at most $K - 1$, so off the store's span the ridge sets the metric and the Mahalanobis rule at its store threshold rejects almost every cue, familiar or not; kNN uses the third neighbour because on normalised vectors the first orders cues as the cosine rule does.

Cues and hypotheses. Familiar cues are the K exemplars with a fraction 0.25 or 0.5 missing as bottom rows, a square block or random pixels, the missing part set to c. Out-of-distribution cues are 500 images of each of the SVHN, CIFAR-10 and DTD test splits (DTD centre-cropped and resized), masked identically. Hypothesis A: on L1 the ERF's balanced accuracy is at least that of every comparator in at least 15 of the 18 conditions (three masks, two fractions, three out-of-distribution datasets); hypothesis B states the same for L2.

Table 5: **Recognition in the geometric regime.** Results of the registered one-exemplar experiment; all methods use the original pixel representation.

			ERF			ND	CA	Best std.	
Mask	Miss.	OOD	BA	FR	NR	BA	BA	BA	rule
CIFAR-100, $K = 100$									
rows	25 %	SVHN	**0.709**	0.000	0.419	0.649	0.685	0.582	cos.
		CIFAR-10	**0.846**	0.000	0.693	0.781	0.803	0.655	cos.
		DTD	**0.786**	0.000	0.572	0.730	0.754	0.613	cos.
	50 %	SVHN	**0.706**	0.052	0.463	0.664	0.680	0.500	Mah.
		CIFAR-10	**0.790**	0.052	0.632	0.748	0.767	0.519	cos.
		DTD	**0.799**	0.052	0.650	0.755	0.775	0.514	cos.
block	25 %	SVHN	**0.687**	0.000	0.375	0.639	0.662	0.597	cos.
		CIFAR-10	**0.826**	0.000	0.653	0.763	0.784	0.651	dist.
		DTD	**0.789**	0.000	0.578	0.734	0.755	0.631	cos.
	50 %	SVHN	**0.711**	0.042	0.464	0.667	0.679	0.508	cos.
		CIFAR-10	**0.826**	0.042	0.694	0.770	0.798	0.511	cos.
		DTD	**0.829**	0.042	0.699	0.780	0.802	0.524	cos.
dropout	25 %	SVHN	**0.730**	0.000	0.460	0.668	0.703	0.613	cos.
		CIFAR-10	**0.891**	0.000	0.782	0.825	0.854	0.692	dist.
		DTD	**0.815**	0.000	0.630	0.756	0.783	0.633	cos.
	50 %	SVHN	**0.881**	0.000	0.761	0.775	0.844	0.500	Mah.
		CIFAR-10	**0.980**	0.000	0.960	0.926	0.969	0.513	dist.
		DTD	**0.930**	0.000	0.860	0.856	0.903	0.502	cos.
Tiny-ImageNet, $K = 200$									
rows	25 %	SVHN	**0.679**	0.000	0.359	0.599	0.646	0.540	cos.
		CIFAR-10	**0.801**	0.000	0.602	0.716	0.757	0.636	dist.
		DTD	**0.743**	0.000	0.485	0.669	0.708	0.581	dist.
	50 %	SVHN	**0.703**	0.027	0.433	0.631	0.674	0.500	Mah.
		CIFAR-10	**0.766**	0.027	0.558	0.694	0.737	0.521	dist.
		DTD	**0.774**	0.027	0.574	0.703	0.747	0.510	dist.
block	25 %	SVHN	**0.655**	0.000	0.310	0.592	0.631	0.560	cos.
		CIFAR-10	**0.788**	0.000	0.576	0.697	0.749	0.632	dist.
		DTD	**0.746**	0.000	0.492	0.669	0.713	0.587	dist.
	50 %	SVHN	**0.701**	0.014	0.416	0.626	0.669	0.500	Mah.
		CIFAR-10	**0.812**	0.014	0.637	0.707	0.781	0.522	dist.
		DTD	**0.813**	0.014	0.640	0.725	0.783	0.508	dist.
dropout	25 %	SVHN	**0.697**	0.000	0.395	0.617	0.668	0.568	cos.
		CIFAR-10	**0.860**	0.000	0.719	0.757	0.812	0.666	dist.
		DTD	**0.767**	0.000	0.534	0.687	0.734	0.601	dist.
	50 %	SVHN	**0.832**	0.000	0.665	0.714	0.796	0.500	Mah.
		CIFAR-10	**0.961**	0.000	0.922	0.876	0.947	0.518	dist.
		DTD	**0.882**	0.000	0.765	0.785	0.850	0.501	dist.

Notes. ERF results use the static rule in Eq. (1); dynamical runs are reported in Tables 7 and 8. BA: balanced accuracy; ND: normalised distance; CA: centred angle. ND and CA use leave-one-out thresholds. Best std. is the highest mean BA among four standard detectors in each condition (dist.: nearest distance; cos.: cosine; Mah.: Mahalanobis). Missing pixels are set to c. Results are means over 5 seeds, with K familiar and 500 out-of-distribution cues per run. The ERF has the highest BA in every condition. Where Mahalanobis is listed, it rejects almost every cue and the other standard detectors are below chance.

Results. Both registered hypotheses hold: the ERF matches or exceeds every comparator in all 18 conditions on each learning set. In every single-seed run it is above the best comparator (smallest margin 0.007). Hypothesis L checks Eq. (9) cue by cue: the verdict of Eq. (1) on a familiar cue equals $q > \cos^2\theta$ for a fraction 0.998 of the cues (at least 0.970 in every run), and every exception is a cue with $q \leq \cos^2\theta$ that lies in another exemplar's support. The following outcomes were registered for reporting alongside the hypotheses. The readout $\hat{\imath}$ returns the source exemplar for a fraction 0.982 (L1) and 0.988 (L2) of the recognised familiar cues, at least 0.887 in every run; and the runs of Eq. (4), which also recognise cues carried into a support along a cone (Eq. (2)), reach 0.796 and 0.762 in balanced accuracy on average, against 0.807 and 0.777 for Eq. (1). A verification notebook recomputes the hypotheses, every threshold and one condition end to end from the stored images.

Post-hoc comparisons. The following comparisons were computed after the registered run, using exactly the same cues (Table 6). Nearest-neighbour data description (NNDD; Tax, 2001) accepts a cue x iff $\|x - m_j\| \leq \|m_j - m_{j'}\|$, with m_j the exemplar nearest to x and $m_{j'}$ the exemplar nearest to m_j; its threshold, 1, and its radii come from the store alone. Ranked without a threshold, the ERF's score $\min_i d_i/R_i$ has an AUROC at or above that of every standard detector and the centred angle in all 180 runs, and at or above NNDD's in 152; where NNDD ranks higher, it does so by at most 0.04. A nearest-neighbour distance over the kept pixels alone, which reads the mask off the cue, reaches an AUROC of 1.000 in every run. This comparator uses the missingness mask in addition to the cue, whereas the reported whole-vector rules receive the corrupted vector alone.

Table 6: **Additional comparisons on the registered cues.** Thresholded decisions and score rankings are reported separately.

(a) Thresholded decisions

Store	Missing	ERF BA	NNDD BA	NNDD FR	ERF wins
CIFAR-100	25%	0.787	0.855	0.025	0 / 9
	50%	0.828	0.675	0.279	9 / 9
Tiny-ImageNet	25%	0.748	0.858	0.017	0 / 9
	50%	0.805	0.717	0.248	9 / 9

(b) Score ranking (AUROC)

Store	Missing	ERF	NNDD	CA	Best std.
CIFAR-100	25%	0.987	0.979	0.980	0.872
	50%	0.904	0.749	0.880	0.540
Tiny-ImageNet	25%	0.989	0.985	0.983	0.885
	50%	0.923	0.796	0.896	0.519

Notes. These post-hoc analyses use the registered run's cues and the static ERF rule. Panel (a) gives balanced accuracy (BA), NNDD's false-rejection rate (FR), and the number of conditions in which ERF mean BA is at least NNDD's. Panel (b) gives AUROC for the ERF score, NNDD ratio, centred angle (CA) and best standard detector. Means cover three masks, three out-of-distribution datasets and 5 seeds. NNDD leads at 25 % missingness; ERF leads at 50 %. The fitted NNDD variant in Table 1 additionally uses a fitted metric and corrupted views; its calibration is given in Section C.

B.3 Runs in three representation spaces

The three spaces are MNIST pixels, ResNet-18 features of MNIST and ImageNet prototypes. Distances outside a memory's own units are reported in units of $\mathrm{nn}_{\mathrm{med}}$, the median nearest-neighbour distance of the learning sample; this reporting scale is not used by the field. A familiar cue is a stored vector displaced by $0.3\,\mathrm{nn}_{\mathrm{med}}$ in a random direction in the feature spaces, or a stored digit with 10–30 % of its rows set to zero or Gaussian noise of 0.08–0.22 times the image's standard deviation in pixel space. Novel cues come from classes absent from the store: MNIST digits 6–9 or 200 held-out ImageNet prototypes.

A run recognises a cue when its final state lies in a support and rejects it otherwise (Section 2.6). Table 7 reports rates, endings and median steps. Every tested run agrees with the static rule; no cue in this pool is carried into a support from outside.

