Report

Current Biology

Spatial and temporal massive memory in humans

Highlights

- People have spatial massive memory for locations of dozens of briefly seen objects
- People also have temporal massive memory for when those objects were seen
- This is recall of where and when items were presented, not familiarity/recognition

Authors

Jeremy M. Wolfe, Farahnaz A. Wick, Maruti Mishra, Joseph DeGutis, Wanyi Lyu

Correspondence

jwolfe@bwh.harvard.edu

In brief

It is well known that humans have massive memory for identities of recently presented visual objects. Wolfe et al. show that there is spatial massive memory (SMM) for where such items were presented and temporal massive memory (TMM) for when the items were presented. SMM and TMM are measured using recall and not familiarity/recognition methods.

Wolfe et al., 2023, Current Biology *33*, 405–410
January 23, 2023
https://doi.org/10.1016/j.cub.2022.12.040

CellPress

Current Biology

CellPress

Report

Spatial and temporal massive memory in humans

Jeremy M. Wolfe,[1,2,3,9,*] Farahnaz A. Wick,[4] Maruti Mishra,[5] Joseph DeGutis,[6,7] and Wanyi Lyu[8]
[1]Visual Attention Lab, Department of Surgery, Visual Attention Lab, Brigham & Women's Hospital, 900 Commonwealth Avenue, Boston, MA 02115, USA
[2]Departments of Ophthalmology & Radiology, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA
[3]Department of Psychological & Brain Sciences, Boston University, Boston, MA, USA
[4]Enveritas, 450 Lexington Avenue, New York, NY 10011, USA
[5]Department of Psychology, University of Richmond, Richmond Hall, 114 UR Drive, Richmond, VA 23173, USA
[6]Boston Attention and Learning Laboratory, VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130, USA
[7]Department of Psychiatry, Harvard Medical School, 940 Belmont Street, Brockton, MA 02301, USA
[8]Department of Biology, Centre for Vision Research, Vision Science to Application, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada
[9]Lead contact
*Correspondence: jwolfe@bwh.harvard.edu
https://doi.org/10.1016/j.cub.2022.12.040

SUMMARY

It is well known that humans have a massive memory for pictures and scenes.[1–4] They show an ability to encode thousands of images with only a few seconds of exposure to each. In addition to this massive memory for "what" observers have seen, three experiments reported here show that observers have a "spatial massive memory" (SMM) for "where" stimuli have been seen and a "temporal massive memory" (TMM) for "when" stimuli have been seen. The positions in time and space for at least dozens of items can be reported with good, if not perfect accuracy. Previous work has suggested that there might be good memory for stimulus location,[5,6] but there do not seem to have been concerted efforts to measure the extent of this memory. Moreover, in our method, observers are recalling where items were located and not merely recognizing the correct location. This is interesting because massive memory is sometimes thought to be limited to recognition tasks based on sense of familiarity.

RESULTS AND DISCUSSION

In experiment 1, as illustrated in Figure 1A, observers were asked to remember a variable number of objects, placed in a jittered 7 × 7 array (observer demographics are in Table 1). Each item was highlighted for 2 s, as shown by the red square surrounding one object in Figure 1A. All items were visible for N × 2 s where N is the set size. Objects were photographs taken from the stimuli used in Brady et al.[3] Set sizes were 5, 15, 25, and 49 items per screen. After the presentation of the items, all items were removed, and observers were tested on their immediate recognition and location memory. One item was presented in a neutral position. For new items, observers simply clicked a "new item" box. For items recognized as previously seen, they responded by clicking on the remembered location of the item. Observers received visual feedback about their old/new response accuracy and no feedback about their localization responses. Figure 1B illustrates a case where the observer responded that the item was "old" and localized (*) near but not exactly on the original location (#).

All encoded items were tested in random order, mixed with the same number of new items for a total of 2N testing trials. After the test period, a new set of stimuli was presented for encoding. This cycle of encoding and testing was repeated until the observer had encoded 300 total items. Thus, there were 60 screens of 5 items, 20 screens of 15 items, and 12 screens of 25 items. For the maximum set size of 49 items, there were six screens for a total of 294 items. Each observer saw just one set of 300 (or 294) unique stimuli. After all items were tested once, there was a second, retest period where all 300 encoded items were presented, one after the other, mixed with 300 fresh, new items. Observers made the same old/new and localization responses as in the initial test. In addition to diverse objects, we also tested face and door stimuli. For these difficult, within-category stimuli, observers were tested for 40 screens at a set size of five items per screen, so the total number of encoded items was 200.

Old/new performance

The signal detection measure of d' quantifies results of the standard old/new massive memory.

Performance on the initial test declined as set size increased (Figure S1) from $d' = \sim 3.0$ at set size 5 to $d' < 2.0$ at set size 49 (one-way ANOVA; $F(3, 55) = 4.9$, $p = 0.0041$, partial eta-sq = 0.21). During retest, where all observers are tested on 300 old versus 300 new items, all set sizes produced d' of approximately 1.4 (one-way ANOVA; $F(3, 55) = 0.38$, $p = 0.77$, partial eta-sq = 0.02). This corresponds to >80% accuracy, comparable to other massive memory studies.[1–4] Faces and doors produced very poor performance, especially at retest (faces, $D' = 0.5$, accuracy = 61%; doors, $D' = 0.22$, accuracy = 55%).

