Evolutionary dynamics of visual memory

Jordan W. Suchow[1], Benjamin Allen[2,3], Martin A. Nowak[3], George A. Alvarez[1]

[1]Department of Psychology, Harvard University
[2]Program for Evolutionary Dynamics, Harvard University
[3]Department of Mathematics, Emmanuel College

Abstract. Visual memory holds in mind details of objects, textures, faces, and scenes. After initial exposure to an image, however, visual memories rapidly degrade because they are transferred from iconic memory, a high-capacity sensory buffer, to working memory, a low-capacity maintenance system. Here, we extend the classic depiction of visual memory maintenance to include competitive interactions between memories and a stability threshold that determines the weakest maintainable memory. The proposed model, based on these principles, can be understood as an evolutionary process with memories competing over a limited memory-supporting commodity. The model reproduces the time course of visual working memory observed through experiment. Notable features of this time course include load-dependent stability and overreaching, in which the act of trying to remember more information causes people to forget faster, and to remember less, respectively. Our results demonstrate that evolutionary models provide quantitative insights into the mechanisms of memory maintenance.

Memories degrade and are eventually forgotten. From its inception, research on memory degradation has characterized 'forgetting functions' that track the downfall of how much is remembered over time[1,2]. A forgetting function is shaped by the processes that degrade and maintain memories, and its functional form is a signature of the underlying mechanisms[3]. Examining forgetting functions can thus reveal important insights, having previously provided some of the primary evidence for iconic memory[4] and for rational theories of adaptive forgetting[5,6].

In the case of visual memory, which holds in mind the details of what was seen, at least three subsystems contribute to storage and maintenance. Each subsystem has a characteristic timescale, format, and neural substrate. Iconic memory, a high-capacity sensory buffer, operates over short time scales (0.05–1 s) and is thought to result from persistence of activation in mid- to high-level visual areas such as the lateral occipital complex and temporal cortex[4,7–8]. Visual working memory, an active maintenance system, operates over moderate time scales (0.5–20 s) and is supported by a network that includes prefrontal cortex, basal ganglia, and parietal cortex[9–12]. Visual long term memory, a high-capacity passive store, operates over lengthy time scales (minutes to decades) and recruits much of the same machinery, such as the hippocampus, that supports more general forms of long term memory[13]. Other subsystems have been proposed, each with its own particular properties and substrates[14–16]. Together, these systems maintain visual memories, allowing us to remember what we see.

The classic forgetting function of visual memory, which is applicable to short and moderate time scales, has a brief period of rapid decline followed by a long plateau, a form that is attributed to the quick fading of iconic memory and the stability of working memory[4]. This model has survived for over 50 years with only slight modification[17,18]. Here, we extend the classic model to account for new data, leveraging the tools of evolutionary biology to model memories as entities that compete for a limited mental commodity that is shared among them.

In a series of experiments, we asked participants to remember a set of objects, and then after a short delay, to report the color of a randomly selected object using a graded continuous-report procedure. We tested a five-hundredfold range of durations (0.03–16 s) and a twelvefold range of loads (1–12 objects), randomly interleaving them all.

Results

Participants' errors on the memory task were used to derive forgetting curves that track the number of remembered objects as it falls over time (Fig. 1A). Fitting an exponential function separately to the data from each memory load, we found that the rate of forgetting depends on the total amount of information held in mind, with lone memories lasting longest (estimated mean lifetime of 157 s) and higher loads leading to progressively shorter lifetimes (Fig. 1B–D). The relationship between memory load and mean lifetime is well described by a power law with exponent −1.7 ($r^2 = -0.98$, $p = 6.5 \times 10^{-8}$), such that halving the load leads to roughly a tripling in mean lifetime (Fig. 1B). This relationship was also found when limiting analysis to durations greater than 1 s, where iconic memory plays no role[4,18] (Fig. 1C). In the initial analysis, we assumed that the forgetting function is exponential-like. To test whether load-dependent stability is robust to this assumption, we also considered another functional form—a power law. Power law forgetting has been observed in long-term memory[2] and is common because it can arise both normatively[5,6] (i.e., as the optimal solution to a task) and as an artifact of averaging exponential-like forgetting functions that differ in timescale. We found a comparable effect of load-dependent stability under power law forgetting (Fig. 1D).

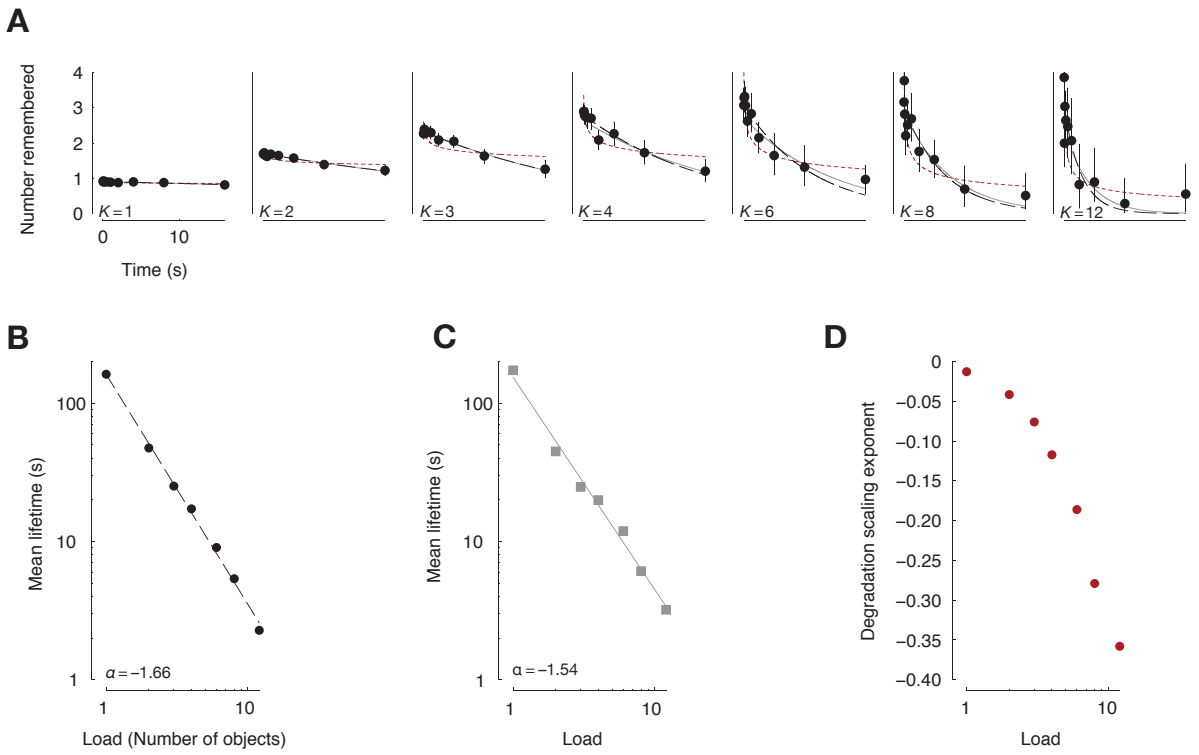

Fig. 1. Load-dependent forgetting: the more you try to remember, the faster you forget. (A) Subplots show the empirical forgetting function for each load (K = 1, 2, 3, 4, 6, 8, or 12 objects), tracking the number of remembered objects as it falls over time. The dashed black and solid grey curves assume an exponential form to the forgetting function; the former is fit to all the data and the latter considers only durations of at least 1 s, where iconic memory plays no role. The dotted red curve assumes that forgetting follows a power law. Curves were fit to the data from each load separately. Error bars here and in other figures are 95% credible intervals. (B) Comparing lifetimes across loads, the relationship is well described by a power law with exponent −1.66. (C) This relationship also holds for the estimates derived from durations of 1s and beyond, where iconic memory plays no role. (D) A similar relationship is found for degradation under power law forgetting, quantified by the scaling exponent, which falls precipitously with load.

The lines of the forgetting functions for each load cross (Fig. 2A). At short durations, presenting a greater number of objects causes more to be remembered. At long durations, however, the opposite is often true: presenting a greater number of objects causes fewer to be remembered (Fig. 2B). Crossovers in the forgetting function imply that the relationship between the number of objects presented and the number remembered changes with time (Fig. 2C). The presence of crossovers suggests a flawed strategy of the participants, who presumably control how many objects they encode and maintain. Like a bodybuilder who herniates a disk by straining to lift too heavy a weight, our participants performed worse because they tried to encode and maintain more than they could handle—they *overreached*. A comparable effect has been reported for tracking many moving objects at once, which is a task that is demanding of attention[19]. Alternatively, it is possible that participants

chose appropriately when deciding how many objects to encode or maintain, but that the presence of distracting objects led to flawed execution of the chosen strategy. Crossovers are inconsistent with the classic model and its variants, whose lines occasionally meet, but never cross (see Supplementary Equations).