Table 7: **Run outcomes on the three representation spaces.** Every measured run agrees with the static decision rule.

	pixels	ResNet	prototypes
θ (degrees)	48.2	40.4	63.6
$R_i/\mathrm{nn}_{\mathrm{med}}$, min–max	0.72–1.27	0.72–1.16	0.77–0.99
FR	0.110 (0.08–0.18)	0.000	0.000
NR	1.000	1.000	1.000
Balanced accuracy (run = Eq. (1))	0.945 (0.91–0.96)	1.000	1.000
Identity accuracy, recognised cues	1.000	1.000	1.000
Familiar endings: fixed / U_c / budget	267 / 33 / 0	300 / 0 / 0	300 / 0 / 0
Novel endings: fixed / U_c / budget	0 / 300 / 0	0 / 300 / 0	0 / 300 / 0
median steps, familiar / novel	5 / 12	6 / 10	6 / 9

Notes. Runs of Eq. (4) are compared cue by cue with Eq. (1): 6 seeds, $K = 6$, and 50 familiar plus 50 novel cues per run. Rates are means over seeds, with min–max ranges when values vary; endings and median steps pool the 300 familiar and 300 novel cues in each space. Radii are in units of $\mathrm{nn}_{\mathrm{med}}$. No cue is carried into a support, and all 36 stored vectors in each space are recognised when presented as their own cues.

B.4 Retrieval of corrupted images

Figure 5 shows the cues of Section 3.2 and Table 8 splits them by family. Each cue also passes through one step of modern Hopfield retrieval. Its inverse temperature β is the largest value in a grid of 11 values, where reconstruction error on 50 held-out familiar cues has saturated. The returned digit is the stored vector nearest the first step x_1; this step is itself a stored vector for 1412 cues and a weighted mixture of memories for 28 cues.

Every recognised ERF cue in this experiment converges numerically to the fixed point of Eq. (7), $s_1\sigma_i = 0.1024\,R_i$ outward of its memory. Its recovery, $1 - \|x_{t_{\mathrm{end}}} - m\|/\|x_0 - m\|$, is therefore $1 - 0.1024\,R_i/d_i(x_0)$ (mean 0.845), and the removed rows of 114 of the 135 recognised occluded cues are filled in. The classical Hopfield network on binarised patterns, run on the same cues, reaches a spurious state on every one (Hamming distance ≥ 55 from every stored pattern).

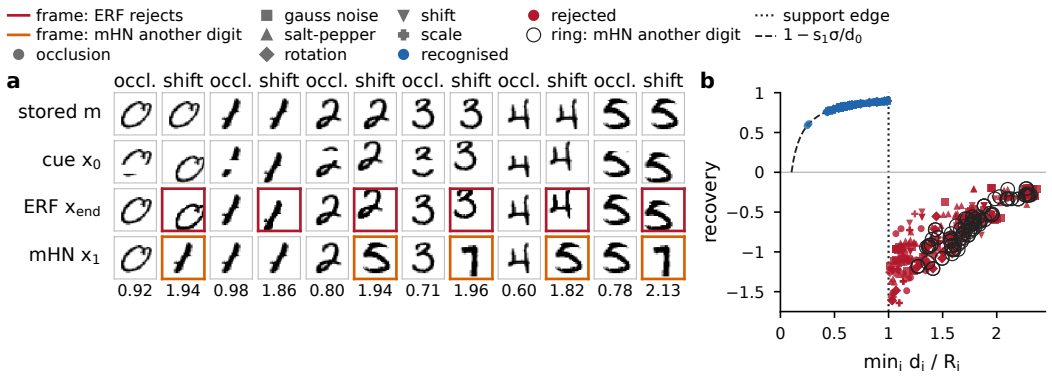


Figure 5: **Retrieval and reconstruction of corrupted MNIST cues.** Ten MNIST digits stored, one per class; $\theta = 48.2°$, $\kappa = 1/10$. **(a)** Digits 0–5 (seed 0), each with a mild cue, the first occlusion of 20 % of the rows that starts inside a support, and a heavy cue, the first shift by 4 pixels that starts outside every support: stored pattern m, cue x_0, the run's end state and modern Hopfield's first step x_1; numbers give $d(x_0, m)/R$. Red frame: rejected; orange frame: x_1 is another digit. **(b)** Recovery against $\min_i d_i(x_0)/R_i$ for the 472 cues of seed 0 with $d_0 > 0$; marker: family; colour: verdict; ring: modern Hopfield returns another digit; dotted: support edge; dashed: $1 - s_1\sigma_i/d_i(x_0)$ from Eq. (7), which every recognised cue follows.

Table 8: **Retrieval and reconstruction by corruption family.** Outcomes of the dynamical runs and the two Hopfield comparators.

	Median distance		ERF		Retrieval errors	
Family	$d_0/\mathrm{nn}_{\mathrm{med}}$	$\min_i d_i/R_i$	Inside support	Recovery	mHN	Classical
occlusion	0.87	0.95	135	0.840	2 (1 / 1)	214
Gaussian noise	1.29	1.34	70	0.825	41 (38 / 3)	214
salt-and-pepper	1.24	1.35	68	0.873	26 (24 / 2)	216
rotation	1.05	1.11	104	0.845	96 (77 / 19)	210
shift	1.54	1.69	10	0.887	170 (169 / 1)	206
scale	0.67	0.75	196	0.845	26 (8 / 18)	216
all six	1.05	1.15	583	0.845	361 (317 / 44)	1276

Notes. Ten MNIST digits, one per class, are stored. Results pool 3 seeds, 20 cues per family and level, and 4 levels (240 cues per family). The first distance column measures distance to the source memory in units of $\mathrm{nn}_{\mathrm{med}}$; the second gives the minimum distance normalised by each support radius. Recovery, $1 - \|x_{t_{\mathrm{end}}} - m\|/\|x_0 - m\|$, is averaged over recognised cues with $d_0 > 0$. Every run from inside a support ends at its memory's fixed point; every run from outside reaches U_c. mHN errors are cues whose nearest memory after modern Hopfield's first step is another digit, split by ERF verdict (rejected / recognised). Classical errors count convergence nearest to another stored pattern.

B.5 Overlapping supports

Figure 6 gives the rates, the agreement with Eq. (1) and the endings of Section 3.2 as the store grows. The schedule $\tau_t = 1/\max(1, W_t)$ turns an update among overlapping wells into a kernel-weighted mean of the active memories plus a drift. In this experiment, every run ends at a numerical fixed point or by entry into U_c within its budget. Acceptance remains stable under this schedule, but retrieval degrades as the wells merge: at $K = 1000$ a recognised ResNet cue ends nearest its source memory for a fraction 0.74 of the cues (0.96 for prototypes at $K = 800$), and a memory given as its own cue returns to itself for a fraction 0.68.

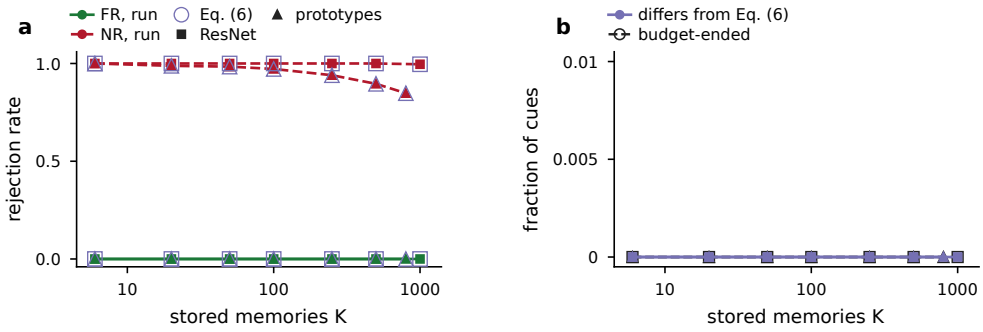


Figure 6: **Acceptance and termination as the store grows.** Means of 5 seeds: ResNet-18 features of MNIST (squares, $K = 6/20/50/100/250/500/1000$, $\theta = 40.4°$) and ImageNet prototypes (triangles, $K = 6/20/50/100/250/500/800$, $\theta = 63.6°$), 50 familiar and 50 novel cues per run, 7000 cues in all. **(a)** FR (solid) and NR (dashed) of the runs; rings: Eq. (1). **(b)** Fraction of cues whose run verdict differs from Eq. (1) (solid) and fraction ending on the budget (dashed): none at any K. Every one of the 3406 escaped cues reaches U_c at the exit step of theorem 2, and $\max_t W_t$ reaches 26.4.