Localization performance

Localization capacity is measured by calculating the proportion of accurate localizing clicks and subtracting the proportion that could

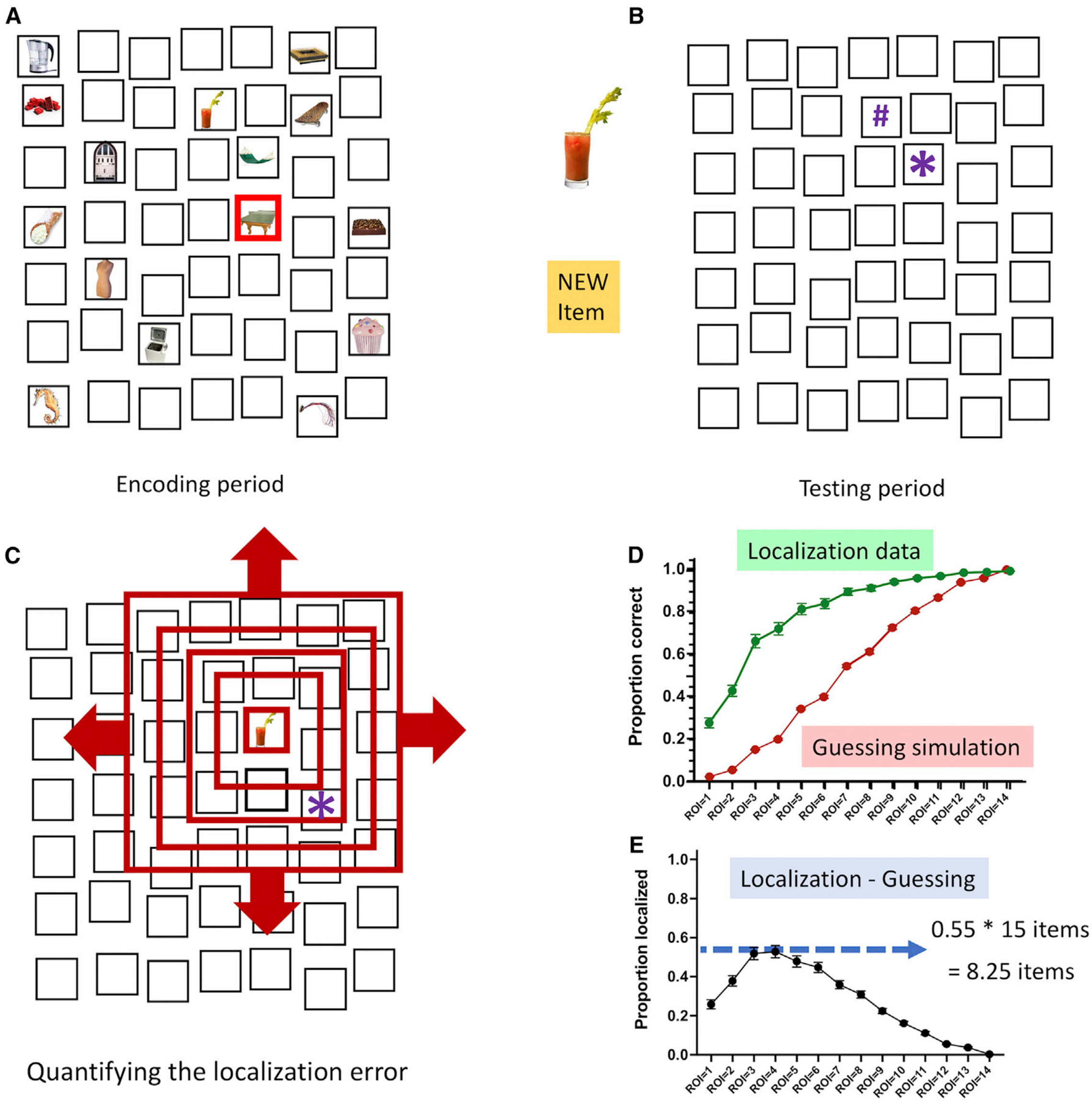

Figure 1. Stimuli and analysis for experiments 1 and 2

(A) Observers see an array of N objects (here 15) presented for 2 × N s. Items are highlighted in red, one by one in random order for 2 s each.

(B) At test, they see objects, placed to the left of the array, that can be old or new. If new, observers click the new item box. If old (previously presented), they are asked to locate the object on the grid. Here, the drink was in row 2, column 4 (#) but is localized in row 3, column 5 (*).

(C) Regions of interest (ROIs) are defined around the location of an old item. The red boxes show square ROIs of 1–5 cells on a side (ROI = 1 to ROI = 5). Each response (asterisk) falls inside some larger ROIs and outside the smaller ones. This response falls inside ROIs $\geq$ 3.

(D) As the ROI increases in size, the proportion of answers inside the ROI rises from 0 to 1 (green) as does the proportion of answers that could fall in the ROI given random guessing (red).

(E) The difference between data and guessing gives a conservative measure of capacity, which is 55% of 15 items, or 8.25 items here, for set size 15 and data as shown in (D).

