Supplementary Material for “Overt attentional correlates of memorability of scene images and their relationships to scene semantics” by Lyu, Choe, Kardan, Kotabe, Henderson, & Berman.

Supplemental Note 1: Experimental details of the Edinburgh dataset

Two groups of 36 undergraduate students (Group 1 and Group 2) from the University of Edinburgh participated in the experiment. All 72 participants had 20/20 corrected or uncorrected vision, were naive to the purposes of the experiment and provided informed consent as administered by the Institutional Review Board of the University of Edinburgh.

Participants sat 90 cm away from a 21-inch CRT monitor and placed their head on a chin and forehead rest. The scenes were displayed fullscreen in their native resolution and subtended 25.8° x 19.4° in visual angle. Eye movements were recorded from the right eye, although viewing was binocular, via an SR Research (Ottawa, Ontario, Canada) Eyelink 1000 eye tracker with a sampling rate of 1000 Hz. The experiment was controlled with SR Research Experiment Builder software. The eye tracker was calibrated using a built-in nine-point calibration routine. The calibration was not accepted until the average error was less than 0.49° and the maximum error was less than 0.99°.

Both Group 1 (G1) and Group 2 (G2) participants performed the first phase (Encoding phase) of the experiment in which the viewing task was manipulated (i.e., visual search task, memorization task or aesthetic preference task). Shortly after the Encoding phase only G1 participants engaged in the second phase testing scene recognition (Test phase). So, the fixation patterns during memorization were obtained from G1 and G2, and the recognition accuracy from memorization encoding was obtained from G1 participants only. In the Encoding phase, 135 full-color (32 bit) 800 x 600-pixel photographs of real-world, indoor and outdoor scenes were presented. The 135 scenes were split into three blocks of 45, and the scene split was the same across participants. During each block, participants were instructed to perform one of three tasks on the scenes presented for 8 s while their eye movements were recorded: (1) memorize the scene for a subsequent old/new recognition test (but no response was required during the encoding period), (2) search for an object, or (3) make an aesthetic preference judgment. The task assignment and order of each block were determined by a dual-Latin square design and counterbalanced across participants. Within each task block, the scenes were presented in random order. The participants completed all three blocks. In this paper, we focus on the intentional scene memorization encoding task and resulting memory performance. For more details and results from the visual search and preference judgment tasks, see Choe et al. (2017).

After completing all three task blocks and a short break, G1 participants engaged in the Test phase, i.e., the scene recognition task. Before the task, participants were informed that their memory would be tested for all of the scenes they had previously encountered, not just the scenes they had been instructed to remember in the memorization block. In each trial, a scene was shown for 3 s, and participants were asked to identify whether the scene was 'old' (encountered in the Encoding phase during *any* block, not just the memorization block, and presented in an identical form), 'altered' (encountered in the Encoding phase but presented in a horizontally-mirrored form), or 'new'. In total, 66 of those scenes were 'old', i.e., seen in the encoding phase, and the other 66 scenes were 'altered' stimuli, and the remaining 3 scenes were not used in the Test phase. In addition, 22 new scenes were never seen before. In this paper, we present the results of the 132 scenes that were used in both the Encoding and Test phases. A total of 154 scenes, consisting of seven categories that contained 22 scenes – old & with memory encoding, old & search encoding, old & preference encoding, altered & memory encoding, altered & search encoding, altered & preference encoding, and new – were used in the recognition task. The recognition accuracy used in this study was based solely on the old & memory and altered & memory trials only (44 trials per participant).

Supplemental Note 2: Comparisons between the Edinburgh and FIGRIM datasets

Fixation map consistency was significantly associated with recognition accuracy in both the Edinburgh and FIGRIM datasets, but fixation counts were only significant in the FIGRIM dataset. Why do these results differ? There are several differences between those two datasets, such as the scene stimuli, experimental paradigms, and participants. However, one notable difference was the viewing duration: 8 s in the Edinburgh dataset vs. 2 s in the FIGRIM dataset. Therefore, we examined the effects of viewing duration of fixation map consistency, fixation count, and their relationship to recognition accuracy. Specifically, we analyzed the Edinburgh data by varying the analysis duration from 1 s to the full 8 s of viewing in 1 s increments (the filled circles in Fig. S2). We then compared the Edinburgh results at 2 s (i.e., the fixations within the first 2 s were analyzed) with the FIGRIM results (the filled stars in Fig. S2).