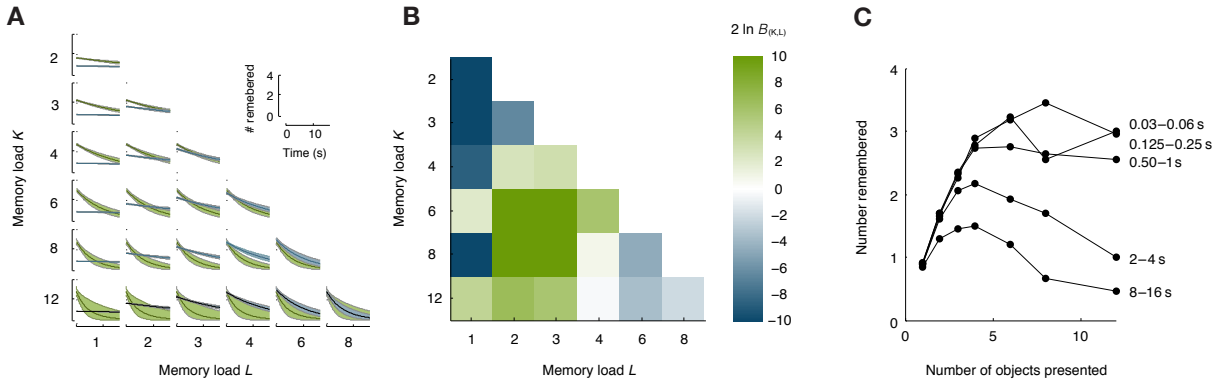

Fig. 2. Crossovers in the forgetting function and mnemonic overreaching. (A) Each subplot is a pairwise comparison of the forgetting functions for two memory loads, with the greater load *K* plotted in green and the lesser load *L* in blue. Shaded error bars show 95% credible intervals at each time point. (B) The strength of evidence for a crossover between the forgetting functions for a pair of loads *K* and *L* is expressed as twice the natural logarithm of the Bayes factor $B_{(K,L)}$ in favor of a model with crossover over one without it[20]. In the heat map, $B_{(K,L)}$ is coded as positive (green) when the evidence favors the crossover model and as negative (blue) when it favors a model without crossover. (C) The crossover effect implies that the relationship between the number of objects presented and the number remembered will change over time. The plateau at ≈ 3 objects for short durations is considered to be the signature of visual working memory's meager capacity. However, the non-monotonic curves seen for durations greater than 1 s are new and suggest a failure on the part of participants, who would have performed better by trying to encode less of the display.

Discussion

Construction of an evolutionary model

To explain these results, we propose a minimal account of visual memory rooted in evolutionary dynamics, a mathematical framework for describing how information is reproduced in a setting that is subject to mutation, selection, and random drift[21]. Specifically, we describe an evolutionary process operating over a commodity that supports memory. Within this framework, units of this memory commodity are assigned to items, and the strength and stability of a memory depends on the number of quanta assigned. This commodity may take any one of a number of forms, including, for example, cycles of a time-based refreshing process[22], distinct phases in phase-dependent coding mechanisms[23], or populations of neurons in prefrontal cortex representing "token" encodings of visual events[24]. Regardless of its particular form, what defines a commodity is being a limited asset, at least partially

shared across memories, whose availability affects performance. A shared commodity stands in contrast to a purely local substrate that represents specific stimulus attributes in particular locations of the visual field[25]. Though such location- and content-based substrates are essential for encoding information into working memory, they are perhaps less relevant to memory maintenance, which may operate over a pluripotent medium[24].

Recent work has sought to determine both the quantization[25] of the commodity and the structure of the memories that it forms (e.g., whether they form bound objects, bags of unbound features, or hierarchical bundles of features[13,26–27]). In the general case, the commodity is divided into N quanta, each of which is dedicated to some information about a mnemonic structure. Discrete "slot"-based models set $N \approx 4$, whereas "continuous resource" models consider the limit as N tends to infinity[25,28–30]. Both classes of model assume that the stability and quality of memory for an object increases as more of the commodity is allocated to it.

We model the evolution of the quantal population using a generalization of the Moran process. The Moran process is a model of evolution in finite populations that was originally used to describe the dynamics of allele frequencies[31], and which has recently been leveraged to describe evolutionary processes in diverse settings, including frequency-dependent selection, emergence of cooperative behavior, and cultural evolution of language[32–34]. The Moran process begins with a population of quanta (the units of the commodity) that have been assigned to structures (which may be objects, features, bundles, etc.). At each time step, a quantum becomes degraded, losing the information that it stores. In the same step, the lost information is replaced by the contents of another quantum, randomly selected from them all (Fig. 3). Our generalization further introduces a stability threshold: if at any point a structure has fewer than s quanta assigned to it, it becomes inaccessible to the maintenance process and the associated quanta lose their assignment, floating freely until they are reassigned (Fig. 3, grey dots). This threshold is comparable to a recently proposed lower bound on the fidelity of an accessible memory[35] and has the effect of limiting the number of structures that can be stored to approximately N/s. When the stability threshold is a single quantum, we can derive the forgetting function analytically (Supplementary Equations); for greater values of the stability threshold, the forgetting function is obtained numerically. Over time, the number of quanta assigned to a structure drifts. Eventually, either a single structure reaches fixation, with all the quanta assigned to it, or corruption prevails and all the quanta are left free-floating and unassigned.

Various cognitive processes could give rise to these dynamics. First consider a process of active maintenance that recycles the mnemonic commodity, repurposing quanta dedicated to lost memories in order to provide redundancy to those that remain. Alternatively, consider a process of interference where at each time step a quantum becomes corrupted, taking on the value and assignment of an

intruding quantum. In these ways, the evolutionary process can be seen as a formal model of memory maintenance in the face of degradation due to interference or decay.

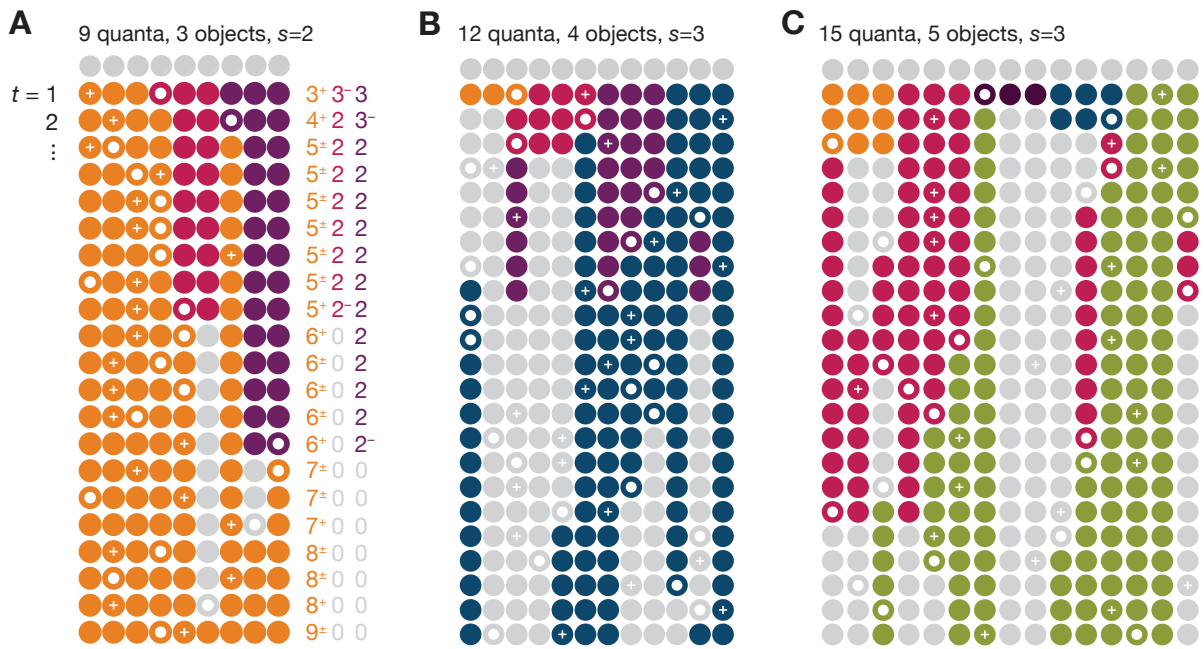

Fig. 3. Modeling the evolution of a mnemonic commodity. (A) In the top row, a pool of 9 unallocated quanta (grey dots) that wait to be assigned. In the second row, each quantum is assigned to one of three structures, labeled in orange, red, and purple. Subsequent rows show the processes as it plays out over time, one time step per row. An empty circle ○ denotes the quantum that died and a circle with a plus mark ⊕ denotes the quantum that was selected to replace it. When the number of quanta assigned to a structure drops below the threshold ($s = 2$ in panel A and $s = 3$ in panels B and C), the remaining quanta become inaccessible to the maintenance process and lose their assignment (greyed-out dots). Numbers to the right of the panel count the number of quanta dedicated to each structure at that time step, with superscripts showing which structure gains, +, loses, −, or both, ±. The number of stored structures corresponds to the number of unique colors in a row. In this run, the orange structure reaches fixation. (B) A second iteration of the process, with 12 quanta and 4 objects. At the last time step that is displayed, the blue structure is the only one left, but it has not yet reached fixation, with much of the commodity left unassigned. (C) A third iteration, with 15 quanta and 5 objects. Red takes an early lead, but is eventually overcome by green.