Similar functions for all conditions in experiment 1 are shown in Figure S2.

be attributed to chance responding. As illustrated in Figure 1C, suppose that the drink picture had been shown in the 3rd row, 5th column of the 7 × 7 array and suppose that the observer localized it to the 4th row, 6th column (*). We score that response as correct if it falls inside a square region of interest (ROI). The proportion of such “correct” responses obviously increases as the size of the ROI increases. In Figure 1C, the response falls outside ROI = 1, but inside ROI = 3 (equivalent to ±1 cell). The rising function of proportion falling inside the ROI is shown in green in Figure 1D using data from the first test for observers seeing set size 15. Of course, an observer’s response could land inside even a small ROI by chance. To measure this chance function that must also rise from 0 to 1, we paired a target location with responses to different target items by the same observer. This takes into account the center bias and

Table 1. Demographic details for experiments 1 and 2

	Female	Male	None or other	Total	Excluded	After exclusion
Experiment 1						
Set size 5	8	8	1	17	2	15
Set size 15	10	3	0	13	0	13
Set size 25	7	7	–	14	2	12
Set size 49	10	13	–	23	5	18
Doors, set size 5	12	9	–	21	8	13
Faces, set size 5	6	10	2	18	3	15
Experiment 2						
Set size 5	4	9	–	13	1	12
Set size 15	4	4	4	12	0	12
Set size 25	8	4	–	12	0	12
Set size 49	4	4	4	12	0	12
Faces, set size 5	4	8	–	12	0	12
Totals	77	79	11	167	21	146

localization biases of individual observers. We repeated this for 100 pairings of target and response for each target. The red guessing simulation line in Figure 1D shows this chance function. The peak of the difference function (test – chance; Figure 1E) can be taken as one, possibly conservative estimate of the capacity of spatial memory. In Figure 1E, the difference in proportions peaked at about 0.55 for an ROIs of 3 or 4 cells. Note that an ROI of 3 measures the proportion of targets localized to within ±1 cell from the true location. For set size 15, this would correspond to localization within ±1 cell of 0.55 × 15 = 8.25 items. The localization-minus-chance functions have the same shape for all conditions (Figure S2). Accordingly, we will use the data for ROI = 3 as our estimate of spatial memory in subsequent analyses.

Figures 2A and 2Cshow test and retest results for experiment 1 while Figures 2B and 2D show results for experiment 2 (replication; see below). Results are expressed as the estimated number of items correctly localized within ROI = 3. These values are obtained by multiplying the ROI = 3 difference proportions (data – chance) by the number of items tested (memory set size for initial test; total number of encoded items for retest). At initial test (Figure 2A), the number of items correctly localized from one screen obviously increased as the number of items presented in that screen increases. Note that observers correctly localize a roughly constant proportion of items at all set sizes. At initial test, at set size 5, object, face, and door conditions do not differ ($F(2, 42) = 0.86$, $p = 0.43$, partial eta-sq = 0.04).

The more interesting results are the retest results (Figure 2C) where observers are tested after seeing 300 objects or 200 faces and doors. Many observers can approximately localize 100 or more objects. Though some observers performed poorly, the average capacity was >70 in three of the four set size conditions. Though, unsurprisingly, this spatial location memory was not as good as the simple old/new recognition memory, the size of what we will call spatial massive memory (SMM) capacity was impressive. A one-way ANOVA showed a significant effect of set size on objects remembered at retest ($F(3, 55) = 4.08$, $p = 0.011$, partial eta-sq = 0.18), entirely due to the effect of the anomalously weaker result for set size 25 (but see experiment 2). The performance with faces and doors was poor across all observers (14 faces and 6 doors on average, out of 200), significantly worse than the performance for diverse objects at the same set size of 5 ($F(2, 42) = 16.97$, $p < 0.0001$, partial eta-sq = 0.45). Post hoc comparisons showed significant differences between face and door results and object results (Tukey's multiple comparisons test; $p < 0.0001$). Doors and faces do not differ from each other ($p = 0.8$).

Apparently, as with simple recognition memory, a few seconds of exposure to diverse, colored objects is enough to encode many locations into some form of long-term memory. Note that the locations were selected in random, and about six objects would have been presented in each of the 49 cells of the display. Exploratory analysis showed no evidence for an interference effect (errors getting larger over the course of 300 trials of retest) or, for that matter, of a learning or recency effect (errors getting smaller).

Experiment 2: Replication

Figures 2B and 2D show the SMM results for a replication of experiment 1. In experiment 2, we increased the highlighting time per item from 2 to 3 s, in an effort to reduce the large performance variation in experiment 1. Faces were tested to see if longer exposure would rescue performance. The old/new recognition results were broadly similar to experiment 1 (Figures S1C and S1D). A one-way ANOVA showed that performance was somewhat better for the smallest set size of 5 ($F(3,44) = 6.056$, $p = 0.0015$, partial eta-sq = 0.29). On retest, average d' in object conditions was 1.8, corresponding to over 80% accuracy. There were no reliable differences between object set sizes ($F(3,44) = 0.6629$, $p = 0.5793$, partial eta-sq = 0.04). The face condition remains significantly worse than the object conditions ($d' = 0.43$, 58% accuracy).

Experiment 2 replicates the results of experiment 1. The SMM capacity estimates for the initial test grow with set size (Figure 2B). At retest (Figure 2D), with 300 objects, many observers could localize over 100 items within ±1 cell of the true object location. Again, there was considerable variation across observers within a condition. Some observers seemed to have decided not to do anything except guess during retest. The results of experiment 2 indicate that the poor performance at set size 25 in experiment 1 was, indeed, an anomaly. In experiment 2, set size 25 produced the highest rather than the lowest average localization memory. As in experiment 1, performance with face stimuli was much worse than performance with diverse objects.