Fig. S2a illustrates that the Edinburgh fixation counts linearly increased with viewing duration, as expected, and the Edinburgh fixation counts at 2 s was not significantly different to the FIGRIM fixation count. Fig. S2b illustrates that the Edinburgh fixation map consistency was relatively stable across viewing duration, and the Edinburgh fixation map consistency at 2 s was not significantly different from the FIGRIM fixation map consistency. These results suggest that the calculation of fixation counts and fixation map consistency was robust across the two datasets despite their many differences.

Fig. S2c illustrates that the correlation between fixation counts and fixation map consistency in the Edinburgh dataset were not significantly different across the different viewing durations. In the FIGRIM dataset, the correlation between fixation counts and fixation map consistency was significantly negative, $\rho(628) = -0.13$, 95% CI [−0.2, −0.07], $p < .001$, but it was within the 95% confidence interval of the Edinburgh correlation at 2 s, $\rho(130) = 0.02$, 95% CI [−0.12, 0.17], $p = .783$.

Fig. S2d illustrates that the correlation values between fixation map consistency and recognition accuracy were significantly positive at 2 s and afterward in the Edinburgh dataset. In particular, the maximum correlation was found at 2 s, $\rho(130) = 0.34$, 95% CI [0.2, 0.46], $p < .001$ (the right panel in Fig. S2d). The FIGRIM correlation value, $\rho(628) = 0.21$, 95% CI [0.15, 0.27], $p < .001$ (the left panel in Fig. 3d), was within the 95% confidence interval of the Edinburgh correlation at 2 s. These results suggest that the fixation map consistency calculated with 2 s of fixation data is reliably associated with scene memorability.

Fig. S2e illustrates that the correlation values between fixation counts and recognition accuracy were positive over time in the Edinburgh dataset; the correlation value at 1 s was significantly positive, $\rho(130) = 0.23$, 95% CI [0.09, 0.36], $p = .007$. Also, the FIGRIM correlation value, $\rho(628) = 0.18$, 95% CI [0.12, 0.25], $p < .001$ (the right panel in Fig. 3d) was within the 95% confidence interval of the Edinburgh correlation at 2 s. The comparable range of correlation values suggests that fixation counts and recognition accuracy are positively and weakly correlated and that a large number of scenes (e.g., 630 scenes in the FIGRIM dataset) is required to detect a significant relationship.

Supplemental Note 3: Examination of scene category

Scene category has been shown to be associated with scene memorability (Bylinskii et al., 2015; Isola et al., 2011), so we examined the effects of scene category on scene memory. To do so, we conducted a scene-level linear regression analysis on the FIGRIM dataset, in which the dependent variable was recognition accuracy from the AMT participants, and the predictors were z-scored fixation map consistency and fixation counts from the lab participants, scene category, the interaction of fixation map consistency and scene category, and the interaction of fixation counts and scene category, i.e., scene memorability ~ scene category * z-scored fixation map consistency * z-scored fixation count. This model explained (df = 546) explained 16.1% of the variance (adjusted R^2) and showed a significant effect of scene category, $F(20,546) = 4.23$, $p < .001$, consistent with the FIGRIM result that some scene categories, such as amusement park and playground, were more memorable than others, such as cockpit and highway (Bylinskii et al., 2015). But, this model showed nonsignificant three-way interactions between scene category, fixation map consistency, and fixation count, $F(20,546) = 0.82$, $p = .686$, nonsignificant interaction between scene category and fixation map consistency, $F(20,546) = 0.42$, $p = .987$, and nonsignificant interaction between scene category and fixation count, $F(20,546) = 0.97$, $p = .503$, suggesting that scene category did not significantly affect the relationships between fixation map consistency and scene memorability and between fixation counts and scene memorability. Fig. S4 illustrates the relationship between fixation map consistency and scene memory for each scene category type. In 19 out of 21 scene categories in the FIGRIM dataset, fixation map consistency and scene memory were positively associated, with the Spearman correlation values ranging from −0.01 to 0.51 (M=0.21, SD=0.15).

Supplemental Note 4: Examination of the interactions between predictors

To examine the relationships between predictors of our models, we further tested the interaction effects of all variables by including interaction terms in the regression model. The analysis code is available at model_information.R on OSF. In the Edinburgh dataset, we added a two-way interaction term, fixation map consistency*fixation count, to the E_{Base} model. This model explained (df = 127) explained 24.4% of the variance (adjusted R^2), with a nonsignificant two-way interaction between fixation map consistency and fixation count, $F(1,127) = 1.06$, p

= .305, suggesting that there was no interaction effect between fixation map consistency and fixation count.

