



Supplementary Information for

Larger images are better remembered during naturalistic encoding

Shaimaa Masarwa, Olga Kreichman, and Sharon Gilaie-Dotan

Sharon Gilaie-Dotan
Email: shagido@gmail.com

This PDF file includes:

Figures S1 to S3
Tables S1 to S5
SI References

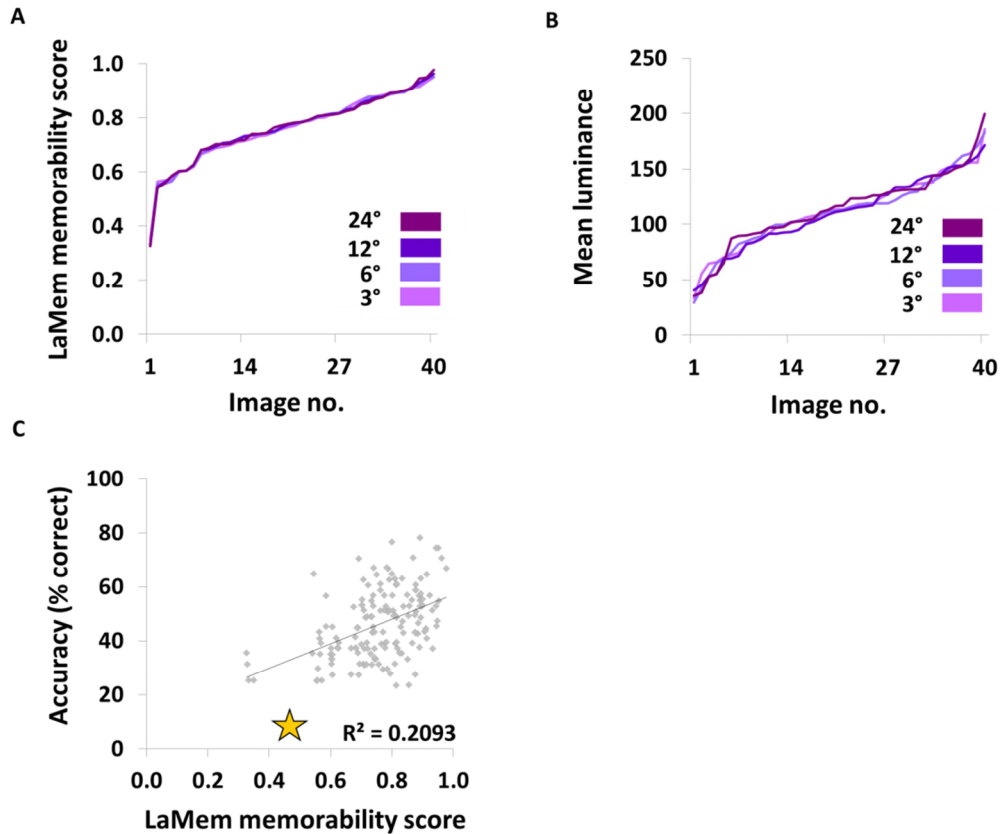


Fig. S1. Experiments 5-7 stimuli – balanced across size conditions and validation. Images in Experiments 5-7 were taken from LaMem Dataset (1) with memorability score for each image. **(A)** We carefully chose the images for each size such that image memorability scores (y axis) would be equally distributed across the experimental size conditions. Here for each size condition, images are presented (on the x axis) in an ascending order according to their LaMem memorability scores (y axis; each size condition indicated by a different shade of purple, see legend). Memorability scores were also equally balanced for each visual category within each size (i.e. for each visual category its 10 images per size had equal memorability scores as those of the other size conditions). **(B)** Image mean luminance levels (y axis) were equally distributed across the experimental size conditions. Here too, images of each size condition are presented (on the x axis) in an ascending order according to their luminance levels (y axis). Each image's luminance level was determined as the mean luminance level over all its pixels after converting it to grayscale (using the `rgb2gray` MATLAB function). **(C)** Per-image accuracy (average across participants, $n=51$) was significantly correlated with LaMem image memorability scores (Experiment 