Each component of the evolutionary model — the commodity, the degradation process, and the stability threshold — contributes to the resulting dynamics. When a memory structure loses a quantum and hits the stability threshold, that structure is lost. This happens quickly at first, but more slowly over time, because the loss of one memory lends stability to those that remain. When there are many objects to remember, the mnemonic commodity is spread thinly, with fewer quanta per memory structure, and so each one stands closer to the stability threshold. In contrast, when there are fewer objects to remember, the representation of each one is more stable. This discrepancy accounts for the

relationship between lifetime and load and may also explain the remarkable stability of lone memories, which need not compete at all for the mnemonic commodity.

Evaluating the evolutionary model

The proposed evolutionary process reproduces the observed forgetting functions of visual memory, showing effects of load-dependent forgetting and mnemonic overreaching, effects that are inconsistent with the classic, pure death, and sudden death accounts, which show neither effect (Fig. 4). In the classic account (Fig. 4, grey dashed lines), only iconic memory degrades; the stability of working memory produces flat forgetting functions with no slope and which do not cross. In the pure death account (Fig. 4, blue dashed lines), working memory decays at a fixed rate that is independent of load; this produces sloped lines that share a common decay rate (mean lifetime) and never cross. The same is also true of the sudden death account (Fig. 4, yellow dashed lines), which extends the pure death account by proposing a 4-second window of time in which working memory is immune to degradation[17]. Only the proposed evolutionary process produces both effects (Fig. 4, green solid lines).

It is conceivable that the proposed process could be used to describe both iconic and working memory, together, as a single process. Iconic memory was initially considered to be a unitary system, but was later fractionated into two distinct subcomponents, one providing visible persistence (i.e., the experience of seeing a stimulus after its removal), the other providing informational persistence (i.e., remembering something about a stimulus after its removal[36]. Visible persistence is distinct in its phenomenology from working memory, as memories are rarely experienced as being seen, but informational persistence and working memory have long been conflated. For example, studies of visual working memory often test at durations of 500–1000 ms, a point in time at which there is a non-negligible contribution of iconic memory to task performance[18]. We find that the evolutionary model provides excellent fits to the forgetting functions of iconic memory that have been measured in previous experiments (Fig. 5). The evolutionary model was fit to data from Yang (1999) by minimizing the squared error between the data and the model's predictions using Nelder–Mead simplex search over the model's parameters[18,37]. Experimental evidence of a distinct iconic storage system underlying informational persistence comes from a variety of experiments, not all of which rely on its timing. However, the closeness of fit between model and data suggests that informational persistence in iconic memory may be the initial moments of maintenance in a lengthier short-term storage system.

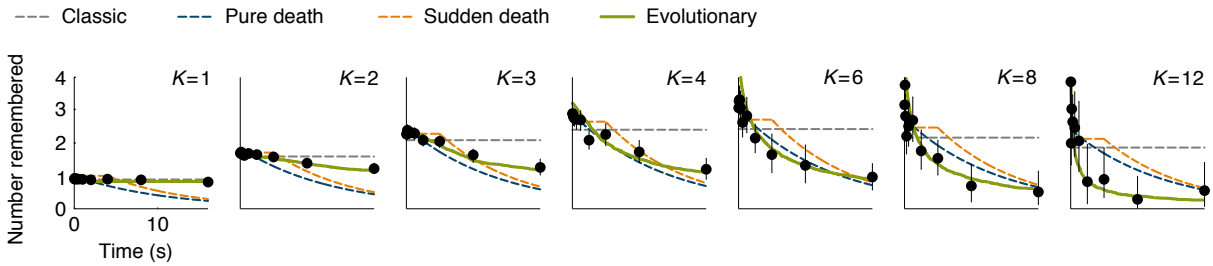

Fig. 4. Comparing the forgetting functions of the classic, pure death, sudden death, and evolutionary models. Subplots show the forgetting function for a particular load (1, 2, 3, 4, 6, 8, or 12 objects). Competing models fail to capture important aspects of the data. The classic model (grey dashed line) does not change over time. The pure death model (blue dashed line) has a fixed rate of forgetting, one that is too quick for low loads (K = 1, 2, and 3) and too slow for high loads (K = 8 and 12). The sudden death model behaves similarly to the pure death model. The evolutionary model succeeds, with slow forgetting at low loads and quick forgetting at high loads (best-fit parameters N = 58, p = 0.82, t_{step} = 0.01, and s = 7). The form of each forgetting function is derived in Supplementary Equations. We also considered the effect of individual differences on the predictions of each model (Figs. S1–6).

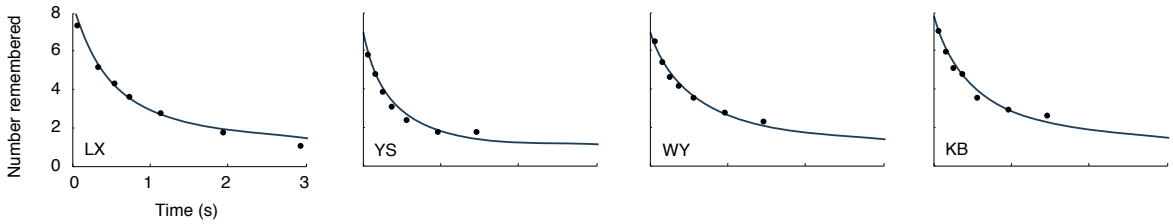

Fig. 5. The evolutionary model can be used fit the full time course of visual memory as a single process. Data are replotted from Yang (1999). Each subplot is data from the participant whose initials appear in the lower left corner of that subplot. The stability threshold was fixed at s = 1.

Extending the evolutionary model

Evolutionary dynamics provides a rich framework in which to extend our account of visual memory. For example, it is likely that the neural substrate over which visual memory maintenance operates is in some way structured—perhaps as a gridded visuotopic maps like those found in visual areas in the brain, or as a scale-free network, like so many other biological systems[38–40]. Evolutionary graph theory, which extends evolutionary dynamics to structured populations, is a natural tool for specifying the interaction network of the mnemonic commodity and exploring how such structure impacts the stability of memories[41]. Similarly, frequency-dependent fitness, where the success of an individual depends on the abundance of that individual's type, is analogous to a memory maintenance policy that selectively maintains memories according to their stability (e.g., by purifying, selectively

maintaining the strongest memories, or balancing, selectively maintaining those memories on the brink).

On a process model of forgetting

In the context of visual working memory, encoding and maintenance are often viewed as a process in which a limited store fills up during encoding and then remains mostly stable, perhaps with whole object representations being lost one by one over time. Importantly, in this view, encoding and maintenance happen independently over stored objects, resulting in exponential decay functions with the same rate for all memory loads. Load-dependent forgetting suggests an alternate view: visual memory representations compete for a commodity that is at least partially shared among them, such that the success of maintenance for one structure is affected by that for the others, thereby introducing a dependency of forgetting on load. Our proposed evolutionary model is the simplest instantiation of this principle, with a mental commodity fully shared across representations.

By constructing an evolutionary model of memory degradation that operates over the natural units of visual memory allocation and maintenance—those of a mnemonic commodity, rather than whole objects—we are able to build better process models of memory maintenance and its dynamics. Here, we focused on short-term visual memories. But just as the framework of evolutionary theory has been applied across many domains and scales, from alleles to words and from cells to societies, so too might our approach, when appropriately extended, be applied to memory maintenance in more complex systems, such as the transactional and collective memories of groups.

Methods

Participants

We recruited 1000 participants using Amazon Mechanical Turk, an online labor market where people perform short computer-based tasks for pay[42–45]. The number of participants was chosen before collecting the data. One thousand trials per condition is 5–10× the typical sample size in comparable studies; simulations suggest it is enough to provide accurate measurements even in cases of moderate to severe degradation of memories. Each participant was paid $0.50 for a few minutes of work. Recruitment and testing was executed in accordance with Harvard University regulations and approved by the Committee on the Use of Human Subjects in Research under the Institutional Review Board for the Faculty of Arts and Sciences.