In the FIGRIM dataset, we added a four-way interaction term, fixation map consistency*fixation count*MOPS*face/human, to the F_{Base} model. This model explained (*df* = 593) explained 30% of the variance (adjusted R^2), with a non-significant four-way interaction between fixation map consistency, fixation count, MOPS, and face/human, $F(1,593) = 0.02$, $p = .879$, non-significant three-way interactions (between fixation map consistency, fixation count, and MOPS, $F(1,593) = 0.28$, $p = .596$, between fixation map consistency, fixation count, and face/human, $F(1,593) = 0.30$, $p = .582$, between fixation map consistency, MOPS, and face/human, $F(1,593) = 3.62$, $p = .058$, and between fixation count, MOPS, and face/human, $F(1,593) = 2.61$, $p = .107$), and non-significant two-way interactions (between fixation map consistency and fixation count, $F(1,593) = 0.86$, $p = .354$, between fixation map consistency and MOPS, $F(1,593) = 1.24$, $p = .266$, between fixation counts and MOPS, $F(1,593) = 0.48$, $p = .489$, between fixation map consistency and face/human, $F(1,593) = 2.41$, $p = .121$, between fixation counts and face/human, $F(1,593) = 1.07$, $p = .300$, and between MOPS and face/human, $F(1,593) = 1.52$, $p = .218$). Together, the results showed that there was no interaction effect between these four predictors.

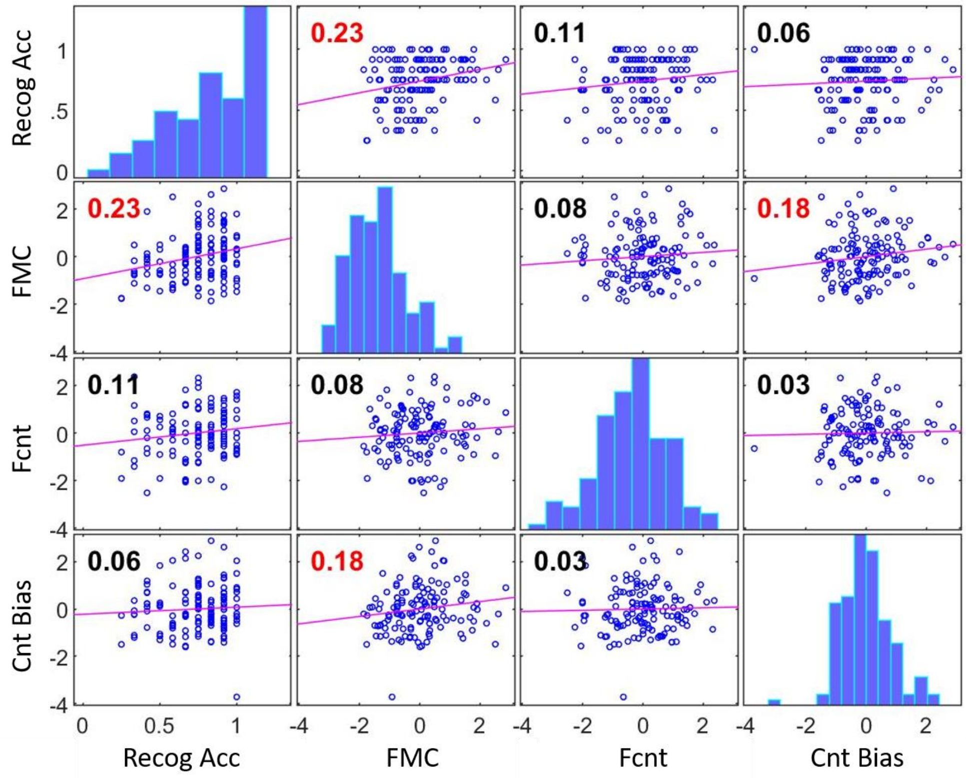

Fig. S1. Relationships between scene memory, fixation map consistency, fixation count, and center bias in the Edinburgh dataset. The names of the variables are abbreviated as Recog Acc, FMC, Fcnt, and Cnt Bias, respectively. This matrix of plots show the correlations among all variables. Histograms of the variables present along the diagonal with scatterplots of the variables off the diagonal. The slopes of the lines in the scatterplots are equivalent to their corresponding correlation coefficients, which are denoted in red if they are significantly different from zero and in black if they are not significantly different from zero.