Experiment 3: Spatial and temporal massive memory

In experiments 1 and 2, observers saw a set of stimuli and were then queried about those stimuli in an old/new testing phase. Experiment 3 used a single stream method to combine exposure and testing. Observers saw an object, placed in the 7 × 7 array. If it was seen for the first time, observers clicked on a "new object" box. If it had been seen, observers clicked on the location in the grid where they believed they had seen the item before. Moreover, they also clicked on a timeline, present on the screen, to indicate "when" they thought they had seen the item. Observers saw 150 items twice. They also saw some filler items that

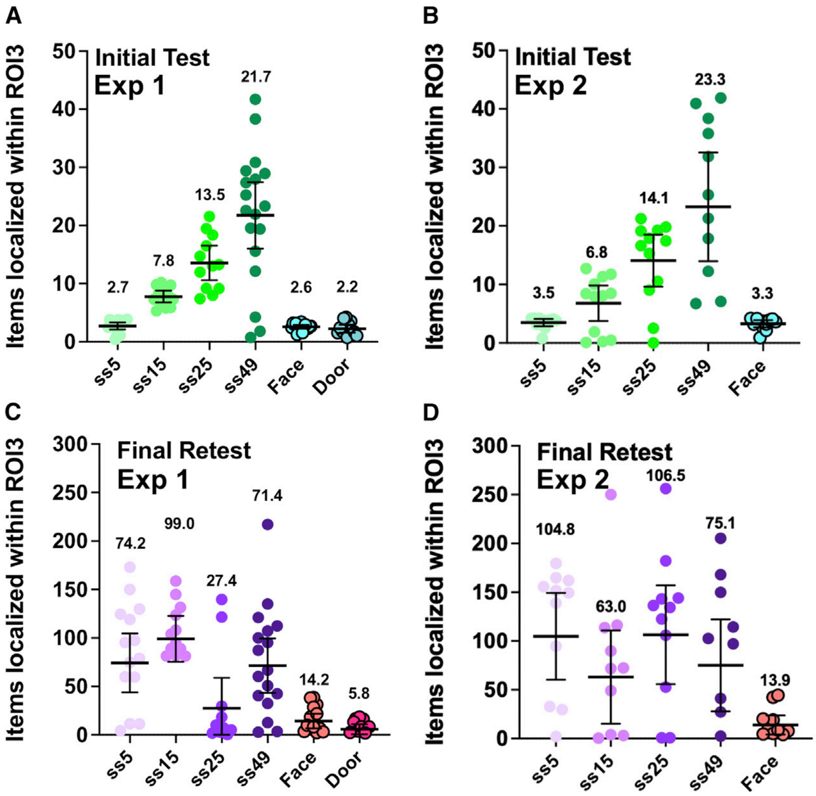

Figure 2. Results for experiments 1 and 2
Estimated number of items localized inside ROI3 at initial test and final retest for experiment 1 (A and C) and experiment 2 (B and D). Each data point represents one observer, error bars show mean ± 95% CI, and numbers show average values for that condition. Doors and faces are tested with set size 5 only. Simple recognition memory results are shown in Figure S1. Effects of grid position for experiment 2 are shown in Figure S4.

appeared only once. Each tested item had slightly less than a 50% chance of being an "old object." On average, observers saw 321 trials. Faces and doors were not tested in experiment 3 since there is no reason to assume that this method would show SMM for those stimuli given experiments 1 and 2.

Performance on the standard old/new test revealed the usual massive memory with a d′ averaging 2.9 (Figure 3A, overall accuracy 91%).

Figure 3B shows clear evidence for SMM, with capacity computed as in experiments 1 and 2. There appear to be three groups of observers. A group at 0% who did not or could not do the task. A group at about 0.33%, corresponding to localizing ∼50 items above the chance prediction and a group around 0.5% (75 items).

Figure 3C shows that there is also a temporal massive memory (TMM). For time, ROIs are defined as a percentage of the total time centered on the time when the object first appeared. Note that, as the guessing function makes clear, it is quite easy to get some trials right by guessing. If you have only completed 20% of the time, you cannot make more than a 20% error. Nevertheless, it is clear that many observers localized 60%–80% of old items to within ±10% of their correct time, against a 40% chance level, indicating a substantial temporal memory. A subset of observers actually did worse than chance, probably by always clicking approximately the same location on the timeline. Spatial and temporal performance were not correlated (r-sq = 0.01, p = 0.59) though this could be an issue of statistical power and would be worth further investigation. Interestingly, the observers who performed at chance in the spatial task were not the same observers who performed at chance in temporal task. The relatively large number of observers who were at chance in one task or the other may reflect the perils of on-line testing.

Experiment 3 makes it possible to look at the effects of the spacing between the first and second appearance of an object ("lag"). The algorithm controlling the sequence of stimuli produced lags that were roughly logarithmically distributed (more short lags than long). Somewhat unsurprisingly, there is an effect of lag (Figure S3). The proportion of targets localized within ±1 cell of the true location declined from 0.54 at $\log_2(\text{lag}) = 1$–0.35 at $\log_2(\text{lag}) = 6$. There was a main effect of $\log_2(\text{lag})$ (mixed effects analysis with Greenhouse-Geisser correction: $F(2.520, 47.46) = 4.663$, $p = 0.009$). The correlation of correctly localized targets with raw lag was significant though the effect was small (r-sq = 0.04, p = 0.027). Performance declined modestly as the lag increased. However, performance was far above chance in all $\log_2(\text{lag})$ bins (all $t(19) > 7$, all $p < 0.0001$).

DISCUSSION

Our spatial and temporal memory for objects may not be as impressive as the specialized memory of food-hoarding birds[7,8] or squirrels.[9] Nevertheless, it is clear from the experiments presented here that human spatial and temporal memory for objects can be quite massive. Unsurprisingly, it is not as robust as simple recognition memory, but it is clear that many observers can recall the location of over 100 items with a precision of ±1 cell in the 7 × 7 grid used here. The poor results for faces and doors suggest that this memory is constrained by the stimulus set. SMM and TMM maybe poor when items are similar to each other, for instance. The more similar the items are, the more likely they will swap locations in memory.[10] It is likely that SMM and TMM would vary with the memorability of the items.[11,12] These data show that SMM and TMM exist. Future work will define their limits.