Stimuli

The stimulus consisted of a set of 1–12 colorful dots. The dots were arranged in a ring around a small central fixation point. Each dot appeared in one of twelve locations spaced equally around the ring,

with the constraint that each dot had its own location. Dots were randomly assigned one of 180 equally spaced equiluminant colors drawn from a circle (radius 59°, center L=54, a=18 and b=–8) cut out from the CIE $L^*a^*b^*$ color space. Stimuli were rendered in a browser. The viewing distance was approximately 50 cm.

Procedure

A schematic diagram of the procedure appears in Fig. 6. The participant pressed the space bar to begin the trial. The stimulus immediately appeared for 250 ms and then disappeared. Participants were asked to remember the colors of the presented dots. The screen remained blank for the retention interval. Once the waiting period was over, a small cue appeared in the location of one of the dots, selected at random. The participant used the mouse to select the remembered color of the cued dot. Colors were selected by moving the mouse in a circle around the center of the display. A dot appeared at the center of the display and was continuously updated with the currently selected color. Participants registered their selection by clicking. No feedback was provided. There were ten possible retention intervals (1/32, 1/16, 1/8, 1/4, 1/2, 1, 2, 4, 8, or 16 s) and seven possible memory loads (1, 2, 3, 4, 6, 8, and 12 objects), for a total of 7 × 10 = 70 test trials. The order of the trials was randomized so that the participant would not know at the time of encoding for how long they would need to remember the objects. There were 6 practice trials, 1–6 objects in ascending order, all with a retention interval of 1 s. There were negligible practice effects during the test trials, suggesting that our training procedure was sufficient for participants to perform the task well (Fig. S9).

Fig 6. The memory task. Participants stare at a small cross at the center of the screen. A set of colorful dots briefly appears. After a delay, the location of one of the dots (selected at random) is marked with a cue. The participant is asked to report the color of the dot that appeared at the marked location. The error is the difference in color between the reported color and the true color. Here, the participant makes a big error, reporting the green object as orange.

Extracting the empirical forgetting functions

First we excluded participants who showed weak evidence of having faithfully completed the task. To do this, for each participant, we compared two models of performance using the Akaike information criterion. The first model was a two-parameter model[25] where with probability 1–g the participant remembers the stimulus with fixed fidelity σ, the dispersion parameter of a von Mises distribution (a circular analogue to the normal distribution), and with probability g guesses blindly. The other model

was a zero-parameter model where the participant always guessed blindly. Since our null model — complete guessing for all 76 trials — is so weak, our criterion for inclusion was strict, $AIC_C \geq 10$, which constitutes strong evidence of the presence of memory[46]. This strict criterion may inadvertently exclude participants with poor working memory, though the results we find are comparable when relaxing the inclusion criterion to $AIC_C \geq 3$, which constitutes moderate evidence of memory.

Next, we combined participants' data into a super-subject. The main manipulations of time and load were performed within each subject — one trial per condition per participant — but the analysis combined the data together. Though this is necessary to achieve sufficiently precise measurements, it leaves open the possibility that variability among people in the form of individual differences will affect the shape of the measured curves (see Supplementary Note). We fit a four-parameter variable-precision model[47,48] to arrive at an estimate of the guess rate g separately for each duration and load K. The product $(1–g)K$, the average number of remembered objects, is plotted in Figs. 2, 3, and 5. Analysis was performed using MemToolbox 1.0.0 [49].

Estimating mean lifetimes

Mean lifetimes were estimated by fitting an exponential decay model to the raw error data. The exponential decay model is a time-based generalization of the two-component model described in the previous section. In the exponential decay model, the number of remembered objects Y falls exponentially with time t, such that $Y(t) = \beta^{-t/\tau}$, where τ is the mean lifetime and β is the number of encoded objects at $t = 0$. Memory quality at each duration, as quantified by the dispersion parameter of the corresponding von Mises distribution, was allowed to vary freely. A loose prior was placed over each parameter for the purposes of estimation. The prior on β was uniform over the full range, 0 to the number of presented objects. The prior on the bias was uniform over the full range, $–\pi$ to π radians. The prior on τ was log-normal with a mean of 20 s and a standard deviation of 2 ln units. The prior on the dispersion parameter of each von Mises distribution was log-normal with a mean of 7.4 and a standard deviation of 1 ln unit. The model was fit with MCMC using PyMC version 2.2 [50].

Strength of evidence for crossover

For each possible pairing of tested set sizes, we measured the strength of evidence in favor of a model where the forgetting function for the greater set size crosses over that for lesser set size (i.e., where it starts higher and ends lower) to one where it does not cross over. Strength of evidence was measured using the Bayes factor, the ratio of the posterior odds to the prior odds. The prior odds were 1 : 1. The prior probabilities on model parameters were the same as in the previous section.

Declaration of competing financial interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Author contributions

J.W.S. developed the study concept. All authors contributed to the study design. Testing and data collection were performed by J.W.S. The form of the forgetting function was derived by B.A. All authors performed data analysis and interpretation. J.W.S. drafted the manuscript and all authors revised it. All authors approved the final version of the manuscript for submission.

Acknowledgments

The authors thank Tim Brady, Patrick Cavanagh, Michael Cohen, Justin Jungé, Dave Rand, Eva Suchow, Steven Suchow, and Yaoda Xu for helpful discussion, and Trinidad Zuluaga for help with data collection.

References

1. Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology*.
2. Wixted, J. T. & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science, 2*, 409−415.
3. Wixted, J. T. (1990). Analyzing the empirical course of forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 927–935.
4. Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*.
5. Anderson, J. R. (1989). A rational analysis of human memory. *Varieties of memory and consciousness: Essays in honour of Endel Tulving*, 195−210.
6. Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396−408.
7. Ferber, S., Humphrey, G. K. & Vilis, T. (2005). Segregation and persistence of form in the lateral occipital complex. *Neuropsychologia*, *43*, 41−51.
8. Keysers, C., Xiao, D.–K., Földiák, P. & Perrett, D. (2005). Out of sight but not out of mind: The neurophysiology of iconic memory in the superior temporal sulcus. *Cognitive Neuropsychology, 22*, 316−332.
9. Baddeley, A. (1992). Working memory. *Science, 255*, 556−559.

10. Todd, J. J. & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*, 751−754.

11. Voytek, B. & Knight, R. T. (2010). Prefrontal cortex and basal ganglia contributions to visual working memory. *Proceedings of the National Academy of Sciences, 107*, 18167−18172.

12. Xu, Y. & Chun, M. M. (2005). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*, 91−95.

13. Brady, T. F., Konkle, T. & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*, 1−34.

14. Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neurosciences, 23*, 247–251.

15. Sligte, I. G., Scholte, H. S. & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE*, *3*, e1699.

16. Wood, J. N. (2009). Distinct visual working memory systems for view-dependent and view-invariant representation. *PLoS ONE*, *4*, e6601.

17. Zhang, W. & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20*, 423−428.

18. Yang, W. (1999). Lifetime of human visual sensory memory: Properties and neural substrate. *University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-99-03.*

19. Holcombe, A. O. & Chen, W.-Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to < 3 Hz when tracking three. *Journal of Vision,* 13, 1–19.

20. Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association,* 90, 773−795.

21. Nowak, M. A. (2006). *Evolutionary dynamics: Exploring the equations of life*. Belknap Press of Harvard University Press.

22. Vergauwe, E., Barrouillet, P. & Camos, V. (2009). Visual and spatial working memory are not that dissociated after all: a time-based resource-sharing account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1012–1028.

23. Siegel, M., Warden, M. R., & Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences*, 106(50), 21341−21346.

24. Bowman, H. & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review, 114*, 38–70.

25. Zhang, W. & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. Nature, 453, 233−235.

26. Luck, S. J. & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279−281.

27. Fougnie, D. & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision, 11*, 1−12.

28. Bays, P. M. & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851−854.

29. Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87−114.

30. Wilken, P. & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*, 1120−1135.

31. Moran, P. A. P. (1958). Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society, 54*, 60−71.

32. Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature, 428*, 646−650.

33. Fudenberg, D., Nowak, M. A., Taylor, C. & Imhof, L. A. (2006). Evolutionary game dynamics in finite populations with strong selection and weak mutation. *Theoretical Population Biology, 70*, 352−363.

34. Komarova, N. L. & Nowak, M. A. (2003). Language dynamics in finite populations. *Journal of Theoretical Biology, 221*, 445−457.

35. Alvarez, G. A., Brady, T. F., Konkle, T. A., Gill, J. & Oliva, A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science, 24*, 981−990.

36. Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, *27*(3), 183–228.

37. Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization*, *9*(1), 112–147.

38. Franconeri, S. L., Alvarez, G. A. & Cavanagh, P. (2013). Flexible cognitive resources: competitive content maps for attention and memory. *Trends in Cognitive Sciences, 17*, 134−141.