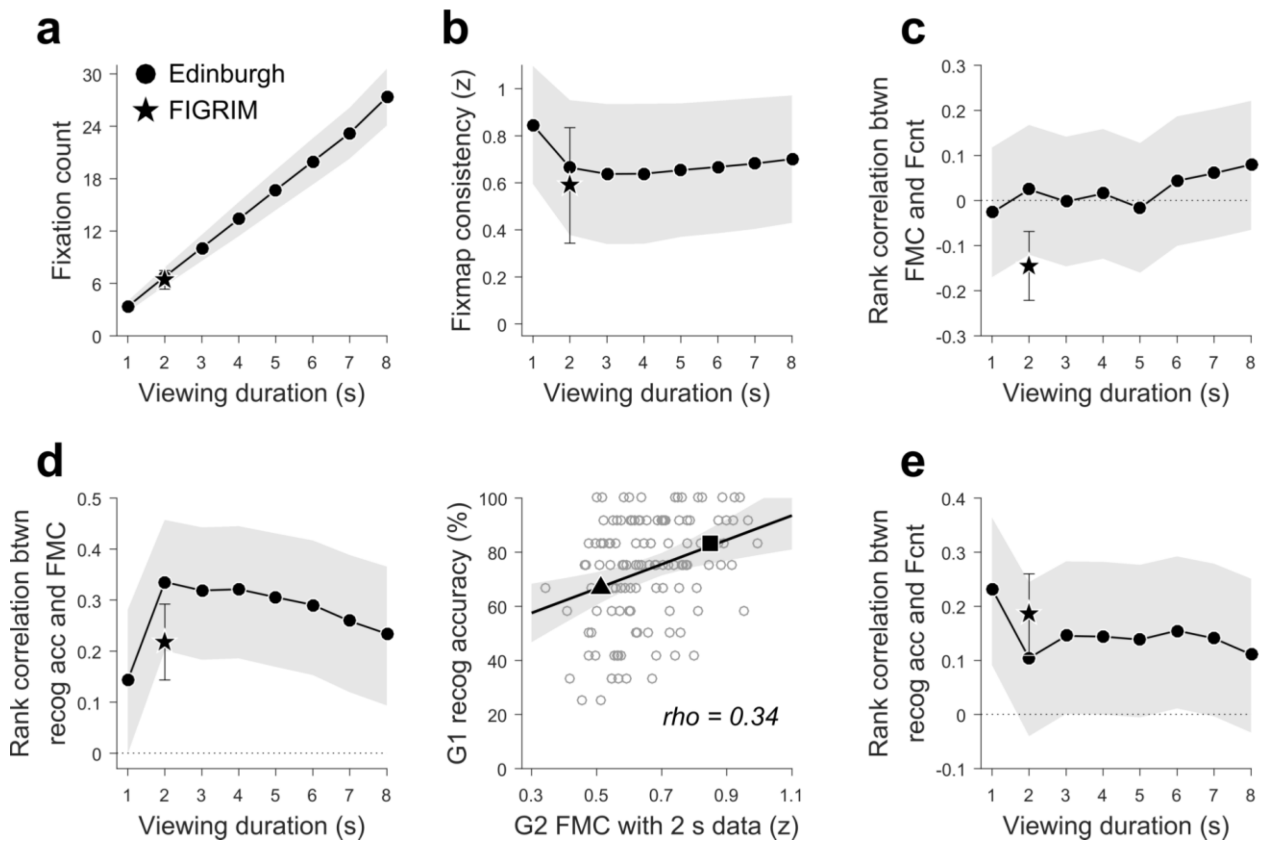

Fig. S2. Relationships between fixation map consistency (FMC), fixation counts (Fcnt), and recognition accuracy (recog acc) with respect to viewing duration. The viewing duration of 4 s indicates that the first 4 s (from the trial onset) of the fixation data were used to calculate fixation map consistency and fixation count. (**a**) fixation counts over viewing duration. The filled circles represent the Edinburgh results, the filled stars represent the FIGRIM results, the gray shades and error bars represent the 95% confidence interval. (**b**) Fixation map consistency over viewing duration. (**c**) Correlations between fixation map consistency and fixation counts over viewing duration. (**d**) The left panel shows correlations between fixation map consistency and recognition accuracy over viewing duration. The right panel illustrates the Edinburgh results at 2 s. The filled square and triangles indicate the scenes presented in Fig. 1. (**e**) Correlations between fixation counts and recognition accuracy over viewing duration.