One important aspect of these measures of SMM and TMM is the recognition that partial or imperfect knowledge is still knowledge. If localization to the exact cell in the 7 × 7 grid was required for a "correct" answer, estimates of spatial memory would be much smaller (Figure S2). However, knowing that an object is roughly "over there" is clearly a real memory and, often, all that is required to be a useful memory in a real-world task.

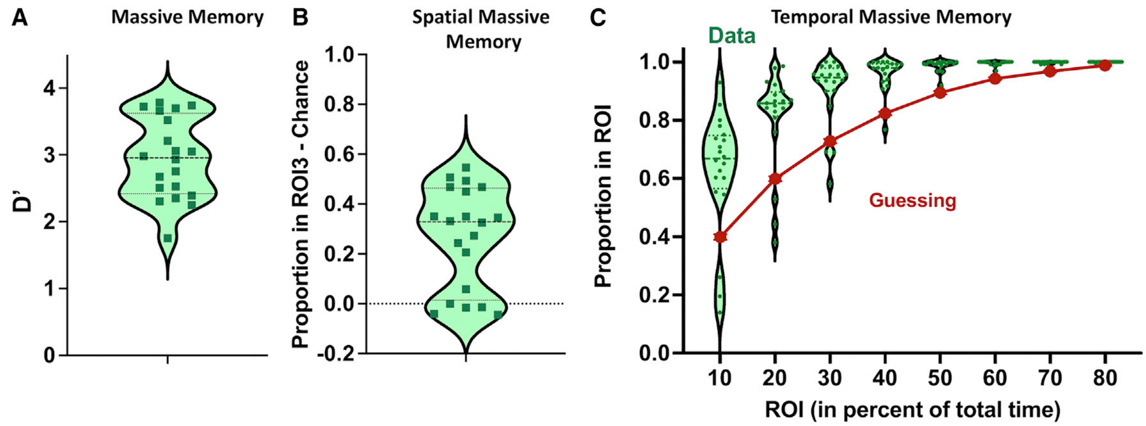

Figure 3. Results of experiment 3

(A) Standard old/new recognition memory, measured in d′ units.

(B) Spatial massive memory, shown as proportion of localizing clicks inside an ROI of ±1 cell (ROI = 3) with simulated chance performance subtracted.

(C) Temporal massive memory: violin plots show proportion of responses inside ROIs defined as percentages of the entire timeline. Red line shows simulated guessing performance. Dots show data from individual observers. Effects of lag between first and second appearance of an item are shown in Figure S3.

Thus, localizing 100 items to within ±1 cell represents a significant act of memory.

Much of the earlier work focused on classic questions in memory finding, for example, clear evidence for recency effects and somewhat weaker evidence for primacy.[13–15] One line of work has focused on the claim that women are better than men at spatial memory tasks and that this superiority is the product of human evolutionary history.[16,17] There appears to be some female advantage, though it might not be specific to location memory,[18] and the evolutionary hypothesis is definitely open to debate.[19] A larger study would be required to assess gender effects in the present experiments.

These results offer multiple avenues for further research. We made one set of methodological choices. They provide an existence proof for SMM and TMM, but other choices might reveal more robust effects. For instance, our experiments use a recall method to assess SMM and TMM. Using a two-alternative forced choice recognition task (was the boot in location A or B?) might be expected to increase SMM. In experiments 1 and 2, observers made two localization responses. It could be that the initial test responses influenced retest, though our exploratory analyses suggest that this is not the case. Interestingly, location in the grid seems to matter. Observers appear to be more accurate in the lower part of the grid (Figure S4), a finding that would be worth examining with other choices about spatial layout (e.g., using a horizontal plane, like a tabletop). The relationship between temporal and SMM should be more closely examined.[5,15,20] Interpreting the lack of correlation between SMM and TMM suffers from the usual problems of a negative result. Similarly, it would be interesting to determine whether memory for faces would improve if we used more distinctive, perhaps, famous faces. It might also be worth the time and effort to ask observers to encode more than the 300 objects used here. Here, we have demonstrated that memory for where and when an object appeared can be encoded and recovered with good, if not perfect precision for a large number of objects.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Demographic details for Experiments 1 and 2
- METHOD DETAILS
  - Stimuli
  - Display – In-person
  - Display – Online
  - Experiment 1
  - Experiment 2
  - Experiment 3
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Spatial memory capacity measure
  - Temporal memory capacity measure

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.12.040.

ACKNOWLEDGMENTS

This work was supported by NSF grant 1848783 to J.M.W. and the Harvard Mind Brain and Behavior program to J.M.W. and J.D.

AUTHOR CONTRIBUTIONS

Conceptualization, J.M.W.; methodology, J.M.W. and W.L.; formal analysis, J.M.W. and W.L.; investigation, W.L.; resources, J.M.W., F.A.W., M.M., J.D., and W.L.; writing – original draft, J.M.W.; writing – review & editing, J.M.W., F.A.W., M.M., J.D., and W.L.; funding acquisition, J.M.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One author self-identifies as an under-represented ethnic minority in science. We support inclusive, diverse, and equitable conduct of research.