39. Gardner, J. L., Merriam, E. P., Movshon, J. A. & Heeger, D. J. (2008). Maps of Visual Space in Human Occipital Cortex Are Retinotopic, Not Spatiotopic. *The Journal of Neuroscience,* 28, 3988–3999.

40. Schira, M. M., Tyler, C. W., Spehar, B. & Breakspear, M. (2010). Modeling magnification and anisotropy in the primate foveal confluence. *PLoS Computational Biology,* 6, e1000651.

41. Lieberman, E., Hauert, C. & Nowak, M. A. (2005). Evolutionary dynamics on graphs. *Nature,* 433, 312–316.

42. Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368.

43. Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5.

44. Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23.

45. Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 411–419.

46. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

47. Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229.

48. van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785.

49. Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), 1–8.

50. Patil, A., Huard, D., & Fonnesbeck, C. J. (2010). PyMC: Bayesian stochastic modeling in Python. *Journal of Statistical Software*, *35*(4), 1–80.

Supplemental Information for

Evolutionary dynamics of visual memory

Jordan W. Suchow, Benjamin Allen, Martin A. Nowak, and George A. Alvarez

Contents

1 Supplementary Equations: Deriving forgetting functions

In the following derivations, we suppose that the participant is asked to remember a set of K things (the memory load), stored as objects, features, or hierarchical bundles of features (hereafter, "mnemonic structures" or just "structures"). We further suppose that visual memory is limited and imperfect, such that only $Y \leq K$ of the structures are stored. The quantity Y is allowed to vary as a function of the time t since the offset of the stimulus. Then, for each model we can define a *forgetting function* that relates to time the expected number of stored structures. For each of the four models of visual memory compared in the main text, we derive expressions for its forgetting function.

1.1 Model #1: Classic

The classic model of the time course of visual memory, still used in modern applications[1, 2, 3], emerged in the 1960s from research using the partial report paradigm[4]. That work revealed the existence of iconic memory, a storage system with a high capacity and whose contents is short-lived, typically fading within a second[4]. Under the classic model, working memory and iconic memory are together responsible for behavioral performance. The contribution of working memory is at most its full capacity β, which is unchanging over time. The contribution of iconic memory above and beyond that of working memory is often called the "partial report superiority effect" and is at most all of the remaining $K - \beta$ things that were not stored in working memory. The partial report superiority effect has been found to decline exponentially as a function of time, and so the forgetting function of the classic model is given by

$$\mathrm{E}[Y(t)] = \begin{cases} \beta + (K - \beta)e^{\frac{-t}{\tau}} & \text{if } \beta \leq K \\ K & \text{if } \beta > K, \end{cases} \tag{1}$$

where τ is the mean lifetime of an item held in iconic memory.

1.2 Model #2: Pure death

The previous model assumed that working memory is stable over time. But working memory is known to degrade [5, 6]. For simplicity, we assume that degradation in working memory is a pure death process in which structures are lost independently over time and independently of each other, each having a mean lifetime of τ_2. First consider the case of $\beta \leq K$, where working memory is exhausted. In this case, the probability that a randomly-chosen structure is stored in working memory is $\frac{\beta}{K}e^{\frac{-t}{\tau_2}}$. The probability that it is stored in iconic memory is $e^{\frac{-t}{\tau}}$. Thus the forgetting function, which tracks the expected number of objects held in at least one of the two systems (those not held in neither system), is given by

$$\mathrm{E}[Y(t)] = K - K\left(1 - \frac{\beta}{K}e^{-t/\tau_2}\right)\left(1 - e^{-t/\tau}\right). \tag{2}$$

In the case of $\beta > K$, where working memory has room to spare, the term $\frac{\beta}{K}$ is replaced by unity because every structure is guaranteed a place. In the limit $\tau_2 \to \infty$, the pure death model reduces to the classic model.

1.3 Model #3: Sudden death

In 2009, Zhang & Luck proposed a "sudden death" model where after a window of initial stability lasting approximately four seconds, entire objects are lost over time, but the quality of those that survive is constant [5]. A reasonable way to formalize this model is to equip the pure death process with an initial grace period that lasts until time t_{death}. The forgetting function for this sudden death model is then governed by the classic model when $t < t_{\text{death}}$ and by the pure death model when $t \geq t_{\text{death}}$. Note that, because of the initial grace period, when used in the sudden death model, the term $\frac{t}{\tau_2}$ in Equation 2 must be replaced by $\frac{t-t_{\text{death}}}{\tau_2}$.

1.4 Model #4: Evolutionary model

Here, we derive the forgetting function of the evolutionary model with the stability threshold set to $s = 1$, i.e., the Moran process [7, 8]. For $s \geq 1$, we determined the forgetting functions numerically.

We consider N quanta, each of which is assigned to one of the K structures at any given time. We suppose that the structures stored in these quanta undergo a process of neutral drift, modeled as a continuous-time Moran or pairwise comparison process. It is convenient to scale time so that one time unit corresponds to N "generations" of this process, so that the contents of each quantum is updated once per unit time, on average.

1.4.1 Decay of founding lineages

When stimuli are first presented to a subject, each quantum is immediately assigned a single structure. We consider this to be the "founding generation" of structures stored in memory. At any subsequent time, the contents of each quantum will be a copy (or a copy-of-a-copy, etc.) of a member of this founding generation. Over time, the lineages (copies and copies-of-copies, etc.) of this founding generation may grow or disappear through random drift. Eventually only one lineage will remain.

We first ask how many lineages from the founding generation will survive to time $t > 0$. This question can be addressed using results from population genetics. We represent the number of founding lineages that persist at time t a the random variable $X(t)$. The expectation of this random variable is [9]:

$$\mathrm{E}[X(t)] = 1 + \sum_{\ell=2}^{M} (2\ell - 1) \frac{\binom{M}{\ell}}{\binom{M+\ell-1}{\ell}} e^{-\binom{\ell}{2}t}.$$

1.4.2 The forgetting function

We now consider the forgetting function—that is, the expected number of distinct structures that survive in memory at a given time. We suppose that, at time $t = 0$, each quantum is assigned randomly to one of K structures. We represent the the number of structures remaining at time $t \geq 0$ by the random variable $Y(t)$. The expected number of structures remembered at time t can be written as

$$\mathrm{E}[Y(t)] = 1 + \sum_{\ell=2}^{N} C_{\ell}^{N,K} e^{-\binom{\ell}{2}t}, \tag{3}$$

with the coefficients $C_\ell^{N,K}$ given by

$$C_\ell^{N,K} = (-1)^\ell (2\ell - 1) \frac{\binom{N}{\ell}}{\binom{N+\ell-1}{\ell}} \frac{K-1}{K} \, {}_2F_1\left(\ell + 1, 2 - \ell, 2, \frac{K-1}{K}\right). \tag{4}$$

Above, ${}_2F_1$ is the hypergeometric function. The derivation of Eq. (3) is given in the next two sections.

In the limit $N \to \infty$ (that is, if memory is regarded as a continuous resource) the forgetting function (3) converges to

$$\mathrm{E}[Y(t)] = 1 + \sum_{\ell=2}^{\infty} C_\ell^K e^{-\binom{\ell}{2} t},$$

with

$$C_\ell^K = (-1)^\ell (2\ell - 1) \frac{K-1}{K} \, {}_2F_1\left(\ell + 1, 2 - \ell, 2, \frac{K-1}{K}\right).$$

1.4.3 Trinomial coefficients

Our derivation of the forgetting function (3) relies on identities involving trinomial coefficients. For nonnegative integers M, i, j with $i + j \le M$, the corresponding trinomial coefficient is defined as

$$\binom{M}{i \quad j \quad M-i-j} = \frac{K!}{i!j!(K-i-j)!}.$$

Trinomial coefficients arise as coefficients in the expansion of $(x + y + z)^M$. In particular, we have

$$(-x + y + 1)^M = \sum_{\substack{i+j=M \\ i \ge 0, j \ge 0}} (-1)^i \binom{M}{i \quad j \quad M-i-j} x^i y^j.$$

From the above expansion, we can derive the following relations:

$$\begin{aligned} \sum_{i=0}^{M-j} (-1)^i \binom{M}{i \quad j \quad M-i-j} &= \frac{1}{j!} \frac{\partial^j}{\partial y^j} (-x + y + 1)^M \Big|_{(x,y)=(1,0)} \\ &= \binom{M}{j} (-x + y + 1)^{M-j} \Big|_{(x,y)=(1,0)} \\ &= \begin{cases} 1 & j = M \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{5}$$

$$\sum_{i=0}^{M-j}(-1)^i i\binom{M}{i \quad j \quad M-i-j} = \frac{1}{j!}\frac{\partial}{\partial x}\frac{\partial^j}{\partial y^j}(-x+y+z)^M\Big|_{(x,y,z)=(1,0,1)}$$
$$= \binom{M}{j}\frac{\partial}{\partial x}(-x+y+z)^{M-j}\Big|_{(x,y,z)=(1,1,0)}$$
$$= -\binom{M}{j}(M-j)(-x+y+z)^{M-j-1}\Big|_{(x,y,z)=(1,1,0)}$$
$$= \begin{cases} -M & j = M-1 \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Combining identities (5) and (6) yields a third identity:

$$\sum_{k=j}^{M}(-1)^{k-j}k\binom{M}{M-k \quad k-j \quad j} = \sum_{i=0}^{M-j}(-1)^i(i+j)\binom{M}{M-i-j \quad i \quad j}$$
$$= \sum_{i=0}^{M-j}(-1)^i i\binom{M}{i \quad j \quad M-i-j}$$
$$+ j\sum_{i=0}^{M-j}(-1)^i\binom{M}{i \quad j \quad M-i-j}$$
$$= \begin{cases} -M & j = M-1 \\ M & j = M \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

1.4.4 Derivation of the forgetting function

We now derive the forgetting function (3). First we suppose that that n of the N founding lineages remain after time t; that is, $X(t) = n$. Since neutral drift does not favor any structure over any other, we can regard these n lineages as being assigned randomly among the K structures. This situation thus reduces to the classical probability problem of randomly partitioning a set of n elements into K or fewer subsets.