B.6 Capacity and tolerance

Throughout, $K \geq 2$, $v_i = m_i - c \neq 0$, $\nu_i = \|v_i\|$, $u_i = v_i/\nu_i$, $\alpha_{ij} = \angle(v_i, v_j) \in [0, \pi]$ and $\alpha_{\min} = \min_{i \neq j} \alpha_{ij}$, where $\angle(a, b) = \arccos\big(\langle a, b\rangle / \|a\|\|b\|\big)$ for $a, b \neq 0$ is the angular distance on the sphere and satisfies the triangle inequality. Memory i receives a half-angle $\theta_i \in (0, \pi/2)$ and the radius $R_i = \nu_i \sin\theta_i$, and σ_i, A_i and λ follow from the R_i by Eq. (6). The derivation of Eq. (4) in Section A uses only $\sigma_i = R_i/\rho^*$ and $A_i = R_i/R_{\text{ref}}$, so the pull of every well at its support edge equals the push of the slope for every choice of radii. Outside every support the length scale is $\ell_t = \|x_t - c\| \sin\bar{\theta}$ for a fixed $\bar{\theta} \in (0, \pi/2)$; with a common half-angle $\theta_i = \theta$ this is Eq. (4) with $\bar{\theta} = \theta$. The proof of theorem 1 (A1) uses only $S(x_t) = \{i\}$ and $\ell_t = R_i$, and the proof of theorem 2 only the update outside every support, so both hold for every choice of radii, the second with $\bar{\theta}$ in place of θ. Well i is *isolated* if $\|m_i - m_j\| \geq R_i + R_j$ for every $j \neq i$, the hypothesis of theorem 1. Two open balls are disjoint iff the distance between their centres is at least the sum of their radii, so well i is isolated iff its support meets no other support. Memory i has the *tolerance* t if every cue with $\|x_0 - m_i\|^2 < t\,\nu_i^2$ lies in $B(m_i, R_i)$ and in no other support; for a stored image whose missing part is set to c this is the loss of any fraction less than t of its squared deviation from c (Eq. (9)).

B.6.1 Separation and the guarantee it gives

Lemma 5 (a support lies in its cone)**.** *Let $a \neq 0$ and $\|x - a\| < \|a\|$. Then $x \neq 0$, $\angle(x, a) < \pi/2$ and $\sin\angle(x, a) \leq \|x - a\|/\|a\|$. In particular, for $\theta \in (0, \pi/2)$ every $x \in B(m_i, \nu_i \sin\theta)$ satisfies $x \neq c$ and $\angle(x - c, v_i) < \theta$.*

Proof of theorem 5. $x = 0$ would give $\|x - a\| = \|a\|$, so $x \neq 0$; let $\varphi = \angle(x, a)$ and $s = \|x\|$. Then $\|x - a\|^2 = s^2 - 2s\|a\|\cos\varphi + \|a\|^2$. If $\cos\varphi \leq 0$ this is at least $\|a\|^2$, which is excluded, so $\varphi < \pi/2$; then the minimum over $s \geq 0$ is attained at $s = \|a\|\cos\varphi$ and equals $\|a\|^2 \sin^2\varphi$, so $\|x - a\| \geq \|a\| \sin\varphi$. For the second claim take $a = v_i$ and $x - c$ in place of x: $\|(x - c) - v_i\| = \|x - m_i\| < \nu_i \sin\theta < \nu_i$, so $\sin\angle(x - c, v_i) < \sin\theta$ with both angles in $[0, \pi/2)$, and $\angle(x - c, v_i) < \theta$. □

Lemma 6 (separation)**.** *Let $i \neq j$. (i) With a common half-angle $\theta \in (0, \pi/2)$,*

$$\|m_i - m_j\|^2 - (R_i + R_j)^2 = (\nu_i + \nu_j)^2 \cos^2\theta - 4\nu_i\nu_j \cos^2(\alpha_{ij}/2), \tag{11}$$

so the supports of m_i and m_j are disjoint iff $\cos\theta \geq g_{ij}\cos(\alpha_{ij}/2)$, where $g_{ij} = 2\sqrt{\nu_i\nu_j}/(\nu_i + \nu_j) \in (0, 1]$. In particular $\alpha_{ij} \geq 2\theta$ makes them disjoint for all $\nu_i, \nu_j > 0$, and if $\nu_i = \nu_j$ they

are disjoint iff $\alpha_{ij} \geq 2\theta$. (ii) With half-angles $\theta_i, \theta_j \in (0, \pi/2)$, $\alpha_{ij} \geq \theta_i + \theta_j$ makes the supports disjoint; so does $\alpha_{ij} \geq 2\max(\theta_i, \theta_j)$. If $\nu_i \cos\theta_i = \nu_j \cos\theta_j$, they are disjoint iff $\alpha_{ij} \geq \theta_i + \theta_j$.

Proof of theorem 6.

(i) Common half-angle. $\|m_i - m_j\|^2 = \|v_i - v_j\|^2 = \nu_i^2 + \nu_j^2 - 2\nu_i\nu_j \cos\alpha_{ij}$ and $(R_i + R_j)^2 = (\nu_i + \nu_j)^2(1 - \cos^2\theta)$, so the difference is $(\nu_i + \nu_j)^2 \cos^2\theta - 2\nu_i\nu_j(1 + \cos\alpha_{ij})$, which is Eq. (11) because $1 + \cos\alpha = 2\cos^2(\alpha/2)$. Since $\cos\theta > 0$ and $\cos(\alpha_{ij}/2) \geq 0$, the right side is nonnegative iff $\cos\theta \geq g_{ij}\cos(\alpha_{ij}/2)$. By the inequality of arithmetic and geometric means $g_{ij} \leq 1$, with equality iff $\nu_i = \nu_j$. If $\alpha_{ij} \geq 2\theta$, then $\alpha_{ij}/2 \in [\theta, \pi/2]$ and $g_{ij}\cos(\alpha_{ij}/2) \leq \cos(\alpha_{ij}/2) \leq \cos\theta$. If $\nu_i = \nu_j$, then $g_{ij} = 1$ and $\cos\theta \geq \cos(\alpha_{ij}/2)$ iff $\alpha_{ij}/2 \geq \theta$, both angles lying in $[0, \pi/2]$.

(ii) Unequal half-angles. Expanding as in (i),

$$\begin{aligned}\|m_i - m_j\|^2 - (R_i + R_j)^2 &= \nu_i^2\cos^2\theta_i + \nu_j^2\cos^2\theta_j \\ &\quad - 2\nu_i\nu_j\big(\cos\alpha_{ij} + \sin\theta_i\sin\theta_j\big) \\ &\geq 2\nu_i\nu_j\big(\cos(\theta_i + \theta_j) - \cos\alpha_{ij}\big),\end{aligned}$$

by $\nu_i^2\cos^2\theta_i + \nu_j^2\cos^2\theta_j \geq 2\nu_i\nu_j\cos\theta_i\cos\theta_j$, with equality iff $\nu_i\cos\theta_i = \nu_j\cos\theta_j$. As $\theta_i + \theta_j \in (0, \pi)$ and cos is strictly decreasing on $[0, \pi]$, $\cos(\theta_i + \theta_j) \geq \cos\alpha_{ij}$ iff $\alpha_{ij} \geq \theta_i + \theta_j$; and $2\max(\theta_i, \theta_j) \geq \theta_i + \theta_j$. Theorem 5 gives the sufficiency in (ii) a second way: the support of m_i lies in the open cone $\{x : \angle(x - c, v_i) < \theta_i\}$, and by the triangle inequality for angles the cones of m_i and m_j are disjoint when $\alpha_{ij} \geq \theta_i + \theta_j$. □

Theorem 7 (guarantee under separation)**.** *Let $\alpha_{\min} > 0$ and let the half-angle be common, $\theta_i = \theta \in (0, \pi/2)$. If $\theta \leq \alpha_{\min}/2$, every pair of supports is disjoint and every well is isolated, so theorem 1 applies to every memory, and (a) and (b) hold; (c) holds for every θ; (e) shows that for equal norms the condition $\theta \leq \alpha_{\min}/2$ cannot be weakened. Under Eqs.* (4) *and* (5) *and the protocol Eq.* (8)*:*

(a) Perturbed cues. *Every cue with $\|x_0 - m_i\| < \nu_i \sin\theta$ lies in $B(m_i, R_i)$ and in no other support. Equation* (1) *recognises it and retrieves i; the run stays in $B(m_i, R_i)$, ends by exit (a) or (c) of Eq.* (8) *inside it, and retrieves i.*

(b) Partial cues. *Let P be an orthogonal projection, $x_0 = c + Pv_i$ and $q = \|Pv_i\|^2/\nu_i^2$; for a coordinate mask P, x_0 is the stored image m_i with its missing part set to c. Then $\|x_0 - m_i\|^2 = (1 - q)\nu_i^2$, so $x_0 \in B(m_i, R_i)$ iff $q > \cos^2\theta$ (Eq.* (9)*), and every such cue is as in (a). If $0 < q \leq \cos^2\theta$ and $\arccos\sqrt{q} < \alpha_{\min} - \theta$, then x_0 is as in (c).*

(c) Cues outside every cone. *Every cue $x_0 \neq c$ with $\angle(x_0 - c, v_j) \geq \theta$ for every j lies in no support, and neither does any point of its ray from c. Equation* (1) *rejects it, the run rejects it at every budget T, and for $\rho_0 = \|x_0 - c\| \leq U$ the state reaches U_c at step $t_U = \lfloor \ln(U/\rho_0)/\ln(1 + \kappa\sin\theta)\rfloor + 1$, where the run ends by exit (b) if $T \geq t_U$.*

(d) Maximal tolerance. *If $\alpha_{\min} < \pi$, the half-angle $\theta = \alpha_{\min}/2$ is admissible, and (a) and (b) hold for every memory with the tolerance*

$$t^* = \sin^2(\alpha_{\min}/2) = \tfrac{1}{2}(1 - \cos\alpha_{\min}) :$$

the ERF retrieves m_i from every cue whose squared distance from m_i is less than $t^*\nu_i^2$, in particular from every partial cue that keeps more than the fraction $1 - t^*$ of m_i's squared deviation from c.