Received: October 12, 2022
Revised: November 30, 2022
Accepted: December 15, 2022
Published: February 6, 2023

REFERENCES

1. Standing, L., Conezio, J., and Haber, R.N. (1970). Perception and memory for pictures: single trial learning of 2500 visual stimuli. Psychon. Sci. *19*, 73–74.
2. Standing, L. (1973). Learning 10,000 pictures. Q. J. Exp. Psychol. *25*, 207–222.
3. Brady, T.F., Konkle, T., Alvarez, G.A., and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. Proc. Natl. Acad. Sci. USA *105*, 14325–14329.
4. Konkle, T., Brady, T.F., Alvarez, G.A., and Oliva, A. (2010). Scene memory is more detailed than you think: the role of categories in visual long-term memory. Psychol. Sci. *21*, 1551–1556.
5. Noack, H., Lövdén, M., Schmiedek, F., and Lindenberger, U. (2013). Age-related differences in temporal and spatial dimensions of episodic memory performance before and after Hundred Days of practice. Psychol. Aging *28*, 467–480.
6. Naveh-Benjamin, M. (1987). Coding of spatial location information: an automatic process? J. Exp. Psychol. Learn. Mem. Cogn. *13*, 595–605.
7. Shettleworth, S.J. (1983). Memory in food-hoarding birds. Sci. Am. *248*, 102–110.
8. Feenders, G., and Smulders, T.V. (2008). Episodic-like memory in food-hoarding birds. In Handbook of Episodic Memory, *18*, E. Dere, A. Easton, L. Nadel, and J.P. Huston, eds. (Elsevier), pp. 197–216.
9. Jacobs, L.F., and Liman, E.R. (1991). Gray squirrels remember the locations of buried nuts. Anim. Behav. *41*, 103–110.
10. Pertzov, Y., Dong, M.Y., Peich, M.C., and Husain, M. (2012). Forgetting what was where: the fragility of object-location binding. PLoS One *7*, e48214.
11. Bainbridge, W.A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. Neuropsychologia *141*, 107408.
12. Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., and Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. Vision Res. *116*, 165–178.
13. Nairne, J.S., and Dutta, A. (1992). Spatial and temporal uncertainty in long-term memory. J. Mem. Lang. *31*, 396–407.
14. Walker, P., Hitch, G.J., and Duroe, S. (1993). The effect of visual similarity on short-term memory for spatial location: implications for the capacity of visual short-term memory. Acta Psychol. *83*, 203–224.
15. van Asselen, M., Van der Lubbe, R.H.J., and Postma, A. (2006). Are space and time automatically integrated in episodic memory? Memory *14*, 232–240.
16. Eals, M., and Silverman, I. (1994). The hunter-gather theory of spatial sex-differences – proximate factors mediating the female advantage in recall of object arrays. Ethol. Sociobiol. *15*, 95–105.
17. Silverman, I., and Eals, M. (1992). Sex differences in spatial abilities: evolutionary theory and data. In The Adapted Mind: Evolutionary Psychology and the Generation of Culture, I. Barkow, L. Cosmides, and J. Tooby, eds. (Oxford University Press), pp. 487–503.
18. Voyer, D., Postma, A., Brake, B., and Imperato-McGinley, J. (2007). Gender differences in object location memory: a meta-analysis. Psychon. Bull. Rev. *14*, 23–38.
19. Newcombe, N.S. (2010). On tending to our scientific knitting: thinking about gender in the context of evolution. In Handbook of Gender Research in Psychology: Volume 1: Gender Research in General and Experimental Psychology, J.C. Chrisler, and D.R. McCreary, eds. (Springer), pp. 259–274.
20. Scofield, J.E., and Johnson, J.D. (2022). The diminishing precision of memory for time. Psychon. Bull. Rev. *29*, 212–219.
21. Kleiner, M., Brainard, D.H., Pelli, D.G., Broussard, C., Wolf, T., and Niehorster, D. (2007). What's new in Psychtoolbox-3? Perception *36*, 1.
22. Ishihara, I. (1980). Ishihara's Tests for Color-Blindness: Concise Edition (Kanehara & Co., Ltd).
23. Baddeley, A.D., Hitch, G.J., Quinlan, P.T., Bowes, L., and Stone, R. (2016). Doors for memory: A searchable database. Q. J. Exp. Psychol. (Hove) *69*, 2111–2118.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Behavioral data: Experiment 1-2 with object	This paper	https://osf.io/t7eb5/
Behavioral data: Experiment 1-2 with faces	This paper	https://osf.io/tu5zp/
Behavioral data: Experiment 1 with doors	This paper	https://osf.io/rdvbe/
Behavioral data: Experiment 3	This paper	https://osf.io/9bshu/
Software and algorithms		
MATLAB	Mathworks	Matlab_R2016a
Psychtoolbox	Kleiner et al.[21]	http://psychtoolbox.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Jeremy M. Wolfe (jwolfe@bwh.harvard.edu).

Materials availability

Studies were preregistered on the Open Science Framework (OSF, see links in key resources table). This study did not generate any unique reagents.

Data and code availability

The experiment was programmed in MATLAB for in-person testing and in JavaScript environment for on-line testing. Raw data can be accessed at the Open Science Framework (OSF) repository and are publicly available as of the date of publication (see key resources table for links). This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The experimental procedures were carried out in accordance with the Institutional Review Board of Brigham and Women's Hospital (IRB #2009P001253). All subjects provided informed consent prior to running in the study. A total of 167 observers participated in Experiments 1 and 2. Of these, 21 were excluded for performing near chance at the old/new task. The demographic Table shows the breakdown by reported gender for each condition. We preregistered an intention to enroll 12 observers per condition. In some cases, more observers remained after exclusions. In Experiment 3, 23 observers were tested (8 female, 15 male, median age, 43 yrs). We preregistered an intention to enroll 20 observers and ended up with 22 observers after excluding one for poor performance. All observers had at least 20/25 acuity with correction if needed and passed the Ishihara color vision test.[22] Observers were paid $12/hour. 154 observers participated on-line. They were recruited via the Amazon Mechanical Turk and tested on CloudResearch online platform. Online observers were restricted to individuals located in the US, with an approval rate above 95%. All observers attested to 20/25 vision with correction and color vision. Online observers were paid $8/hour.