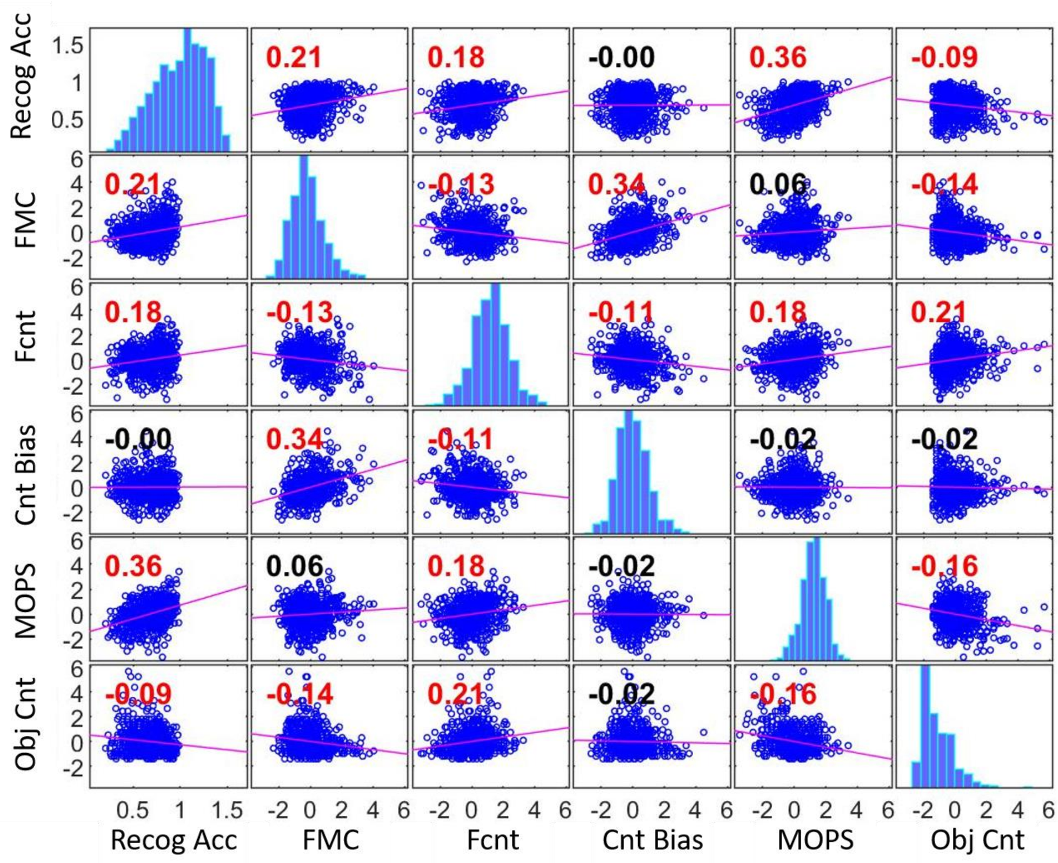

Fig. S3. Relationships between scene memory, fixation map consistency, fixation count, center bias, MOPS, and object counts in the FIGRIM dataset. The names of the variables are abbreviated as Recog Acc, FMC, Fcnt, Cnt Bias, MOPS, and Obj Cnt, respectively. This matrix of plots show the correlations among all variables. Histograms of the variables present along the diagonal with scatterplots of the variables off the diagonal. The slopes of the lines in the scatterplots are equivalent to their corresponding correlation coefficients, which are denoted in red if they are significantly different from zero and in black if they are not significantly different from zero.