(e) Tightness. *Let $\nu_i = \nu_j$ for a pair with $\alpha_{ij} = \alpha_{\min}$, and let $\theta \in (\alpha_{\min}/2, \pi/2)$. Then $B(m_i, R_i)$ and $B(m_j, R_j)$ intersect, so neither well is isolated; some cue with $\|x_0 - m_i\| < R_i$ is recognised by Eq.* (1) *and not retrieved as i; and for some orthogonal projection P of rank $d - 1$ the partial cue $c + Pv_i$ keeps $q > \cos^2\theta$ and is recognised by Eq.* (1) *and not retrieved as i.*

Proof of theorem 7. Let $\theta \leq \alpha_{\min}/2$. For $i \neq j$, $\alpha_{ij} \geq \alpha_{\min} \geq 2\theta$, so the supports of m_i and m_j are disjoint by theorem 6(i), and every well is isolated.

(a) Perturbed cues. $x_0 \in B(m_i, R_i)$, and $x_0 \notin B(m_j, R_j)$ for $j \neq i$ because the supports are disjoint. Hence $d_i(x_0)/R_i < 1 \leq d_j(x_0)/R_j$ for every $j \neq i$: Eq. (1) recognises x_0 and retrieves

i. By theorem 1 every state of the run lies in $B(m_i, R_i)$; there $\|x_t - c\| < \nu_i + R_i \le U$, so exit (b) never occurs, the run ends by (a) or (c) at a state in $B(m_i, R_i)$ and in no other support, and it retrieves $\arg\min_j d_j(x_{t_{\text{end}}})/R_j = i$.

(b) Partial cues. P is symmetric and idempotent, so

$$\begin{aligned}\|x_0 - m_i\|^2 &= \|(I-P)v_i\|^2 \\ &= \nu_i^2 - \|Pv_i\|^2 = (1-q)\nu_i^2,\end{aligned}$$

which is less than $R_i^2 = \nu_i^2 \sin^2\theta$ iff $q > \cos^2\theta$. For the second claim, $q > 0$ gives $x_0 \neq c$, and $\langle Pv_i, v_i\rangle = \|Pv_i\|^2$ gives $\cos\angle(x_0 - c, v_i) = \sqrt{q}$, so $\angle(x_0 - c, v_i) = \arccos\sqrt{q} \ge \theta$ because $\sqrt{q} \le \cos\theta$. For $j \neq i$ the triangle inequality gives

$$\begin{aligned}\angle(x_0 - c, v_j) &\ge \alpha_{ij} - \arccos\sqrt{q} \\ &\ge \alpha_{\min} - \arccos\sqrt{q} > \theta.\end{aligned}$$

(c) Cues outside every cone. This part uses neither the separation nor theorem 1. Every point of the ray $\{c + s(x_0 - c) : s > 0\}$ makes the angle $\angle(x_0 - c, v_j) \ge \theta$ with every v_j, and by theorem 5 every point of $B(m_j, R_j)$ makes an angle less than θ with v_j; so no point of the ray lies in a support, and Eq. (1) rejects x_0. If $x_s - c = (1 + \kappa\sin\theta)^s(x_0 - c)$ for all $s \le t$, then $S(x_s) = \emptyset$ for $s \le t$ and theorem 2(i) gives $x_{t+1} - c = (1 + \kappa\sin\theta)^{t+1}(x_0 - c)$; by induction every state lies on the ray and in no support. The run therefore never ends by exit (a), its final state lies in no support, and it rejects x_0 at every budget. $\|x_t - c\| = (1 + \kappa\sin\theta)^t\rho_0$ first exceeds U at $t = t_U$.

(d) Maximal tolerance. $\theta = \alpha_{\min}/2 < \pi/2$, and $\sin^2\theta = (1 - \cos 2\theta)/2$. A cue with $\|x_0 - m_i\|^2 < t^*\nu_i^2$ satisfies (a); a partial cue with $q > 1 - t^* = \cos^2\theta$ satisfies (b).

(e) Tightness for equal norms. Write $\nu = \nu_i = \nu_j$, $R = R_i = R_j$ and $\alpha = \alpha_{ij} \in (0, \pi)$; $\alpha < \pi$ because $\alpha/2 < \theta < \pi/2$. By theorem 6(i) with $\nu_i = \nu_j$, $D = \|m_i - m_j\| < 2R$, and $D > 0$. For $s \in (D/2, R)$ the cue $x_0 = m_i + s(m_j - m_i)/D$ has $d_i(x_0) = s < R$ and $d_j(x_0) = |D - s| < s$, so $d_j(x_0)/R_j < d_i(x_0)/R_i < 1$: Eq. (1) recognises x_0 and $\hat{\imath}(x_0) \neq i$. For the partial cue, let $e_1 = u_i$ and let e_2 be the unit vector in the plane of u_i, u_j orthogonal to e_1 with $u_j = \cos\alpha\, e_1 + \sin\alpha\, e_2$. Choose $\varphi \in (\alpha/2, \theta)$, let $w = -\sin\varphi\, e_1 + \cos\varphi\, e_2$ and $P = I - ww^\top$, of rank $d-1$. Then $Pv_i = \nu(e_1 + \sin\varphi\, w) = \nu\cos\varphi\,(\cos\varphi\, e_1 + \sin\varphi\, e_2)$, so $q = \cos^2\varphi > \cos^2\theta$, $d_i(x_0) = \nu\sin\varphi < R$ and

$$d_j(x_0)^2 - d_i(x_0)^2 = 2\nu^2\cos\varphi\,\big(\cos\varphi - \cos(\alpha - \varphi)\big) < 0,$$

because $0 < \alpha < 2\varphi$ gives $|\alpha - \varphi| < \varphi < \pi/2$. Hence $d_j(x_0)/R_j < d_i(x_0)/R_i < 1$ and $\hat{\imath}(x_0) \neq i$. □

The guarantee of theorem 7(a) is Euclidean in each memory's own units: the ERF retrieves m_i from every cue within $\sqrt{t^*}\,\nu_i$ of it, the missing part of a partial cue and additive noise alike. Part (e) shows that for equal norms no common half-angle above $\alpha_{\min}/2$ keeps this guarantee, whether the cues are perturbed or partial; for unequal norms and $K \ge 3$ theorem 6(i) gives the largest common half-angle that keeps every well isolated in closed form, $\min_{i<j}\arccos\big(g_{ij}\cos(\alpha_{ij}/2)\big) \ge \alpha_{\min}/2$.

Corollary 8 (per-memory half-angles)**.** *Let $K \ge 3$, let the directions u_i be pairwise distinct, and set $\theta_i = \frac{1}{2}\min_{j\neq i}\alpha_{ij}$. Then $\theta_i \in (0, \pi/2)$, $\theta_i \ge \alpha_{\min}/2$, and for every such store every pair of supports is disjoint and every well is isolated. Every cue with $\|x_0 - m_i\| < \nu_i\sin\theta_i$, and every partial cue $c + Pv_i$ with $q > \cos^2\theta_i$, lies in $B(m_i, R_i)$ and in no other support: Eq.* (1) *retrieves i and the run stays in $B(m_i, R_i)$ and retrieves i. Every cue $x_0 \neq c$ with $\angle(x_0 - c, v_j) \ge \theta_j$ for every j is rejected by Eq.* (1) *and by the run, and its distance from c grows by the factor $1 + \kappa\sin\bar{\theta}$ per step until it reaches U_c. Memory i tolerates the loss of every fraction less than*

$$t_i = \sin^2\theta_i = \tfrac{1}{2}\big(1 - \max_{j\neq i}\cos\alpha_{ij}\big) \;\ge\; t^*$$

of its squared deviation from c.

Proof of theorem 8. $\theta_i > 0$ because the directions are distinct. If $\min_{j\neq i}\alpha_{ij} = \pi$, then $u_j = -u_i$ for every $j \neq i$, and two of the $K-1 \geq 2$ memories $j \neq i$ share a direction, which is excluded; so $\theta_i < \pi/2$. For $i \neq j$, $\theta_i + \theta_j \leq \alpha_{ij}/2 + \alpha_{ij}/2 = \alpha_{ij}$, and theorem 6(ii) makes the supports disjoint. The remaining claims follow as parts (a)–(c) of the proof of theorem 7, with θ_i for memory i, theorem 5 applied to each support with its own half-angle, and theorem 2 with $\bar{\theta}$ in place of θ. □

The half-angles of theorem 8 are measured on the store, as θ is, and they make every well isolated on every store, including the one-exemplar stores of Section B.2, where no well is isolated at the common measured θ.

B.6.2 Capacity

For $d \geq 2$ let u be uniform on the unit sphere S^{d-1}, e a unit vector, and $p_d(\varepsilon) = \Pr(\langle u, e\rangle \geq \varepsilon)$, the normalised measure of a spherical cap.

Lemma 9 (spherical caps)**.** *For $d \geq 2$ and $\varepsilon \in [0, 1)$,*

$$\textit{(i)}\quad p_d(\varepsilon) \leq \tfrac{1}{2}(1-\varepsilon^2)^{(d-1)/2} \leq e^{-d\varepsilon^2/2},$$

$$\textit{(ii)}\quad p_d(\varepsilon) \geq \frac{(1-\varepsilon^2)^{(d-1)/2}}{\sqrt{2\pi d}}.$$

The bound $e^{-d\varepsilon^2/2}$ is Ball (1997, Lemma 2.2).