Experiment 1	Female	Male	none or other	Total	Excluded	after exclusion
set size 5	8	8	1	17	2	15
set size 15	10	3	0	13	0	13
set size 25	7	7		14	2	12
set size 49	10	13		23	5	18
Doors, set size 5	12	9		21	8	13
Faces, set size 5	6	10	2	18	3	15
Experiment 2						
set size 5	4	9		13	1	12

(Continued on next page)

Continued						
Experiment 1	Female	Male	none or other	Total	Excluded	after exclusion
set size 15	4	4	4	12	0	12
set size 25	8	4		12	0	12
set size 49	4	4	4	12	0	12
Faces, set size 5	4	8		12	0	12
Totals	77	79	11	167	21	146

Demographic details for Experiments 1 and 2

Different groups of observers were tested in each condition in order to keep the length of the experiment tractable (important for on-line studies). Within-subject comparisons between conditions would be more robust, but these were not central to the measurement of SMM and TMM.

METHOD DETAILS

Stimuli

Stimuli were colored photographs of objects, faces, and doors. Object images were taken from Brady et al.[3] Stimuli used in the experiments were randomly drawn from a set of 2400 available images. The face images were pooled from two image sources. To acquire an adequate number of unique faces (600), we included faces with either smiling (326) or neutral (274) expressions. The background and area below the neck were removed using a deep neural network trained on background vs foreground images. The door photographs were compiled by Baddeley et al. and have been used to study visual episodic memory and long-term memory.[23] They served as a comparison for the face stimuli. We selected a subset of the 600 images to match with the number of face stimuli set. Rectangular and plain doors were selected. Doors with bright colors or covered with writing were removed.

Display – In-person

Experiment 1 (set size 15) was run on a 21.5-inch iMac with a screen resolution of 1920 x 1080 pixels (refresh rate 60 Hz). The experiment was programmed in MATLAB (The MathWorks, Inc., Natick MA) with PsychToolBox extensions.[21] Items were displayed on a slightly irregular and visible 7 by 7 grid with black outline. The grid filled a square field with each side equal to 80% of the height of the monitor and each box subtended 11% of that field height. At an approximate viewing distance of 60 cm, this would correspond to approximately 3.9 deg boxes arranged in a 7x7 grid subtending approximately 30 deg on a side. The image stimuli were placed randomly within the available boxes. During the test phase, an image of the test item was presented to the left side of the display field along with an instruction “Click old location or click in new box”. The new box was located on the right side of the display field with text “click here if this is a new item”.

Display – Online

Experiment 1 (set size 5, 25, 49) and Experiment 2 (all set sizes) was written in JavaScript and hosted on-line. Online and MATLAB display were visually similar. The grid filled a square field with white background placed in the center of the browser. Each side of the field equaled to 50% of the browser viewport width and each box subtended 11% of that field height. The locations of the boxes are slightly jittered. The Experiment was presented in full screen. Due to the online nature of the task, the remote viewing conditions varied, but observers attested to completing the experiment on a desktop or laptop, not on a cell phone or other small display.

The choice of on-line vs in person testing was driven by Covid concerns.

Experiment 1

An example of stimuli display is shown in Figure 1A. The experiment consisted of one test period and one retest period. Observers were informed about both parts at the beginning of the experiment. They needed to achieve an overall accuracy of 60% for their old/new response (see below) in order to be compensated. As noted in the results, some Os seem to perform at chance on either the spatial or the temporal localization task. Because of the essentially exploratory nature of this experiment, we did not exclude observers for ’bad behavior’ on the localization tasks.

During the test period, on each screen, observers saw a set of colored objects, placed in a jittered 7 x 7 array of cells. The cells were visible to the observers throughout the experiment. The items were highlighted one by one by a red outline in random sequence for 2 seconds each. All items were visible for N x 2 seconds where N is the set size. Set sizes were 5, 15, 25, and 49 items per screen. Each observer was tested on one of the four set sizes. Observers were instructed to pay attention to the highlighted item and remember its identity and location. The set size manipulation is akin to a "blocked" vs "spaced" manipulation in a memory experiment. Did it matter if observers tried to encode larger or smaller numbers of items at one time? Beyond that, each set size condition acts as a replication of the basic SMM experiment.

After the presentation, all items were hidden. The cell array background remained visible. One test item was presented at a time at a neutral location to the left of the array. Observers were instructed to click on a ''new item'' box if they had not previously seen the object. If they recognized the object as part of the encoded set, they responded by clicking on the remembered location of this old item. This allows us to collect old/new accuracy data as well as measuring the precision of recall of the encoded item location. Observers were instructed to respond as quickly and accurately as possible, but there was no time limit. All encoded items were tested in random order, mixed with the same number of new items for a total of 2N testing trials. Observers received both visual ('correct' vs 'wrong', presented at the top of the screen) and audio feedback on their old/new response accuracy and did not receive feedback on the localization response.