For $k \leq n$, the probability that k of the K items are represented in these n lineages is

$$\Pr[Y(t) = k | X(t) = n] = \frac{\binom{K}{k}\left\{ n \atop k \right\} k!}{K^n}. \tag{8}$$

Above, $\left\{ n \atop k \right\}$ denotes the (n,k)th Stirling number of the second kind—that is, the number of ways to partition a set of n elements into k non-empty subsets. This Stirling number can be obtained by the formula

$$\left\{ n \atop k \right\} = \frac{1}{k!}\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}j^n. \tag{9}$$

Combining Eqs. (8) and (9) yields

$$\Pr[Y(t) = k | X(t) = n] = \binom{K}{k} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \left(\frac{j}{K}\right)^n,$$

or equivalently, upon rearranging,

$$\Pr[Y(t) = k | X(t) = n] = \sum_{j=0}^{k} (-1)^{k-j} \binom{K}{K-k \quad k-j \quad j} \left(\frac{j}{K}\right)^n. \tag{10}$$

The trinomial coefficient in Eq. (10) arises via the relation

$$\binom{K}{k}\binom{k}{j} = \binom{K}{K-k \quad k-j \quad j}.$$

Now we consider the overall expected number of items remembered at time t by summing Eq. (10) over values of n weighted by their probabilities:

$$\begin{aligned} \mathrm{E}[Y(t)] &= \sum_{k=1}^{K} k \sum_{j=0}^{k} (-1)^{k-j} \binom{K}{K-k \quad k-j \quad j} \sum_{n=1}^{N} \left(\frac{j}{K}\right)^n \Pr[X(t) = n] \\ &= \sum_{k=1}^{K} k \sum_{j=0}^{k} (-1)^{k-j} \binom{K}{K-k \quad k-j \quad j} G(j/K; t), \end{aligned} \tag{11}$$

Above, $G(x;t)$ is the probability generating function of $X(t)$:

$$G(x;t) = \sum_{n=1}^{N} x^n \Pr[X(t) = n].$$

We use a previously discovered [9] formula for this generating function:

$$G(x;t) = x + x(1-x) \sum_{\ell=2}^{N} (2\ell - 1)(-1)^{\ell+1} \frac{\binom{N}{\ell}}{\binom{N+\ell-1}{\ell}} \, {}_2F_1(\ell + 1, 2 - \ell, 2, x) \, e^{-\binom{\ell}{2} t}. \tag{12}$$

Substituting in Eq. (11), we obtain

$$\begin{aligned} \mathrm{E}[Y(t)] = &\sum_{k=1}^{K} k \sum_{j=0}^{k} (-1)^{k-j} \binom{K}{K-k \quad k-j \quad j} \\ &\times \left[\frac{j}{K} + \frac{j}{K}\left(1 - \frac{j}{K}\right) \sum_{\ell=2}^{N} (2\ell - 1)(-1)^{\ell+1} \frac{\binom{N}{\ell}}{\binom{N+\ell-1}{\ell}} \, {}_2F_1(\ell + 1, 2 - \ell, 2, j/K) \, e^{-\binom{\ell}{2} t} \right] \end{aligned} \tag{13}$$

Using identity (6) from section 1.4.3, we can simplify the term that is linear in j/K:

$$\sum_{k=1}^{K} k \sum_{j=0}^{k} \frac{j}{K} (-1)^{k-j} \binom{K}{K-k \quad k-j \quad j} = 1.$$

Eq. (13) therefore reduces to

$$\mathrm{E}[Y(t)] = 1 + \sum_{k=1}^{K} k \sum_{j=0}^{k} (-1)^{k-j} \binom{K}{K-k \quad k-j \quad j} \times \frac{j}{K}\left(1-\frac{j}{K}\right) \sum_{\ell=2}^{N} (2\ell-1)(-1)^{\ell+1} \frac{\binom{N}{\ell}}{\binom{N+\ell-1}{\ell}} \, {}_2F_1(\ell+1, 2-\ell, 2, j/K)\, e^{-\binom{\ell}{2}t}.$$

In summary, the expected number of items remembered can be written as

$$\mathrm{E}[Y(t)] = 1 + \sum_{\ell=2}^{M} C_\ell^{N,K} e^{-\binom{\ell}{2}t},$$

with

$$C_\ell^{N,K} = (-1)^{\ell+1}(2\ell-1)\frac{\binom{N}{\ell}}{\binom{N+\ell-1}{\ell}} \times \sum_{k=1}^{K} k \sum_{j=0}^{k} (-1)^{k-j} \binom{K}{K-k \quad k-j \quad j} \frac{j}{K}\left(1-\frac{j}{K}\right) \times \; {}_2F_1(\ell+1, 2-\ell, 2, j/K). \quad (14)$$

To simplify this expression for $C_\ell^{N,K}$ we reorder sums:

$$\begin{aligned} &\sum_{k=1}^{K} k \sum_{j=0}^{k} (-1)^{k-j} \binom{K}{K-k \quad k-j \quad j} \frac{j}{K}\left(1-\frac{j}{K}\right) \, {}_2F_1(\ell+1, 2-\ell, 2, j/K) \\ =&\sum_{j=0}^{K} \frac{j}{K}\left(1-\frac{j}{K}\right) \, {}_2F_1(\ell+1, 2-\ell, 2, j/K) \sum_{k=j}^{K} (-1)^{k-j} k \binom{K}{K-k \quad k-j \quad j}. \end{aligned} \quad (15)$$

Simplifying the second (nested) sum according to identity (7) from section 1.4.3, we obtain (4).

2 Supplementary Figures: Individual differences

People vary considerably in the capacity of their working memory systems, and these individual differences are correlated with intelligence, reasoning abilities, and reading comprehension [10, 11, 12, 13]. Our analysis procedure, which combines data from multiple participants into a single super-subject, masks such variability, and it is therefore important to consider the ways in which the presence of individual differences might impact our results.

First, variability might alter the predictions of the classic, sudden death, or pure death models, undermining our claim that they fail to capture features of the empirical forgetting curves. Second, variability might alter the predictions of the proposed evolutionary model, undermining the logic whereby a tight fit between model and data lends support to the model. We examine each of these possibilities below.

2.1 Individual differences in the classic model

In the classic model, variability can arise through individual differences in the initial capacity K, which is the number of structures encoded in working memory. Through simulation, we inject individual differences by drawing $1 - g$, the probability of encoding each object, from a Beta distribution with parameters chosen to cover a reasonable range of variability. Figure S1 shows that individual differences of this sort have no impact on the resulting curves.

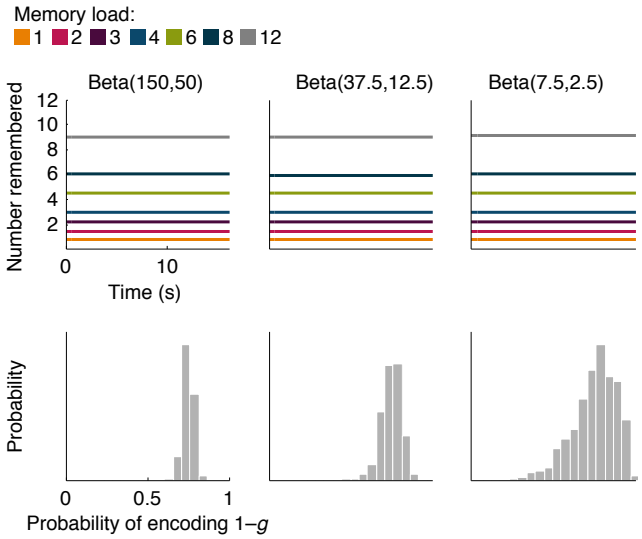

Fig. S 1: Individual differences in the classic model. The bottom row shows histograms of $1 - g$, the probability of successfully encoding an object in working memory. The top row shows the resulting forgetting functions, averaged over participants. Moving rightward, columns have greater individual differences.