Proof of theorem 9. $\langle u, e\rangle$ has the density $c_d(1-s^2)^{(d-3)/2}$ on $(-1, 1)$, $c_d = \Gamma(\frac{d}{2})/(\sqrt{\pi}\,\Gamma(\frac{d-1}{2}))$ (its square is Beta$(\frac{1}{2}, \frac{d-1}{2})$), so $p_d(\varepsilon) = c_d\int_\varepsilon^1 (1-s^2)^{(d-3)/2}ds$. Since $\int_\varepsilon^1 s(1-s^2)^{(d-3)/2}ds = (1-\varepsilon^2)^{(d-1)/2}/(d-1)$ and $\varepsilon \leq s \leq 1$ on the range,

$$\varepsilon\, p_d(\varepsilon) \;\leq\; \frac{c_d\,(1-\varepsilon^2)^{(d-1)/2}}{d-1} \;\leq\; p_d(\varepsilon). \tag{12}$$

Γ is log-convex, so

$$\Gamma(x+\tfrac{1}{2}) \leq \Gamma(x)^{1/2}\Gamma(x+1)^{1/2} = \sqrt{x}\,\Gamma(x),$$

$$x\Gamma(x) = \Gamma(x+1) \leq \Gamma(x+\tfrac{1}{2})^{1/2}\Gamma(x+\tfrac{3}{2})^{1/2} = \sqrt{x+\tfrac{1}{2}}\,\Gamma(x+\tfrac{1}{2});$$

with $x = (d-1)/2$ these give $(d-1)/\sqrt{2\pi d} \leq c_d \leq \sqrt{(d-1)/2\pi}$. (ii) follows from the right inequality of Eq. (12) and the lower bound on c_d. For (i) let $h(\varepsilon) = p_d(\varepsilon)(1-\varepsilon^2)^{-(d-1)/2}$; $h(0) = \frac{1}{2}$ by symmetry and

$$h'(\varepsilon) = (1-\varepsilon^2)^{-(d+1)/2}\Big[(d-1)\,\varepsilon\, p_d(\varepsilon) - c_d(1-\varepsilon^2)^{(d-1)/2}\Big] \leq 0$$

by the left inequality of Eq. (12), so $h \leq \frac{1}{2}$ on $[0, 1)$. Last,

$$\tfrac{1}{2}(1-\varepsilon^2)^{(d-1)/2} \leq \tfrac{1}{2}e^{-(d-1)\varepsilon^2/2} = \tfrac{1}{2}e^{\varepsilon^2/2}e^{-d\varepsilon^2/2} < e^{-d\varepsilon^2/2}$$

because $\frac{1}{2}e^{1/2} < 1$. □

Theorem 9(i) and (ii) differ by the factor $\sqrt{\pi d/2}$, so the exponent $\frac{d-1}{2}\ln(1-\varepsilon^2)$ of the cap measure is exact; the weaker form $e^{-d\varepsilon^2/2}$ loses it for large ε.

The random store has directions $u_1, \ldots, u_K$ i.i.d. uniform on S^{d-1} and arbitrary norms $\nu_i > 0$. This is the case of memories drawn i.i.d. from any law that is invariant under rotations about c and puts no mass at c, such as $\mathcal{N}(c, s^2 I_d)$, with c the mean of that law. Theorem 12 treats $c = \bar{m}$, the mean of the store. For $t \in (0, \frac{1}{2})$ let $\theta_t = \arcsin\sqrt{t}$, the half-angle at which theorem 7 certifies the tolerance t.

Theorem 10 (capacity)**.** *Let $d \geq 2$, $K \geq 2$, and let the directions u_i be i.i.d. uniform on S^{d-1}, with arbitrary norms. For $\varepsilon \in [0, 1)$, $\delta \in (0, 1)$ and $t \in (0, \frac{1}{2})$:*

(i) The upper-tail probability satisfies

$$\begin{aligned}\Pr(\cos\alpha_{\min} \geq \varepsilon) &\leq \frac{K(K-1)}{4}(1-\varepsilon^2)^{(d-1)/2} \\ &\leq \frac{K(K-1)}{2}e^{-d\varepsilon^2/2}.\end{aligned}$$

Hence with probability at least $1 - \delta$,

$$\cos\alpha_{\min} < \varepsilon(K, d, \delta) = \sqrt{2\ln\big(K(K-1)/2\delta\big)/d},$$

and theorem 7 holds with $\theta = \alpha_{\min}/2$ and the tolerance $t^* > \frac{1}{2}\big(1 - \varepsilon(K, d, \delta)\big)$ for every memory.

(ii) The lower-tail probability satisfies

$$\Pr(\cos\alpha_{\min} < \varepsilon) \leq \frac{2\sqrt{2\pi d}}{K(K-1)\,(1-\varepsilon^2)^{(d-1)/2}}.$$

(iii) If $K \leq 2\sqrt{\delta}\,\big(4t(1-t)\big)^{-(d-1)/4}$, then with probability at least $1-\delta$, $\alpha_{\min} > 2\theta_t$: every well of the field with half-angle θ_t is isolated, and theorem 7(a)–(c) hold with the tolerance t for every memory. Conversely, if $\Pr(\alpha_{\min} \geq 2\theta_t) \geq 1 - \delta$, then

$$K < 1 + \big(2\sqrt{2\pi d}/(1-\delta)\big)^{1/2}\big(4t(1-t)\big)^{-(d-1)/4}.$$

For equal norms $\alpha_{\min} \geq 2\theta_t$ is the event that every well is isolated at θ_t (theorem 6(i)).

(iv) Let $x_0 = c + \rho w$ with $\rho > 0$ and w uniform on S^{d-1}, independent of the store. For every store, the probability over w that $\angle(x_0 - c, v_i) < \theta_t$ for some i is at most $Kp_d(\sqrt{1-t}) \leq \frac{K}{2}t^{(d-1)/2}$; on the complement Eq. (1) *and the run reject x_0 and it escapes (theorem 7(c)). At $K = 2\sqrt{\delta}\,(4t(1-t))^{-(d-1)/4}$ this bound equals $\sqrt{\delta}\,\big(t/4(1-t)\big)^{(d-1)/4}$.*

Proof of theorem 10.

(i) Upper tail. $\cos\alpha_{\min} = \max_{i<j}\langle u_i, u_j\rangle$. Given u_j, u_i is uniform, so $\Pr(\langle u_i, u_j\rangle \geq \varepsilon) = p_d(\varepsilon)$, and the union bound over the $K(K-1)/2$ pairs with theorem 9(i) gives the tail bound. At $\varepsilon = \varepsilon(K, d, \delta)$ the right side equals δ (for $\varepsilon(K, d, \delta) \geq 1$ the claim holds because $\cos\alpha_{\min} < 1$ almost surely). Almost surely $0 < \alpha_{\min} < \pi$, so $\theta = \alpha_{\min}/2$ is admissible in theorem 7(d), and $t^* = \frac{1}{2}(1 - \cos\alpha_{\min})$.

(ii) Lower tail. Let $I_{ij} = \mathbf{1}[\langle u_i, u_j\rangle \geq \varepsilon]$ for $i < j$, $N = \sum_{i<j} I_{ij}$ and $p = p_d(\varepsilon)$, so $\{\cos\alpha_{\min} < \varepsilon\} = \{N = 0\}$ and $\mathbb{E}N = \binom{K}{2}p$. The I_{ij} are pairwise independent: indicators of disjoint pairs are independent, and for $j \neq k$, conditionally on u_i the events $\{\langle u_i, u_j\rangle \geq \varepsilon\}$ and $\{\langle u_i, u_k\rangle \geq \varepsilon\}$ are independent with probability p each, so $\Pr(I_{ij} = I_{ik} = 1) = p^2$. Hence $\operatorname{Var} N = \binom{K}{2}p(1-p)$, and Chebyshev's inequality gives

$$\begin{aligned}\Pr(N = 0) &\leq \Pr(|N - \mathbb{E}N| \geq \mathbb{E}N) \\ &\leq \frac{\operatorname{Var} N}{(\mathbb{E}N)^2} \leq \frac{1}{\binom{K}{2}p},\end{aligned}$$

which theorem 9(ii) bounds as stated.

(iii) Capacity at fixed tolerance. $\alpha_{\min} \leq 2\theta_t$ iff $\cos\alpha_{\min} \geq \cos 2\theta_t = 1 - 2t$, and $1 - (1-2t)^2 = 4t(1-t)$. By (i), $\Pr(\alpha_{\min} \leq 2\theta_t) \leq \frac{K^2}{4}(4t(1-t))^{(d-1)/2} \leq \delta$. On the complement $\theta_t < \alpha_{\min}/2$, and theorem 7 applies with $\theta = \theta_t < \pi/4$ and the tolerance $\sin^2\theta_t = t$. Conversely, each $\langle u_i, u_j\rangle$ has a density, so $\Pr(\alpha_{\min} \geq 2\theta_t) = \Pr(\cos\alpha_{\min} < 1 - 2t)$, which (ii) bounds by

$2\sqrt{2\pi d}\,/\big(K(K-1)(4t(1-t))^{(d-1)/2}\big)$; if this is at least $1-\delta$, then $(K-1)^2 < K(K-1) \le 2\sqrt{2\pi d}\,(4t(1-t))^{-(d-1)/2}/(1-\delta)$.