5, $n=160$ images). For this analysis, accuracies of both versions of Experiment 5 were averaged to compute the overall per-image memory performance (same images were used in both versions but presented in a different size, see Table S3). Different versions of Experiment 5 swapped image sets (keeping each image set unchanged) between the size conditions (image set used for 3° in version 1 was used for 24° in version 2 (and vice versa), image set used for 6° in version 1 was used for 12° in version 2 (and vice versa). N.B. In Experiment 6, with a slightly more complicated design, we also made sure memorability scores (as in (A)) were equalized across conditions and that average luminance (as in (B)) was counter-balanced across conditions. In Experiment 7, images of each of the 4 main conditions (3° Blurred, 3° Sharp, 24° Blurred, 24° Sharp) were taken from one of the image sets of Experiment 5 such that these

analyses (balanced memorability scores in (A) and balanced luminance in (B)) will hold for them as well.

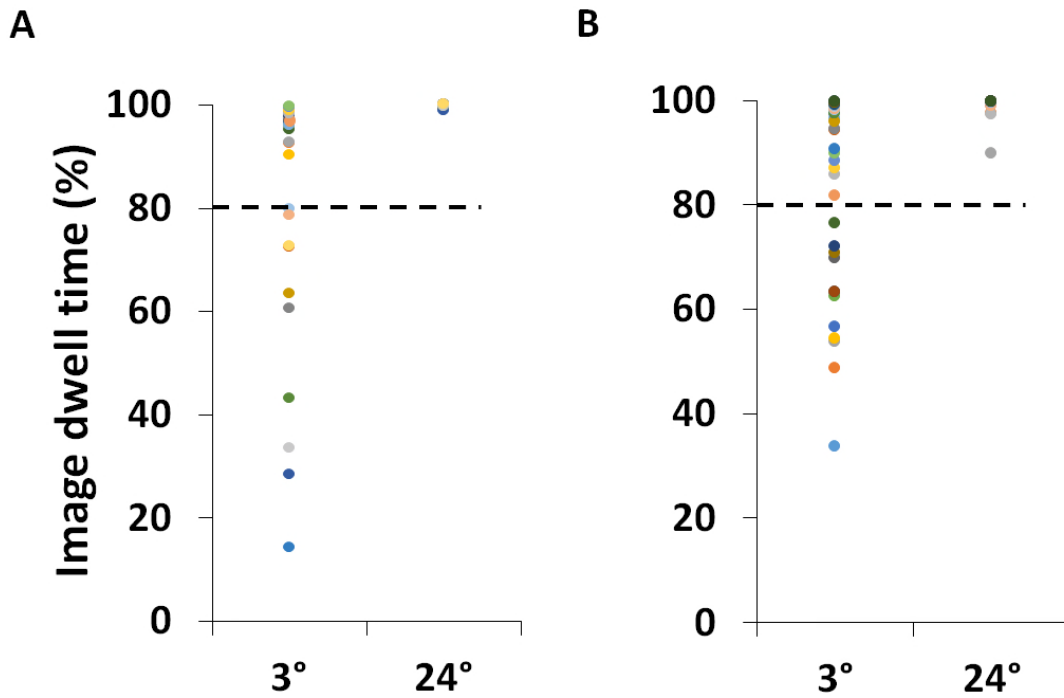


Fig. S2. Eye tracking analysis (Experiments 6 and 7) and participant exclusion based on dwell time on 3° images. (A) Experiment 6: total n=30, 10 participants excluded due to low dwell time on the smaller 3° images condition. **(B)** Experiment 7: total n=35, 12 participants excluded due to low dwell time on the smaller images conditions (3° Blurred, 3° Sharp). Each dot represents one participant. Dashed line represents exclusion cutoff criterion of less than 80% dwell time on 3° images (participants whose data were below the line were excluded). See Methods for more details.