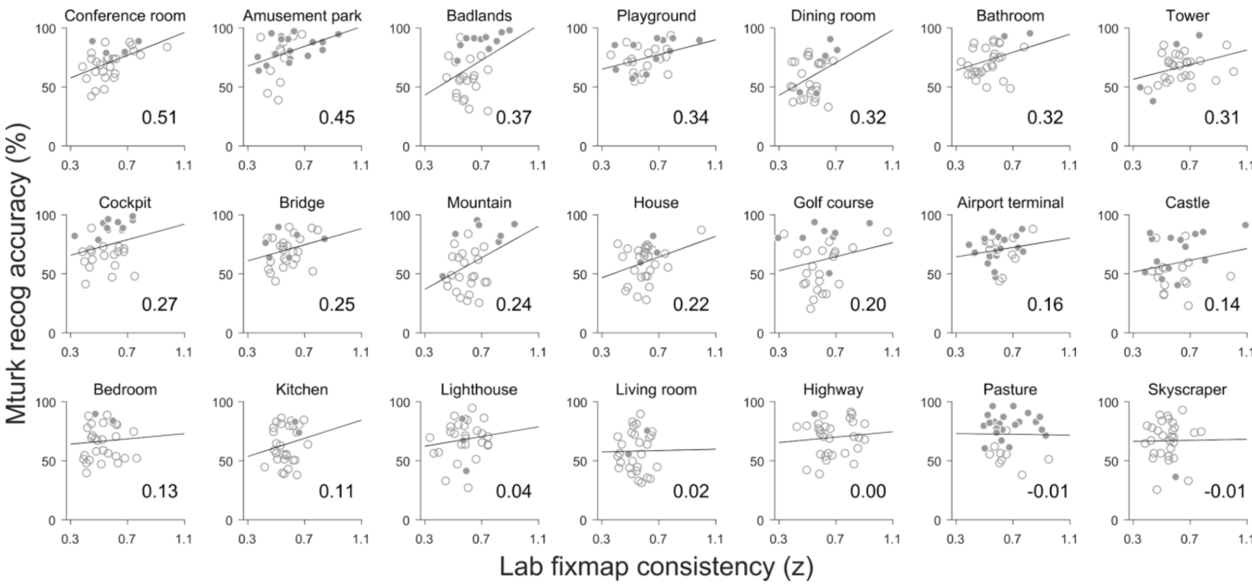

Fig. S4. Relationships between fixation map consistency, scene category, scene semantics, and scene memory. The effect of scene category in the FIGRIM dataset. The name of each category is shown at the top of each panel. A filled circle represents a scene with face/human, and an open circle represents a scene without face/human. The line represents a linear regression.

Table S1. Model comparisons in the Edinburgh dataset.

Model name	*df*	Adjusted R^2	AIC	BIC	Compare to	F value	*p* value
E_{Base}	130	0.17	−75.8	−67.2	–	–	–
E_{Fcnt}	129	0.18	−76.7	−65.2	E_{Base}	3.10	.081
E_{FMC}	129	0.24	−86.4	−74.9	E_{Base}	12.95	<.001
E_{Both}	128	0.24	−86.6	−72.2	E_{FMC}	2.09	.151

The analysis code is available at model_information.R on OSF.

Table S2. Model comparisons in the FIGRIM dataset.

Model name	*df*	Adjusted R^2	AIC	BIC	Compare to	F value	*p* value
F_{Base}	608	0.11	−511	−409	–	–	–
F_{Fcnt}	607	0.14	−531	−424	F_{Base}	22.99	<.001
F_{FMC}	607	0.16	−543	−436	F_{Base}	33.79	<.001
F_{Both}	606	0.20	−573	−462	F_{FMC}	31.60	<.001
F_{Both} + MOPS	605	0.23	−595	−479	F_{Both}	23.16	<.001
F_{Both} + face/human	605	0.29	−651	−536	F_{Both}	82.16	<.001
F_{Both} + face/human + motion	604	0.29	−650	−530	F_{Both} + face/human	0.30	.584
F_{Both} + MOPS, face/human	604	0.30	−654	−534	F_{Both} + face/human	4.42	.036

The analysis code is available at model_information.R on OSF.

Table S3. Top 20 objects/features in the descending order of mean difference in scene memorability. Manually defined features are underlined.

Object name	Number of scenes with object	Average memorability with object	Average memorability without object	t-test p-value (uncorrected)	Corrected p-value
face/human	158	0.78	0.64	.000	.000
person	101	0.74	0.66	.000	.000
pilot	10	0.90	0.67	.000	.000
motion	119	0.74	0.66	.000	.000
plant	71	0.59	0.68	.000	.001
cloud	17	0.51	0.68	.000	.034
playground	24	0.76	0.67	.001	.081
decor	26	0.55	0.68	.001	.121
child	11	0.83	0.67	.001	.127
bush	27	0.56	0.68	.001	.127
chair	97	0.62	0.68	.002	.163
roller.coaster	10	0.76	0.67	.002	.186
chandelier	24	0.57	0.67	.003	.300
house	34	0.60	0.68	.003	.302
table	101	0.63	0.68	.003	.303
door	97	0.63	0.68	.004	.377
frame	40	0.59	0.68	.005	.396
cushion	17	0.55	0.67	.006	.523
lamp	26	0.58	0.67	.006	.532
mountain	50	0.59	0.68	.007	.554