(iv) Rejection of an independent cue. $\angle(w,u_i) < \theta_t$ iff $\langle w,u_i\rangle > \cos\theta_t = \sqrt{1-t}$. Given the store, each of these K events has probability at most $p_d(\sqrt{1-t}) \le \frac{1}{2}t^{(d-1)/2}$ by theorem 9(i); the union bound gives the claim, and theorem 7(c) applies on the complement. At the stated K,

$$\begin{aligned}\frac{K}{2}t^{(d-1)/2} &= \sqrt{\delta}\,(4t(1-t))^{-(d-1)/4}t^{(d-1)/2}\\ &= \sqrt{\delta}\,\big(t/4(1-t)\big)^{(d-1)/4}.\end{aligned}$$ □

Since $4t(1-t) = 1-(1-2t)^2 \le e^{-(1-2t)^2}$ and $2e^{-1/4} > \sqrt{2}$, the capacity of (iii) is at least $\sqrt{2\delta}\,e^{d(1-2t)^2/4}$, the form given by the bound $e^{-d\varepsilon^2/2}$.

Corollary 11 (the trade-off is sharp)**.** *In the setting of theorem 10 let $d\to\infty$ and $K = K(d)\to\infty$.*

(i) If $(\ln K)/d \to \lambda \in (0,\infty)$, then $\cos\alpha_{\min} \to \sqrt{1-e^{-4\lambda}}$ in probability, and the maximal tolerance t^* of theorem 7(d) converges in probability to $\tau(\lambda) = \frac{1}{2}\big(1-\sqrt{1-e^{-4\lambda}}\big)$.

(ii) If $(\ln K)/d \to 0$ and $(\ln d)/\ln K \to 0$, then $\cos\alpha_{\min}\big/\sqrt{4\ln K/d} \to 1$ in probability, and $t^* = \frac{1}{2} - (1+o_P(1))\sqrt{\ln K/d}$.

(iii) For fixed $t\in(0,\frac{1}{2})$ and $\delta\in(0,1)$ let $K_\delta(t)$ be the largest K with $\Pr(\alpha_{\min}\ge 2\theta_t)\ge 1-\delta$. Then

$$\frac{1}{d}\ln K_\delta(t) \to \Lambda(t) = \frac{1}{4}\ln\frac{1}{4t(1-t)},$$

and Λ and τ are inverse to each other.

Proof of theorem 11. Fix $\eta > 0$.

(i) Exponential growth. Let $s = \sqrt{1-e^{-4\lambda}}$. For $\varepsilon = s+\eta < 1$, $\ln(1-\varepsilon^2) < -4\lambda$, and by theorem 10(i)

$$\frac{1}{d}\ln\Pr(\cos\alpha_{\min}\ge\varepsilon) \le \frac{2\ln K}{d} + \frac{d-1}{2d}\ln(1-\varepsilon^2),$$

whose limit $2\lambda + \frac{1}{2}\ln(1-\varepsilon^2)$ is negative. For $\varepsilon = s-\eta > 0$, $\ln(1-\varepsilon^2) > -4\lambda$, and by theorem 10(ii) with $K(K-1)\ge K^2/2$,

$$\begin{aligned}\frac{1}{d}\ln\Pr(\cos\alpha_{\min}<\varepsilon) \le{}& \frac{\ln(4\sqrt{2\pi d})}{d} - \frac{d-1}{2d}\ln(1-\varepsilon^2)\\ &- \frac{2\ln K}{d},\end{aligned}$$

whose limit $-\frac{1}{2}\ln(1-\varepsilon^2) - 2\lambda$ is negative. Both probabilities tend to 0, and $t^* = \frac{1}{2}(1-\cos\alpha_{\min})$.

(ii) Subexponential growth. Let $\varepsilon_0 = \sqrt{4\ln K/d} \to 0$. For $\varepsilon = (1+\eta)\varepsilon_0 < 1$, theorem 10(i) and $1-\varepsilon^2 \le e^{-\varepsilon^2}$ give $\Pr(\cos\alpha_{\min}\ge\varepsilon) \le \frac{1}{4}K^{2-2(1+\eta)^2(d-1)/d} \to 0$. For $\varepsilon = (1-\eta)\varepsilon_0$, $-\ln(1-\varepsilon^2) \le \varepsilon^2/(1-\varepsilon^2)$ and theorem 10(ii) give

$$\begin{aligned}\ln\Pr(\cos\alpha_{\min}<\varepsilon) \le{}& \ln(4\sqrt{2\pi}) + \tfrac{1}{2}\ln d\\ &- 2\ln K\Big[1-\frac{(1-\eta)^2}{1-\varepsilon^2}\Big] \to -\infty,\end{aligned}$$

because $\varepsilon\to 0$ and $\ln d = o(\ln K)$. The expression for t^* follows.

(iii) Capacity at fixed confidence. For d large, $2\sqrt{\delta}\,(4t(1-t))^{-(d-1)/4} \ge 2$, so by theorem 10(iii)

$$\begin{aligned}\big\lfloor 2\sqrt{\delta}\,(4t(1-t))^{-(d-1)/4}\big\rfloor &\le K_\delta(t)\\ &< 1 + \big(2\sqrt{2\pi d}/(1-\delta)\big)^{1/2}(4t(1-t))^{-(d-1)/4};\end{aligned}$$

the logarithms of both bounds divided by d tend to $\Lambda(t)$. Last, $e^{-4\Lambda(t)} = 4t(1-t)$ and $1-4t(1-t) = (1-2t)^2$, so $\tau(\Lambda(t)) = \frac{1}{2}(1-(1-2t)) = t$. □

Theorem 11 is a sharp threshold: with $K = e^{\lambda d}$ memories, every well is isolated with probability tending to one at every tolerance $t < \tau(\lambda)$, and for equal norms the closest pair's supports intersect with probability tending to one at every $t > \tau(\lambda)$.

B.6.3 THE STORE'S OWN MEAN

For the CIFAR-100 and Tiny-ImageNet stores the reference point is the mean of the store, $c = \bar{m}$. The centred vectors $m_i - \bar{m}$ are not independent, and their pairwise correlation is $-1/(K-1)$. For an isotropic Gaussian store this centring lowers every pairwise cosine in distribution, and it moves every angle by at most an explicit amount.

Proposition 12 (the store's own mean)**.** *Let $m_1, \dots, m_K$ be i.i.d. $\mathcal{N}(\mu, s^2 I_d)$, $d \geq 2$, $K \geq 3$, write $g_k = (m_k - \mu)/s$, and let $\alpha'_{ij} = \angle(g_i, g_j)$ be the angles of the same store about μ, which follow the model of theorem 10.*

(i) With $c = \bar{m}$, for $i \neq j$ and every ε, $\Pr(\cos\alpha_{ij} \geq \varepsilon) \leq \Pr(\cos\alpha'_{ij} \geq \varepsilon) = p_d(\varepsilon)$. Hence theorem 10(i) and the first statement of theorem 10(iii) hold for $c = \bar{m}$ as stated, and so does theorem 10(iv) for a cue $x_0 \sim \mathcal{N}(\mu, s_0^2 I_d)$, $s_0 > 0$, independent of the store, with the probability taken over the store and the cue.

(ii) For any reference point c let $h = (c - \mu)/s$, so that $v_k = s(g_k - h)$. If $\|h\| < \min_k \|g_k\|$, then $|\alpha_{ij} - \alpha'_{ij}| \leq 2\beta$ for all $i \neq j$ and $|\cos\alpha_{\min} - \cos\alpha'_{\min}| \leq 2\beta$, where $\beta = \arcsin(\|h\|/\min_k \|g_k\|)$. For $c = \bar{m}$, and for c the mean of n further vectors of the same law drawn independently of the store, $n\|h\|^2$ is χ^2_d with $n = K$ respectively n, and for every $x \in (0,1)$ with $xn > 2$,

$$\Pr\left(\beta \leq \arcsin\sqrt{2/(xn)}\right) \;\geq\; 1 - (2/e)^{d/2} - K\big(xe^{1-x}\big)^{d/2}.$$

(iii) Parts (i) and (ii) of theorem 11 hold for $c = \bar{m}$.

Proof of theorem 12.

(i) Pairwise comparison. For each coordinate ℓ the pair $(g_i^\ell - \bar{g}^\ell,\, g_j^\ell - \bar{g}^\ell)$ is centred bivariate normal with variances $1 - 1/K$ and covariance $-1/K$, independently over ℓ. So (v_i, v_j) has the law of $s\sqrt{1-1/K}\,(x,\, \rho x + \sqrt{1-\rho^2}\, y)$ with $\rho = -1/(K-1) \in [-\frac{1}{2}, 0)$ and x, y i.i.d. $\mathcal{N}(0, I_d)$. Let $w = \rho x + \sqrt{1-\rho^2}\, y$, so y is a positive multiple of $w + |\rho| x$ ($x, y, w \neq 0$ almost surely). Writing $w = a\, x/\|x\| + b\, n$ with $n \perp x$ a unit vector and $b \geq 0$,

$$\cos\angle(x, w + \gamma x) = \frac{a + \gamma\|x\|}{\sqrt{(a + \gamma\|x\|)^2 + b^2}}$$

is nondecreasing in $\gamma \geq 0$, so $\cos\angle(x,y) \geq \cos\angle(x,w)$. As $x/\|x\|$ and $y/\|y\|$ are independent and uniform, $\Pr(\cos\angle(x,y) \geq \varepsilon) = p_d(\varepsilon)$, and the same holds for α'_{ij}. The union bound in the proof of theorem 10(i) uses only these pairwise probabilities, and almost surely $0 < \alpha_{\min} < \pi$, because at most two of $K \geq 3$ nonzero vectors are pairwise antipodal. For the cue, $x_0 - \bar{m} = s_0 h_0 - s\bar{g}$ with $h_0 \sim \mathcal{N}(0, I_d)$ independent of the store, and each coordinate of $x_0 - \bar{m}$ has covariance $-s^2/K + s^2/K = 0$ with the same coordinate of v_i; the two vectors are independent isotropic Gaussians, so $\Pr(\angle(x_0 - \bar{m}, v_i) < \theta_t) = p_d(\sqrt{1-t})$ for each i, and the union bound of theorem 10(iv) applies.