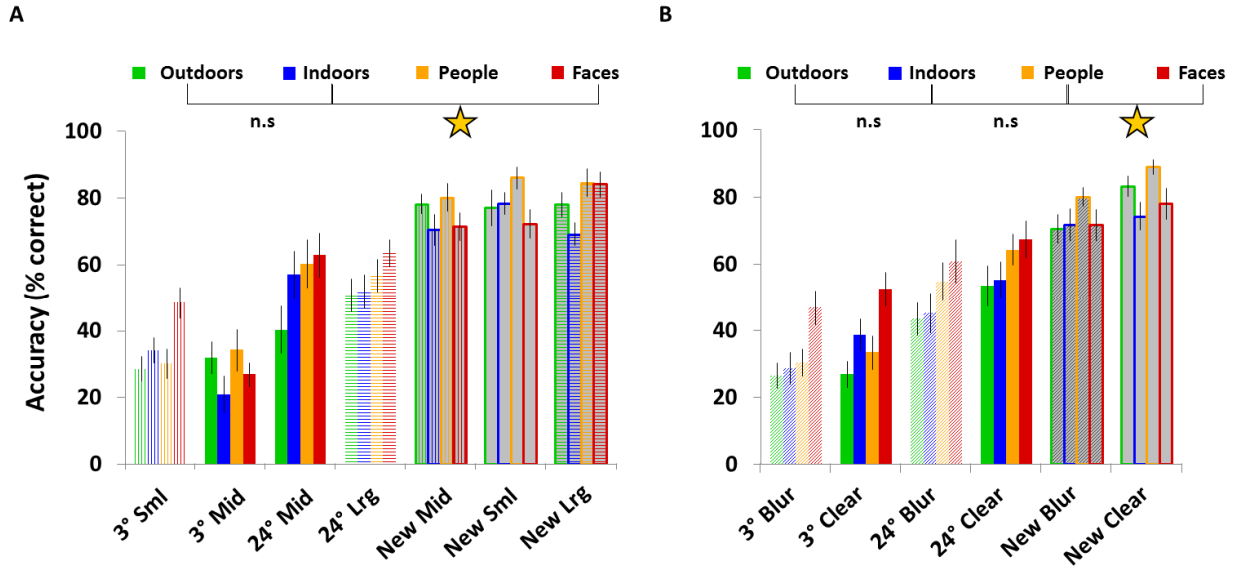


Fig. S3. Significant visual category effects on memory in Experiments 6 and 7. Accuracy results (% correct on the y axis) according to the different conditions (x axis) and visual categories (faces in red, people in orange, indoors in blue, outdoors in green). 2-way ANOVA analyses of data of Experiments 6 (**(A)**, $n=18$) and 7 (**(B)**, $n=23$) with factors of condition (x axis) and category revealed significant effect of condition (Exp. 6: $F(3,51)=27$, $p<0.0001$; Exp. 7: $F(3,66)=37.36$, $p<0.0001$) and of category (Exp. 6: $F(3,51)=7.15$, $p=0.0004$; Exp. 7: $F(3,66)=12.51$, $p<0.0001$) on memory during naturalistic encoding (in Exp. 6 significant interaction between condition and category $F(9,153)=3.539$, $p=0.0005$, in Experiment 7 no interaction was found $F(9,198)=1.685$, $p=0.0947$). In both experiments faces were best remembered (in Experiment 6 significantly better than indoors (and outdoors) and in Experiment 7 significantly better than people (and the other categories)). Note that the New condition(s) appearing in the graphs were not included in the statistical analyses.