After all 2N items were tested, a new set of stimuli was presented for encoding. The cycle of encoding and testing was repeated until the observers had been tested on 300 encoded items in total. Thus, observers either saw 60 screens of 5 items, 20 screens of 15 items, 12 screens of 25 items, or 6 screens of 49 items (49 that was the maximum number of items that can fit on the screen; when set size was 49, observer saw a total number of 294 items). Next, observers performed a retest where all encoded items were presented a second time with 300 entirely new distractors. At retest, items were presented in the groups in which they were originally presented (e.g., all the items from screen #5). Order of items within each screen as well as the screen order were randomized. Observers were not told about this structure. Observers saw one test item at a time and made the same old/new and localization responses as in the initial test. They received accuracy feedback on their old/new response and received no feedback on the localization response.

In addition to the object stimuli, run at four set sizes, we also used two sets of stimuli that were expected to produce much poorer memory performance; faces and doors. For these stimuli, observers were tested for 40 screens at a set size of five items per screen, so the total number of 'old' items was 200 rather than 300.

Experiment 2

Experiment 2 was run on-line. It was a replication of the Experiment 1 on-line version. The methods were the same except that the presentation time for each image was increased from 2 second to 3 seconds. The stimuli were object and face images. 300 objects were tested with four different set sizes (5, 15, 25, 49) and 200 faces were tested, using a set size of 5 item/screen. Faces were included in order to test if additional time would reveal SMM for these face stimuli (It did not). Doors were not included. They are only of interest in comparison to faces and it seemed clear that those stimuli did not produce SMM.

Experiment 3

In Experiment 1-2, observers were instructed to attend to the cued image while several images were simultaneously available. Due to the on-line nature of the task, we have no control over where observers were attending during the encoding period. It is possible that they distributed attention unevenly across the available items on screen and/or followed a reading pattern. Experiment 3 aimed to further validate the SMM capacity measured obtained in Experiment 1-2 using a n-back design. In addition to the recognition and spatial memory, Experiment 3 allowed an assessment of observers' temporal memory capacity

In Experiment 3, there was no separate encoding and testing period. Observers saw the same grid array. This time, only one item appeared at a time at a random location for 3 seconds. After an item appeared, Obeservers responded whether they had previously seen the item in the image stream. They clicked on the 'new item' box if the item was new. If the item was old, they clicked on remembered location where the item first appeared. Additionally, observers were prompted to estimate *when* the item first appeared by clicking on a time bar. On this time bar, time zero marked the beginning of the experiment. A green progress bar grew as the experiment proceeded, indicating to observers their current temporal location in the experiment. That is, the bar would be half its maximum length when observers had seen half of the trials. Observers were instructed to click on a time bar location that corresponded to the time when the current ''old'' item had first appeared.

Observers could respond at any time after the image onset and there was no time limit. After all the possible responses were given for each item, observers received visual and auditory feedback on the old/new response accuracy. They received no feedback on the spatial or temporal response. After the feedback, the next item appeared at a new random location. Unbeknownst to the observers, the first 10 item were always unrepeated (new). This kept the distributions of lags between first and second appearances of an item from being too strongly skewed to short lags. After the 10th image, an old image appeared with a 50% probability on each trial. On old item trials, one of the previously shown images was randomly selected without replacement and presented at a new random location. If the next image was new, a new image was presented at a random location. The length of the experiment varied slightly among observers. The experiment ended when observers had seen 150 pairs of items. Observers completed a practice block before the main experiment. The first five items in the practice block were always new. The practice block ended after observers responded to five old items. They could proceed to the main experiment if their old/new decision accuracy was above 80%.

QUANTIFICATION AND STATISTICAL ANALYSIS

Spatial memory capacity measure

To assess spatial memory, we calculated the percentage of clicks/responses that fell within an ROI around the true location of an old object (for Experiment 3, correct location is where the old item first appeared). We defined ROI as a square region centered on the target with x cell(s) on each side. For example, ROI = 1 is the cell containing the original target and ROI = 3 is equivalent to an area extending 3 cells both horizontally and vertically; that is, +/- one cell around the original location. We calculated the percentage of correct clicks for ROIs from 1 to 14, which covers the entire array, even when the target is located in a corner of the 7x7 grid. By definition, this function must rise from 0 to 100% as the ROI diameter is increased from arbitrarily small to very large. At any size, some clicks could land in an ROI by chance. That percentage also rises from 0 to 100%. To calculate this chance function, we paired a target location with the responses to different target items by the same observer to take into account the center bias and any idiosyncratic localization bias (e.g., always clicking in the lower left, if guessing). We repeated this for 100 pairings of target and response for each old target. Subtracting the two functions gave a difference function that provided a conservative estimate of capacity. It is conservative because it always assumes that some responses are lucky guesses. If an observer knew where every item had been and placed all clicks within +/- 1 cell of the correct location, this method would still attribute $\sim$15% of responses to guessing. In our experiments, the difference functions peaked around ROI diameter = 3 to 4 cells for all conditions (see Figure S2).

Temporal memory capacity measure

Time responses ranged from 0 (experiment start) to 100 (experiment end). Temporal ROIs were defined as a percentage of the total time centered on the time when the object first appeared. For example, an ROI = 8 is equivalent to a window +/- 4% of total time around the true time. We calculated the percentage of clicks in ROIs ranging from 0 – 100 in width. As in the spatial analysis, this generates a function rising from 0 to 1. We estimated a guessing function by pairing each old item time location with randomly-generated guessing clicks. Each guess click was drawn from a random time location prior to the onset of the current item. We calculated the percentage of guess clicks that fall into each ROI. We subtracted the guessing function from the correct function to get the temporal memory capacity measure. For the TMM estimates, it is possible for the observer to perform worse than "chance". If an observer gives up on the task and always clicks, for example, on the beginning of the time bar, that will yield worse performance than actual, random guessing.