(ii) Perturbation by centring. By theorem 5 with $a = g_i$ and $x = g_i - h$, $\angle(g_i - h, g_i) \leq \arcsin(\|h\|/\|g_i\|) \leq \beta$, and the triangle inequality for angles gives $|\angle(g_i - h, g_j - h) - \angle(g_i, g_j)| \leq 2\beta$. The minimum over pairs moves by at most 2β, and $|\cos a - \cos b| \leq |a - b|$. For $c = \bar{m}$, $h = \bar{g} \sim \mathcal{N}(0, I_d/K)$; for the mean of n independent vectors, $h \sim \mathcal{N}(0, I_d/n)$. For $X \sim \chi^2_d$ the moment generating function $\mathbb{E}e^{\gamma X} = (1 - 2\gamma)^{-d/2}$, optimised over γ, gives $\Pr(X \geq 2d) \leq (2/e)^{d/2}$ and $\Pr(X \leq xd) \leq (xe^{1-x})^{d/2}$ for $x \in (0,1)$. With probability at least $1 - (2/e)^{d/2} - K(xe^{1-x})^{d/2}$ both $\|h\|^2 < 2d/n$ and $\min_k \|g_k\|^2 > xd$ hold, and then $\|h\|^2/\min_k \|g_k\|^2 < 2/(xn) < 1$.

(iii) Asymptotic rates. By (ii) with $n = K$ it suffices that $\beta \to 0$ in probability in case (i) of theorem 11 and $\beta/\sqrt{\ln K/d} \to 0$ in probability in case (ii). In case (i) take $x = e^{-2-4\lambda}$: then $xe^{1-x} \leq e^{-1-4\lambda}$, $K(xe^{1-x})^{d/2} \leq \exp\big(\ln K - (\frac{1}{2} + 2\lambda)d\big) \to 0$, and $\arcsin\sqrt{2/(xK)} \to 0$. In

case (ii) take $x = \frac{1}{2}$: $K(\frac{1}{2}e^{1/2})^{d/2} = \exp\big(\ln K - \frac{d}{2}(\ln 2 - \frac{1}{2})\big) \to 0$ because $\ln K = o(d)$, and $\arcsin(2/\sqrt{K}) \le \pi/\sqrt{K}$, so $\beta/\sqrt{4\ln K/d} \le \frac{\pi}{2}\sqrt{d/(K\ln K)} \to 0$, because $\ln d = o(\ln K)$ gives $d \le \sqrt{K}$ for large K. □

For $c = \bar{m}$, theorem 12(ii) transfers the lower bound of theorem 10(ii) with the correction 2β: $\cos\alpha_{\min} \ge \cos\alpha'_{\min} - 2\beta$, where $\beta \le \arcsin\sqrt{2/(xK)}$ with the stated probability. The correction is of order $K^{-1/2}$ and is small against $\sqrt{\ln K/d}$ when d is small against $K\ln K$. For a reference point that is the mean of n further vectors drawn independently of the store, as for the stores whose learning sample is drawn separately (Section 2), the centred vectors have pairwise correlation $1/(n+1) > 0$ and theorem 12(i) does not apply; on the event of theorem 12(ii), every well at the half-angle θ_t is isolated whenever $\alpha'_{\min} > 2\theta_t + 2\arcsin\sqrt{2/(xn)}$.

C Experiments with fitted radii

C.1 Fitting and evaluation protocol

Design and registration. The confirmation experiment uses three stores under two encoders, giving six store–encoder settings. Each setting has twenty corruption conditions: six families at two levels, with three mask shapes for each of the two missing-pixel families. A reported condition specifies the family, level, mask shape where applicable, and distractor set. CIFAR-100 has three distractor sets and each DISC21 store has one, giving $2 \times (20 \times 3 + 20 + 20) = 200$ conditions. The rule, fitting procedure, stores, encoders, corruption families and accuracy hypotheses were recorded before the run in `preregistration_round14.md`. The separation threshold 0.9 was selected after the results were seen; its diagnostic use on this experiment is exploratory. The subsequent fixed-threshold validation is reported in Section D.

Stores and representations. The encoders are frozen DINOv2 ViT-S/14 ($d = 384$) and ResNet-50 ($d = 2048$). Stores contain one CIFAR-100 training image per class ($K = 100$), or $K = 1000$ or 5000 distinct DISC21 reference images. The 5 seeds 60–64 select the smaller stores; seeds 60–62 select the $K = 5000$ stores. The fill value c is the mean of the stored images in image space. The centred-angle comparator uses the mean of the encoded stored items, a different point.

Fitting views and optimisation. Each stored image is masked at 10, 20, 30, 40, 50 or 60 %, using bottom rows, a square block or random pixels, with the missing part set to c. Gaussian noise has standard deviation 0.05, 0.1 or 0.2. All corruptions act before encoding. The map $W \in \mathbb{R}^{128\times d}$ is fitted by InfoNCE, with a view as query, the K stored images as candidates and the source memory as target. Optimisation uses temperature 0.07, Adam at learning rate 10^{-3}, 600 steps and minibatches of 4096 views. The projected vectors are normalised and the radii follow Eq. (10).

Evaluation cues and rates. Evaluation uses masks at 25 and 50 %, filled with c or zero; Gaussian noise at standard deviation 0.1 or 0.2; shifts of 2 or 4 pixels; rotations of 10 or 20 degrees; and Gaussian blur at standard deviation 1 or 2. Zero fill, shift, rotation and blur are absent from fitting. Distractors are 500 held-out references for each DISC21 store, and 500 images from each of SVHN, CIFAR-10 and DTD for CIFAR-100. They receive the same corruptions and encoding as familiar cues. Fitting and evaluation use separate random streams; row masks, shifts, rotations and blur are fixed given their parameters. No distractor or evaluation cue enters any fitted parameter or threshold. FR, NR and balanced accuracy follow Section 3. Identity accuracy uses all familiar cues, including rejected ones.

C.2 Comparators and calibration

Comparators in the fitted space share W with the ERF; the fitted NNDD variant also uses the same views. Other comparators use the frozen encoder's original space. Calibration uses stored items and, where specified, their fitting views.

Global distance, cosine and angle thresholds. Nearest distance uses one radius: the median leave-one-out nearest-neighbour distance among stored items, computed separately in the original and fitted spaces. Modern Hopfield in the original space uses maximum cosine similarity, with the median leave-one-out maximum cosine as its threshold. Centred angle uses the median leave-one-out nearest angle about the encoded store mean. In the normalised fitted space, cosine and Euclidean distance give the same identity ordering; the thresholded global-distance rule represents this fitted-space comparison.

NNDD variants. Let δ_i be the distance from stored item i to its nearest other stored item in the relevant space. The NNDD comparator in the original encoder space accepts when the minimum distance-to-δ_i ratio over memories is at most one. The variant labelled *NNDD, fitted* uses the same acceptance construction in the fitted space, with radius $\min(a_i, \delta_i)$ and a strict ratio threshold of one. Its own-view quantile a_i is the same as the ERF's; its cap uses the nearest stored item instead of the ERF's nearest other-item view. These local-radius rules do not use a global leave-one-out median threshold. Their acceptance construction is specified here separately from the nearest-exemplar NNDD rule in Section B.2.

C.3 Radius and identity readout

For $0 < a_i < b_i$, Eq. (10) gives $a_i < r_i < b_i$. Thus every sampled own view at distance at most a_i lies inside the support, and every sampled view of another item lies outside it. The 0.9-quantile does not determine whether the remaining own views lie inside or outside r_i. The minimum defining b_i concerns only sampled views, so neither statement guarantees behaviour on unseen cues. Separation, the fraction with $a_i < b_i$, ranges from 0.13 to one across the confirmation settings.

These fitted supports overlap. Separation of fitting views is therefore distinct from pairwise disjointness of supports, and the isolation guarantees in theorems 1 and 7 do not apply. The fitted experiments evaluate acceptance and identity directly. Since z is normalised, minimum distance and maximum cosine select the same stored item. The identity columns of Tables 1 and 9 accordingly compare the fitted readout with modern Hopfield in the original encoder space; giving it W makes the identity readouts coincide.

C.4 Corruption families and radius ablations

Table 9 groups the confirmation results by corruption family, restricted to settings with separation above 0.9. The ERF exceeds the strongest mean comparator in all six families in this subset, including zero fill, shift, rotation and blur, which are absent from fitting. This grouping does not describe the low-separation settings.