Table S1. Experiments 1-7 accuracy performances at test according to size at exposure

Visual angle at exposure		3°	6°	12°	21° / 24°	New				
Exp. 1 (n = 17)		44.63 ± 4.35			62.86 ± 3.65	81.02 ± 5.26				
Exp. 2 (n = 16)		38.67 ± 4.05			56.38 ± 5.05	82.81 ± 3.32				
Exp. 3 (n = 17)		53.24 ± 3.85			63.53 ± 3.27	80 ± 3.68				
Exp. 4 (n = 16)		Sml			61.70 ± 4.94	79.78 ± 4.25				
		Sml (highRes)								
Exp. 5 (n = 51)		All categories	33.1 ± 2.27	42.9 ± 2.48	51.96 ± 2.93	56.6 ± 2.49	81.6 ± 1.8			
		Faces	40.6 ± 3.46	54.1 ± 3.71	58.63 ± 3.74	67.25 ± 3.18	81.23 ± 2.18			
		People	31.4 ± 3.52	42.35 ± 2.93	58.43 ± 3.63	59.61 ± 3.49	86.67 ± 1.92			
		Indoors	35.09 ± 3.24	40.59 ± 3.23	50 ± 4.09	55.69 ± 3.38	74.56 ± 2.58			
		Outdoors	25.29 ± 3.47	34.7 ± 3.8	40.78 ± 3.48	43.92 ± 3.77	83.82 ± 2.4			
Exp. 6 (n = 18)		3° Sml	3° Mid			24° Mid	24° Lrg	New Sml	New Mid	New Lrg
		35.3 ± 3.04	28.47 ± 3.66			55.1 ± 6.15	55.67 ± 4.11	75.05 ± 3.22	78.5 ± 2.95	78.97 ± 2.82
Exp. 7 (n = 23)		Blurred	Sharp			Blurred	Sharp	Blurred		Sharp
		33.15 ± 3.64	37.93 ± 3.62			51.09 ± 4.77	60.11 ± 4.5	73.59 ± 3.56		81.2 ± 2.71

For each experiment and each condition mean accuracy (% correct) at test across participants ± SEM. Columns represent the size at exposure; the “New” column represents images not seen in exposure as additional control. All images at test apart from Experiment 6 were at fixed size of 8°. Exp. 4: Sml is the regular 3° condition as in the other experiments, Sml (highRes) is the condition that was seen from 4m and occupied 3°. Exp. 6: 3° Sml - images were presented as small (3°) at both exposure and test, 24° Lrg – images were presented as large (24°) at both exposure and test, 3° Mid and 24° Mid – as in the other experiments, tested with mid-sized images, New Sml, New Mid, New Lrg – new (unseen) images at test presented at different sizes. Exp. 7 – Blurred and Sharp refer to sharpness level of images, images that were shown as blurred in exposure were also shown as blurred at test and so goes for the sharp images. See Methods for more details.

Table S2. Experiments 1-7 accuracy statistical analyses

		T-test				
		Contrast	Results			
Exp. 1 (n = 17)		3°, 21°	t(15) = 4.732	p = 0.000267		
Exp. 2 (n = 16)		3°, 21°	t(14) = 6.84	p < 10⁻⁵		
Exp. 3 (n = 17)		3°, 21°	t(15) = 4.043	p = 0.001063		
		ANOVA				
		Factors	Main effects	Interaction	Post-hoc (Bonferroni/Dunn)	
Exp. 4 (n = 16)		Condition Sml, Sml(highRes) Sml(highRes), 21°	F(3,45) = 21.13 p < 0.0001		p=0.425 p=0.0009	
Exp. 5 (n = 51)	Version X Size	Version Version 1, Version 2	F(1,196) = 0.21 p = 0.646	F(3,196)=0.12 p=0.94		
		Size 3°, 6°, 12°, 24°	F(3,196) = 16.55 p < 0.0001			
	Size X Category	Size 3°, 6° 6°, 12° 12°, 24°	F(3,150) = 57.31 p < 0.0001	F(9,450)=1.84 p=0.0593		p < 0.0001 p < 0.0001 p = 0.0177
		Category Faces, People Faces, Indoors Indoor, Outdoors	F(3,150) = 15.76 p < 0.0001			p = 0.0101 p = 0.0004 p = 0.0018
Exp. 6 (n = 18)	Size (at exposure) X Size (at test)	Size (at exposure) 3°, 24°	F(1,17)=40.032 p<0.0001	F(1,17)=2.93 p=0.1047		
		Size (at test) 8°, As in exposure	F(1,17)=3.206 p=0.0912			
		Condition 3° Sml, 3° Mid (*) 3° Sml, 24° Mid 24° Mid, 24° Lrg (*)	F(3,51)=26.996 p<0.0001		p = 0.0775 p < 0.0001 p = 0.8795	
Exp. 7 (n = 23)	Size X Sharpness level	Size 3°, 24°	F(1,22)=89.270 p<0.0001	F(1,22)=1.22 p=0.2814		
		Sharpness level Blurred, Sharp	F(1,22)=11.999 p=0.0022			
		Condition 3° Blurred, 3° Sharp 3° Sharp, 24° Blurred (*) 24° Blurred, 24° Sharp	F(3,66)=37.364 p<0.0001		p = 0.0977 p < 0.0001 p = 0.0023	