2.2 Variability in the pure death model

In the pure death model, variability can arise in two ways: through individual differences in the initial capacity β, and through individual differences in the mean lifetime τ. Variability in β is modeled in the same way as in the classic model. Through simulation, we inject variability into the mean lifetime by drawing t from a log normal distribution. Figure S2 shows that variability in β has no impact on the resulting curves and that variability in τ bends each curve, but does not change the relationship between them, which would be needed to reproduce the effects of load-dependent stability or crossover.

2.3 Variability in the sudden death model

In the sudden death model, variability can arise in three ways: through individual differences in (1) the initial capacity β, (2) the mean lifetime τ, and (3) the length of the window of initial stability t_d. Variability in β and τ are modeled in the same way as in the classic and pure death models.

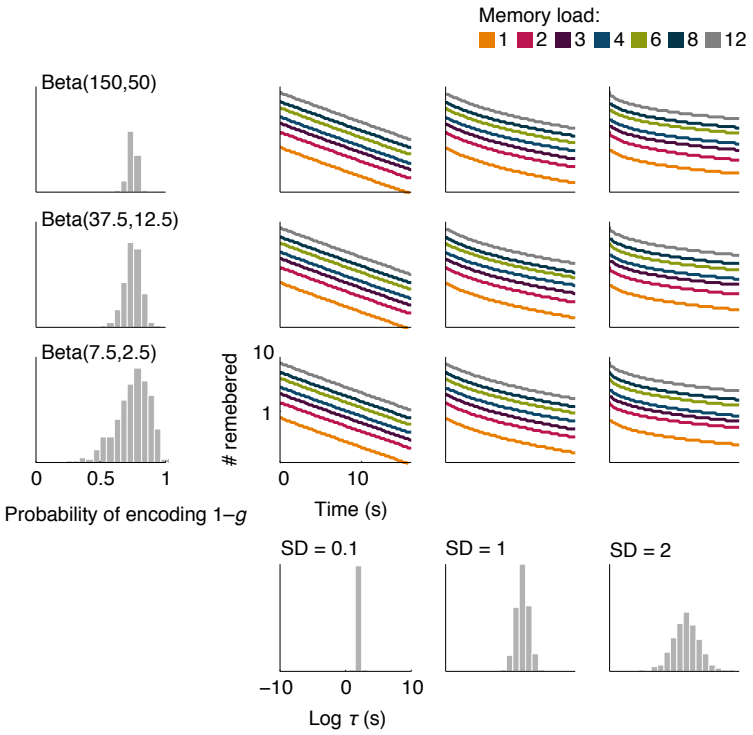

Fig. S 2: Individual differences in the pure death model. The leftmost column shows histograms of $1-g$, the probability of successfully encoding an object in working memory. The bottommost row show histograms of τ, the mean lifetime. There are nine plots, one for each pair of distributions on $1-g$ and τ. Moving rightward, columns have greater individual differences in τ. Moving downward, rows have greater individual differences in $1-g$. The y-axis is logarithmic to highlight shifts away from an exponential function (a straight line).

Through simulation, we inject variability into t_d by drawing it from a log normal distribution. As before, variability in β has no impact on the resulting curves. Figure S3 shows that variability in τ has the same effects as it does in the pure death model and that variability in t_d softens the corner at time points directly before and after the cutoff. As with the pure death model, these individual differences change the shape of the curves, but do not impact the relationship between them.

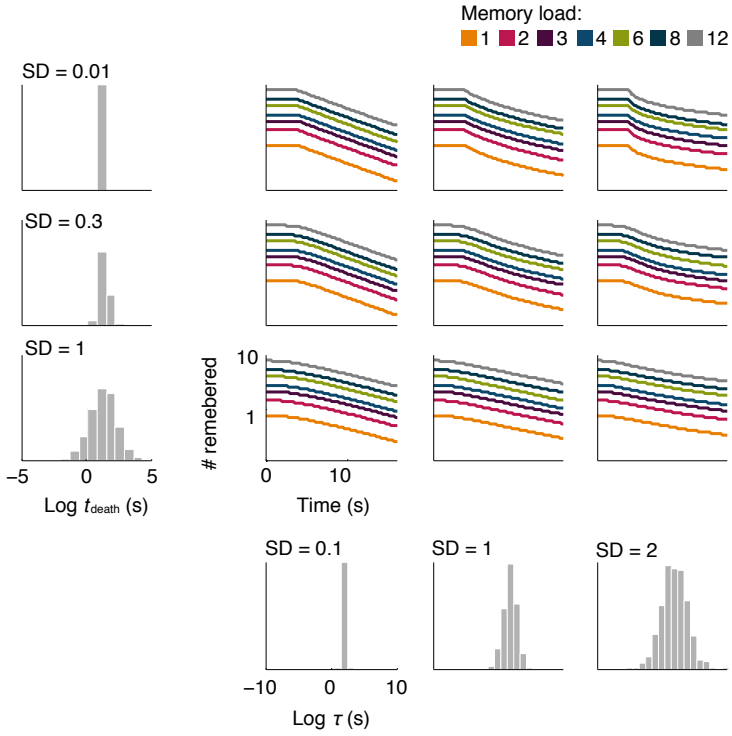

Fig. S 3: Individual differences in the sudden death model. The leftmost column shows histograms of t_{death}, the probability of successfully encoding an object in working memory. The bottommost row show histograms of τ, the mean lifetime. There are nine plots, one for each pair of distributions on t_{death} and τ. Moving rightward, columns have greater individual differences in τ. Moving downward, rows have greater individual differences in t_{death}. The y-axis is logarithmic to highlight shifts away from an exponential function (a straight line).

2.4 Variability in the evolutionary model

In the proposed evolutionary model, variability can arise in three ways: through individual differences in (1) the number of quanta N, (2) the duration of one time step t_{step}, or (3) the stability threshold s. Through simulation, we inject variability into each parameter and observe the effects on the predicted forgetting functions. Drawing N from a (discretized) normal distribution, we find that individual differences have a greater benefit to high memory loads than to low loads, thereby leading to a slight weakening of the crossover effect and load-dependent stability (Fig. S4). How-

ever, a crossover is seen even with high levels of individual differences (SD of ± 2 ln units). Next, drawing the stability threshold from a discrete uniform distribution, we again find that individual differences have a greater benefit to high memory loads than to low loads, with considerably less crossover, but only a miniscule effect on the presence of load-dependent stability (Fig. S5). Lastly, drawing t_{step} from a log normal distribution, we once again find the same result, with slight weakening of both load-dependent stability and crossover (Fig. S6). Together, these results suggest that the predictions of the proposed evolutionary model are tolerant to large individual differences in N, moderate individual differences in k, and large individual differences in t_{step}.

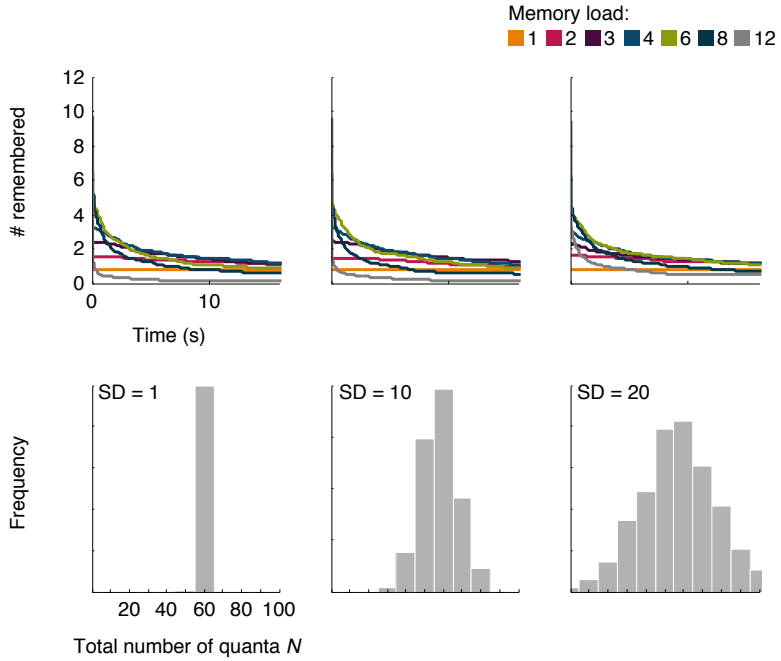

Fig. S 4: Individual differences in the number of quanta of the evolutionary model. The bottom row shows histograms of N, the total number of quanta. Moving rightward, columns have greater individual differences in N. The top row shows the corresponding forgetting functions.