The per-memory radius ablation uses the earlier CIFAR-100/DINOv2 setting, with $20 \times 3 = 60$ evaluation conditions. Both alternatives use the same fitted metric and masking views; the global radius is the 0.9-quantile of the pooled distances from views to their source memories. Per-memory radii match or exceed it in 60 conditions, with a mean gain of 0.084. A separate earlier comparison on CIFAR-100 and Tiny-ImageNet pixels gives 0.660 for geometric radii without metric fitting and 0.902 for the full fitted method on the same cues. That comparison changes both the metric and the radius construction, so it does not isolate the effect of per-memory radii.

C.5 Fitting tolerance

The tolerance experiment varies the maximum fitting-mask fraction τ on DISC21 with $K = 5000$ under ResNet-50, using two seeds and fixed evaluation conditions. The metric and radii are refitted for each τ; the Gaussian-noise fitting views remain part of the family. The median own-view distance a_i rises from 0.835 to 0.900, while the median nearest-other-view distance b_i falls. Separation decreases monotonically from 0.820 at $\tau = 0.2$ to 0.366 at $\tau = 0.6$ (Table 10).

Detection does not decrease monotonically with separation: it peaks at $\tau = 0.4$ with balanced accuracy 0.734 and reaches 0.700 at the widest tolerance, while identity accuracy rises throughout. Widening the fitting family changes both the metric and the radii, so this experiment does not isolate a causal effect of separation on accuracy. It shows a trade-off within the tested fitting procedure, rather than an application of the isolated-support capacity theorem.

Table 9: **Recognition by corruption family on high-separation stores.**

Corruption of the cue	ERF	Best comparator		Cells	Identity	
	BA	BA	rule	won	ERF	mHN
missing, filled with c	**0.972**	0.848	NNDD, fitted	32/42	0.999	0.693
missing, filled with zero	**0.791**	0.686	centred angle	33/42	0.781	0.607
shift	**0.992**	0.961	NNDD, fitted	14/14	1.000	0.999
rotation	**0.970**	0.902	NNDD, fitted	14/14	0.996	0.980
blur	**0.760**	0.689	centred angle	10/14	0.796	0.599
Gaussian noise	**0.749**	0.582	centred angle	13/14	0.902	0.186

Notes. Only confirmation conditions with separation above 0.9 are included. BA is balanced accuracy. The best comparator is selected by its mean BA within each family; Cells won counts conditions where the ERF matches or exceeds every comparator after averaging over seeds. Identity uses all familiar cues: ERF uses the fitted metric and mHN the original encoder space. Four families are absent from fitting.

Table 10: **Recognition as the fitting tolerance changes.**

Declared tolerance τ	0.2	0.3	0.4	0.5	0.6
Separation of the store	0.820	0.744	0.646	0.516	0.366
Median own-view distance a_i	0.835	0.844	0.861	0.880	0.900
Median nearest other view b_i	0.924	0.907	0.897	0.886	0.876
Detection (balanced accuracy)	0.715	0.731	0.734	0.730	0.700
Identity	0.668	0.707	0.729	0.743	0.743

Notes. DISC21 with $K = 5000$, ResNet-50, two seeds and 20 fixed evaluation conditions. The fitting masks are capped at fraction τ. Wider fitting tolerance enlarges the median radius and reduces separation in this experiment.

D Validation of the separation diagnostic

D.1 Threshold selection and validation design

The threshold 0.9 was selected from the confirmation results, so its performance on those settings is exploratory. It was then fixed in `preregistration_round16.md`, before evaluation on five further store sources. The registered prediction is that the ERF matches or exceeds every comparator if and only if the store's separation exceeds this threshold. No threshold is refitted during validation.

The stores are Tiny-ImageNet ($K = 200$, one image per class), DTD ($K = 500$), CIFAR-10, SVHN and 3D Shapes ($K = 1000$ each), under DINOv2 and ResNet-18. ResNet-18 was absent from the fitted confirmation runs. Each of the 10 settings uses seeds 70 and 71 and twenty corruption conditions, giving 200 conditions after averaging over seeds. Distractors are 500 held-out images from the same source. The fitting procedure, radii, comparators and calibration follow the confirmation protocol in Section C.

D.2 Aggregation of comparator performance

For each condition, we first average every method's balanced accuracy over seeds and then subtract the largest comparator accuracy from the ERF accuracy. Figure 3a shows these condition-level leads; panel b and Table 11 average them within each store–encoder setting. The strongest comparator may therefore differ between conditions.

In contrast, Table 1 selects the comparator with the largest mean accuracy over a setting, and Table 9 selects it over a corruption family in the high-separation subset. A mean of condition-wise maxima is at least as large as a maximum of method means. Thus a small lead over the comparator in Table 1 can coexist with a negative mean diagnostic lead, as for DINOv2 on DISC21 with $K = 1000$. All diagnostic results in Section 3.4 and this appendix use the condition-wise reference.

D.3 Registered hypotheses and results

Two of the three registered targets fail. H16a requires correct prediction in at least 0.75 of the 200 conditions; the observed fraction is 0.68. H16b requires the ERF to match or exceed every comparator in at least 0.75 of the conditions above the threshold; the observed fraction is 0.67.

H16c requires the sign of the mean condition-wise lead to agree with the threshold in at least eight settings. It agrees in 10 of 10 settings (Table 11). Mean leads are +0.047 above the threshold and -0.034 below it. Individual corruption conditions can change the sign within a setting. These results support the diagnostic at the setting level for the tested stores and encoders, while failing the two condition-level prediction targets.

Table 11: **Validation of the fixed separation threshold.**

Store	Sep.	ERF	Best rival	Lead	Prediction
DINOv2					
Tiny-ImageNet, $K = 200$	1.00	0.903	0.830	+0.073	lead, correct
DTD, $K = 500$	0.98	0.856	0.831	+0.025	lead, correct
CIFAR-10, $K = 1000$	0.62	0.824	0.833	-0.009	no lead, correct
SVHN, $K = 1000$	0.35	0.755	0.787	-0.032	no lead, correct
3D Shapes, $K = 1000$	0.00	0.599	0.620	-0.021	no lead, correct
ResNet-18					
Tiny-ImageNet, $K = 200$	1.00	0.824	0.780	+0.043	lead, correct
DTD, $K = 500$	0.84	0.763	0.765	-0.003	no lead, correct
CIFAR-10, $K = 1000$	0.25	0.688	0.726	-0.037	no lead, correct
SVHN, $K = 1000$	0.02	0.521	0.649	-0.128	no lead, correct
3D Shapes, $K = 1000$	0.00	0.603	0.608	-0.006	no lead, correct

Notes. ERF and Best rival report balanced accuracy. Best rival is the mean of the strongest comparator accuracy in each condition; Lead is the mean condition-wise difference. Values average twenty corruption conditions after averaging over two seeds (five sources, two encoders, 200 conditions). These store sources were not used to select the threshold.

E Related work

Associative memory, rejection and local basins. Classical, dense and modern Hopfield networks retrieve stored patterns through energy-based updates (Hopfield, 1982; Krotov & Hopfield, 2016; Ramsauer et al., 2021), but their standard mechanisms do not specify when an arbitrary cue should be rejected. Rejection can be incorporated into the dynamics by adding a trivial attractor for out-of-distribution inputs (Moriai et al., 2025). Locality has separately been studied through finite attraction basins in kernel memory networks (Iatropoulos et al., 2022) and compactly supported interactions in Epanechnikov energy (Hoover et al., 2025). ERF combines these two directions: each memory receives a compact well with its own radius, while one shared outward drift acts when local support is absent. Thus a supported cue can be retained and an unsupported cue can escape under the same update, rather than competing with an additional rejection attractor. These guarantees are conditional: retention requires an isolated support, and geometric escape requires a ray that intersects no support.

Score-based rejection and calibration. Score-based detectors compare confidence, energy, distance or similarity with an acceptance threshold (Hendrycks & Gimpel, 2017; Liu et al., 2020b; Lee et al., 2018; Ren et al., 2021; Sun et al., 2022; Techapanurak et al., 2021; Yang et al., 2024). Thresholds are often calibrated to a target acceptance rate on familiar validation inputs, and auxiliary outliers may be used during training (Hofmann et al., 2024). Our setting excludes separate familiar calibration images and distractors from adaptation. In the geometric regime, comparator thresholds are leave-one-out statistics of the store; in the fitted regime, comparators using the adapted metric receive the same representation and corrupted views (Appendices B.2 and C). The comparisons therefore isolate the choice of acceptance region rather than access to additional calibration data.

Local acceptance regions around exemplars. RCE networks place hyperspheres around prototypes and leave inputs outside them unclassified (Reilly et al., 1982); nearest-neighbour data description sets a local acceptance scale from neighbour distances (Tax, 2001); and the extreme value machine fits inclusion probabilities around individual examples (Rudd et al., 2018). Open-set recognition likewise studies the risk of assigning known labels far from training data (Scheirer et al., 2013; Mendes Júnior et al., 2017). ERF adopts local regions for instance memory but connects them to an associative field. Its geometric radii give a closed-form missing-information tolerance and capacity relation, while its fitted radii combine the spread of each memory's own views with the proximity of other memories' views. Comparisons with NNDD and a shared-radius control hold the fitted metric and views fixed, isolating how the regions are chosen. Unlike few-shot classification (Vinyals et al., 2016; Snell et al., 2017; Liu et al., 2020a), the task is to recognise corrupted copies of stored instances, not new instances of a stored class.