For each experiment the statistical analyses comparing performance at test according to the experimental design are presented. Experiments 1-3 included only 2 conditions and therefore 2-tailed paired t-test results are presented, Experiment 4 included 3 conditions thus 1-way ANOVA results are presented, Experiment 5 first presents version X size 2-way ANOVA to examine the possibility of a version effect, and then a 2-way ANOVA to examine size and category effects across the 2 experimental versions (data collapsed across versions since there was no effect of version). While Experiments 1-5 and 7 examined the effects of image size at exposure only (size at test was constant at ~8°), Experiment 6 examined the possibility of an interaction between size

at exposure (“Size(at exposure)”) and size at test (“Size(at test)”). This was examined both in a 2-way ANOVA testing each factor, and also in a 1-way ANOVA testing directly the influence of same size at test and exposure for the smaller images (3°) and the larger ones (24°) relative to different size at test (3° Sml vs 3° Mid, 24° Mid vs 24° Lrg, both contrasts of interest are marked with an asterisk). Experiment 7 tested whether size influences image memory even when there was no added information. The main contrast of interest between 3° Sharp and 24° Blurred conditions that had the same amount of information (see Methods) is indicated by (*) in the 1-way ANOVA at the bottom. We also examined if sharpness level and size as factors influenced performance in a 2-way ANOVA above. Significant results (including Bonferroni/Dunn post-hoc corrected for multiple comparisons) appear in bold. See Methods for more details. Additional notations as in Table S1.

Table S3. Experiment 5 per-image analysis (n=160 images)

Per image memorability (Exp. 5) to memorability scores (LaMem)		Correlation		
Exp. 5 (n = 160)	Ver. 1	r = 0.389	p < 0.0001	t(158) = 5.309
	Ver. 2	r = 0.276	p = 0.0003	t(158) = 3.622
	Mean (Ver. 1, Ver. 2)	r = 0.457	p < 0.0001	t(158) = 6.467

For each image we calculated its average memorability in each version of Experiment 5 across all participants and compared these results to the LaMem image memorability scores (1). Since each image appeared in the 2 versions in 2 different sizes (either 3° and 24°, or 6° and 12°), for the averaging across the versions (line 3), for each image we took the average memorability (mean across the 2 versions) as an approximation for its memorability for an average sized image (see Figure S1 panel C). In each row: correlation coefficient (r) and its corresponding p and t values (significant results in bold).