3 Supplementary Figure: The effects of practice

Here, we consider the effects of practice by tracking performance as it changes over the course of the experiment's 70 trials (Fig. S7). The number of remembered objects dropped slightly (linear correlation, $r = -0.30, p = 0.013$), roughly 0.01% per trial (slope of linear regression, -0.002 object/trial; intercept, 2.2 objects). There were no significant changes in memory quality ($r = 0.15, p = 0.22$) or bias ($r = 0.03, p = 0.81$). This suggests that our training procedure was sufficient for participants to perform the task well.

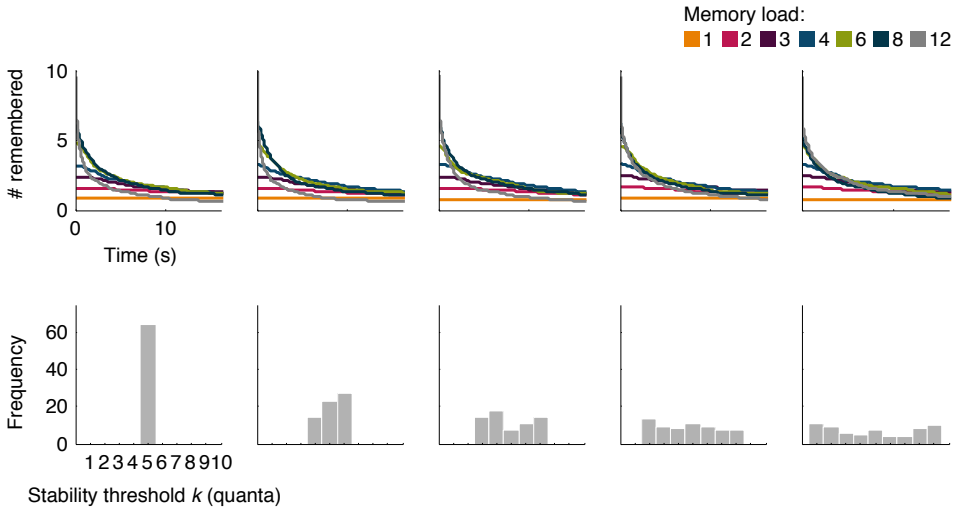

Fig. S 5: Individual differences in the stability threshold. The bottom row shows histograms of k, the stability threshold. Moving rightward, columns have greater individual differences in k. The top row shows the corresponding forgetting functions.

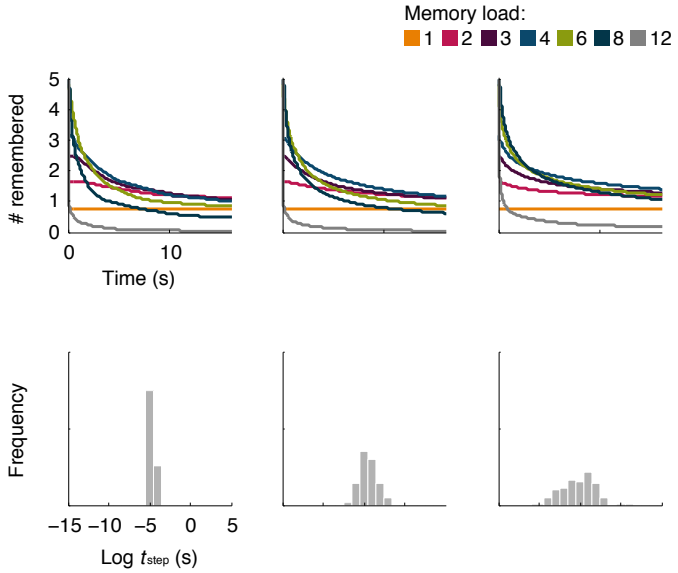

Fig. S 6: Individual differences in the rate of degradation. The bottom row shows histograms of t_{step}, the duration of a time step. Moving rightward, columns have greater individual differences in t_{step}. The top row shows the corresponding forgetting functions.

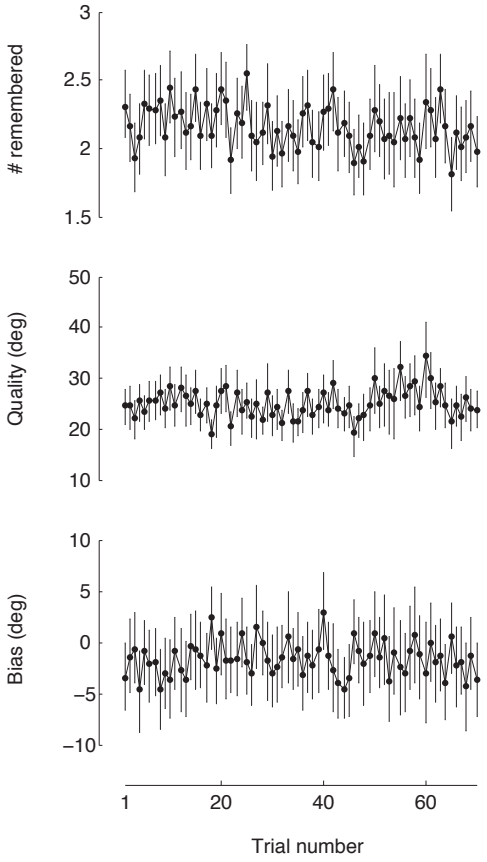

Fig. S 7: The effects of practice. Each subplot shows changes in performance as a function of trial number. The upper plot shows changes in the number of remembered objects. The middle plot shows changes in memory quality (lower is better). The lower plot shows changes in bias.

4 Supplementary Figure: Laboratory replication

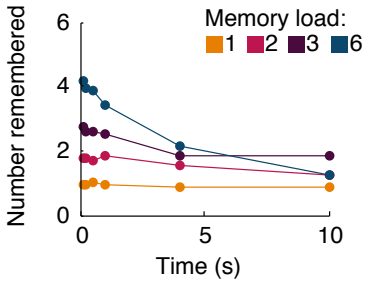

Fig. S 8: Replication in the lab. We replicated the online experiments in the lab with a group of six participants. The tested memory loads were 1, 2, 3, and 6. The tested durations were 0.125, 0.25, 0.5, 1, 4, and 10 s. Participants were each tested for 6-8 sessions of 360 trials, 15 trials per condition (pairing of duration and memory load), in random order. Data were fit with a hierarchical version of the 2-component model of Zhang & Luck [14]. The plotted data is the population mean. Data were fit using the MemToolbox 1.0.0. [15].

References

[1] Lu, Z.-L., Neuse, J., Madigan, S., and Dosher, B. A. Fast decay of iconic memory in observers with mild cognitive impairments. *Proceedings of the National Academy of Sciences* **102**(5), 1797–1802 (2005).

[2] Kuhbandner, C., Spitzer, B., and Pekrun, R. Read-out of emotional information from iconic memory. *Psychological Science* **22**(5), 695–700 (2011).

[3] Hahn, B., Kappenman, E. S., Robinson, B. M., Fuller, R. L., Luck, S. J., and Gold, J. M. Iconic decay in schizophrenia. *Schizophrenia Bulletin* **37**(5), 950–957 (2011).

[4] Sperling, G. The information available in brief visual presentations. *Psychological Monographs* **74** (1960).

[5] Zhang, W. and Luck, S. J. Sudden death and gradual decay in visual working memory. *Psychological Science* **20**(4), 423–428 (2009).

[6] Yang, W. *Lifetime of human visual sensory memory: Properties and neural substrate.* PhD thesis, University of Pennsylvania, (1999).

[7] Moran, P. A. P. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society* **54**(1), 60–71 (1958).

[8] Moran, P. A. P. *The statistical processes of evolutionary theory.* Clarendon Press, (1962).

[9] Tavaré, S. Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoretical Population Biology* **26**(2), 119–164 (1984).

[10] Unsworth, N. and Engle, R. W. Working memory capacity and fluid abilities: Examining the correlation between operation span and raven. *Intelligence* **33**(1), 67–81 (2005).

[11] Daneman, M. and Carpenter, P. A. Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior* **19**(4), 450–466 (1980).

[12] Fukuda, K., Vogel, E., Mayr, U., and Awh, E. Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic bulletin & review* **17**(5), 673–679 (2010).

[13] Kyllonen, P. C. and Christal, R. E. Reasoning ability is (little more than) working-memory capacity?! *Intelligence* **14**(4), 389–433 (1990).

[14] Zhang, W. and Luck, S. J. Discrete fixed-resolution representations in visual working memory. *Nature* **453**, 233–235 (2008).

[15] Suchow, J. W., Brady, T. F., Fougnie, D., and Alvarez, G. A. Modeling visual working memory with the memtoolbox. *Journal of Vision* **13**(10) (2013).