Table S4. Experiments 1-7 reaction time performances

Visual angle at exposure		3°	6°	12°	21° / 24°	New			
Exp. 1 (n = 17)		1041.62 ± 53.48			984.5 ± 41.96	947.647 ± 33.33			
Exp. 2 (n = 16)		1056.53 ± 53.80			1005.72 ± 50.96	1005 ± 62.99			
Exp. 3 (n = 17)		954.56 ± 41.58			934.412 ± 40.69	943.12 ± 38.39			
Exp. 4 (n = 16)		Sml			1071.78 ± 61.36	1112.84 ± 66.97			
		Sml (highRes)							
Exp. 5 (n = 51)	All categories	1055.16 ± 37.02	1041.8 ± 34.11	1003.8 ± 30.83	1034.1 ± 36.31	1047.2 ± 33.16			
	Faces	1049 ± 39.93	1000.4 ± 38.21	1007.3 ± 60.38	1030.3 ± 48.83	1030.6 ± 32.53			
	People	1043.5 ± 41.14	1078.2 ± 54.69	1005.9 ± 46.83	1017.3 ± 30.64	1023.6 ± 32.92			
	Indoors	1053.6 ± 46.61	1015.7 ± 38.71	1014.2 ± 33.98	1048.8 ± 43.64	1078.7 ± 36.33			
	Outdoors	1074.6 ± 53.08	1073.1 ± 48.91	987.98 ± 33.72	1040.2 ± 44.78	1055.7 ± 44.28			
Exp. 6 (n = 18)		3° Sml			24° Mid	24° Lrg	New Sml	New Mid	New Lrg
		3° Mid			1020.41 ± 62.9	1006.03 ± 53.16	1079.62 ± 75.86	1001.17 ± 61.82	1093.25 ± 59.94
Exp. 7 (n = 23)		Blurred			Blurred	Sharp	Blurred		Sharp
		Sharp			1147.7 ± 63.32	1176 ± 74.96	1178.1 ± 65.2		1171.4 ± 61.8

For each experiment and each condition mean RTs across participants ± SEM (data in ms). Notations as in Tables S1, S2. We did not find any explicit or implicit priming effects (see Table S5).

Table S5. Experiments 1-7 reaction time (at test) statistical analyses

		T-test		
		Contrasts	Results	
Exp. 1 (n = 17)		3°, 21°	t(15) = 2.380 p = 0.031	
Exp. 2 (n = 16)		3°, 21°	t(14) = 2.186 p = 0.046	
Exp. 3 (n = 17)		3°, 21°	t(15) = 1.025 p = 0.321	
		ANOVA		
		Factors	Main effects	Interaction
Exp. 4 (n = 16)		Condition	F(3,45) = 1.321 p = 0.279	
Exp. 5 (n = 51)	Size X Category	Size	F(3,150) = 1.859 p = 0.139	F(9,450) = 0.67 p = 0.7365
		Category	F(3,150) = 0.212 p = 0.8883	
Exp. 6 (n = 18)	Size (at exposure) X Size (at test)	Size (at exposure)	F(1,68) = 0.400 p = 0.529	F(1,68) = 0.646 p = 0.424
		Size (at test)	F(1,68) = 0.346 p = 0.558	
Exp. 7 (n = 23)	Size X Sharpness level	Size	F(1,88) = 0.014 p = 0.905	F(1,88) = 0.028 p = 0.867
		Sharpness level	F(1,88) = 0.060 p = 0.807	

Analyses were done on mean RTs. Significant effects are indicated in bold. Note that in the 2 significant findings (Experiments 1 and 2), the larger images showed faster RTs. Further analyses to examine possible priming effects (2) ((i) whether RTs of old images were faster than RTs of new images, and (ii) whether RTs of correctly remembered old images were faster than RTs of non-remembered old images) did not yield any significant findings in any of the experiments. Additional notations as in Tables S1, S2, and S4.

SI References

1. A. Khosla, A. S. Raju, A. Torralba, A. Oliva, Understanding and predicting image memorability at a large scale. *Proc. IEEE Int. Conf. Comput. Vis.* 2015 Inter, 2390–2398 (2015).
2. E. Tulving, D. Schacter, Priming and human memory systems. *Science*. 247, 301–306 (1990).