nature human behaviour

Article

https://doi.org/10.1038/s41562-024-01870-3

Images with harder-to-reconstruct visual representations leave stronger memory traces

In the format provided by the authors and unedited

Effect	DFn	DFd	F	p	η^2
Distinctiveness (Dist.)	1	44	197.187	$< .001$	0.353
Reconstruction error (RE)	1	44	27.408	$< .001$	0.029
Time	2	88	106.498	$< .001$	0.285
Dist. $\times$ RE	1	44	0.160	.691	0.000
Dist. $\times$ Time	2	88	4.414	.015	0.001
RE $\times$ Time	2	88	4.192	.018	0.001
Dist. $\times$RE $\times$ Time	2	88	3.909	.024	0.001

Table S1: To formally test how distinctiveness and reconstruction error influence the effect of encoding time on memory performance, we ran a three-way repeated-measures ANOVA with the dependent variable being hit rate and the three factors being distinctiveness (Dist), reconstruction error (RE) and encoding duration (Time). The table shows the results. As expected, there were main effects of distinctiveness, reconstruction error and encoding duration. More critically, we also saw an interaction between reconstruction error and encoding duration, meaning that images with large reconstruction error benefited even more from longer encoding times, relative to images with smaller reconstruction error. Interestingly, we also observed an interaction between distinctiveness and encoding duration and a three-way interaction between distinctiveness, reconstruction error and encoding duration. These interactions can arise from the differential time-courses of the memory benefit (see Fig. 5C) brought by large reconstruction error as a function of the distinctiveness levels.

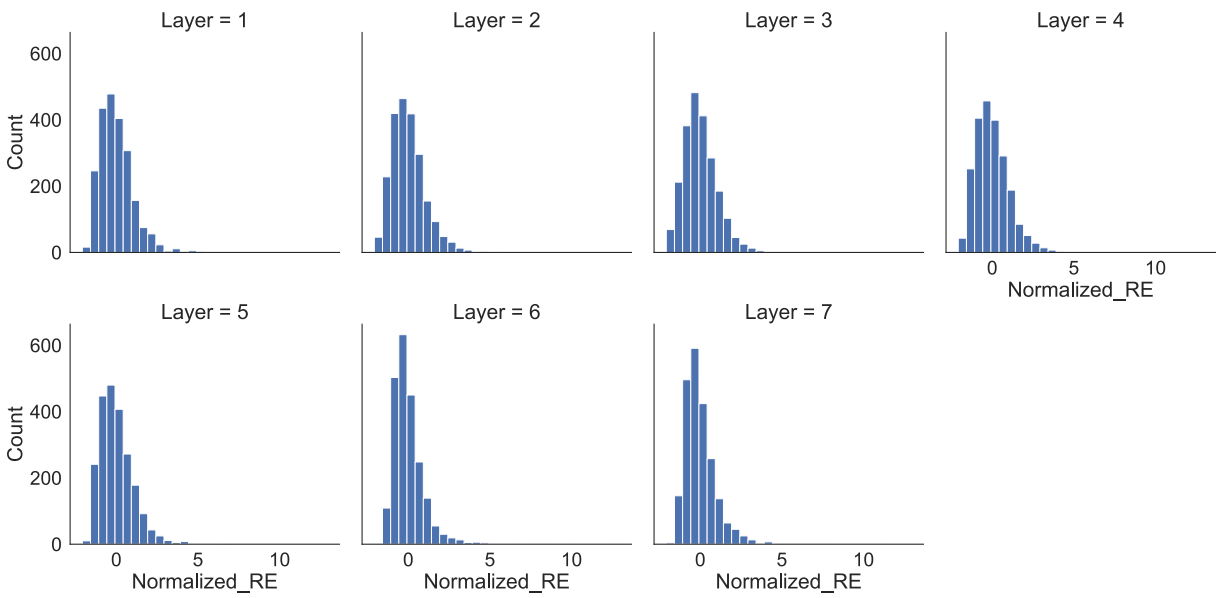

Figure S1: The distributions of normalized reconstruction errors for the 2221 target images from the seven sparse coding models (each trained to reconstruct a different layer of VGG-16).

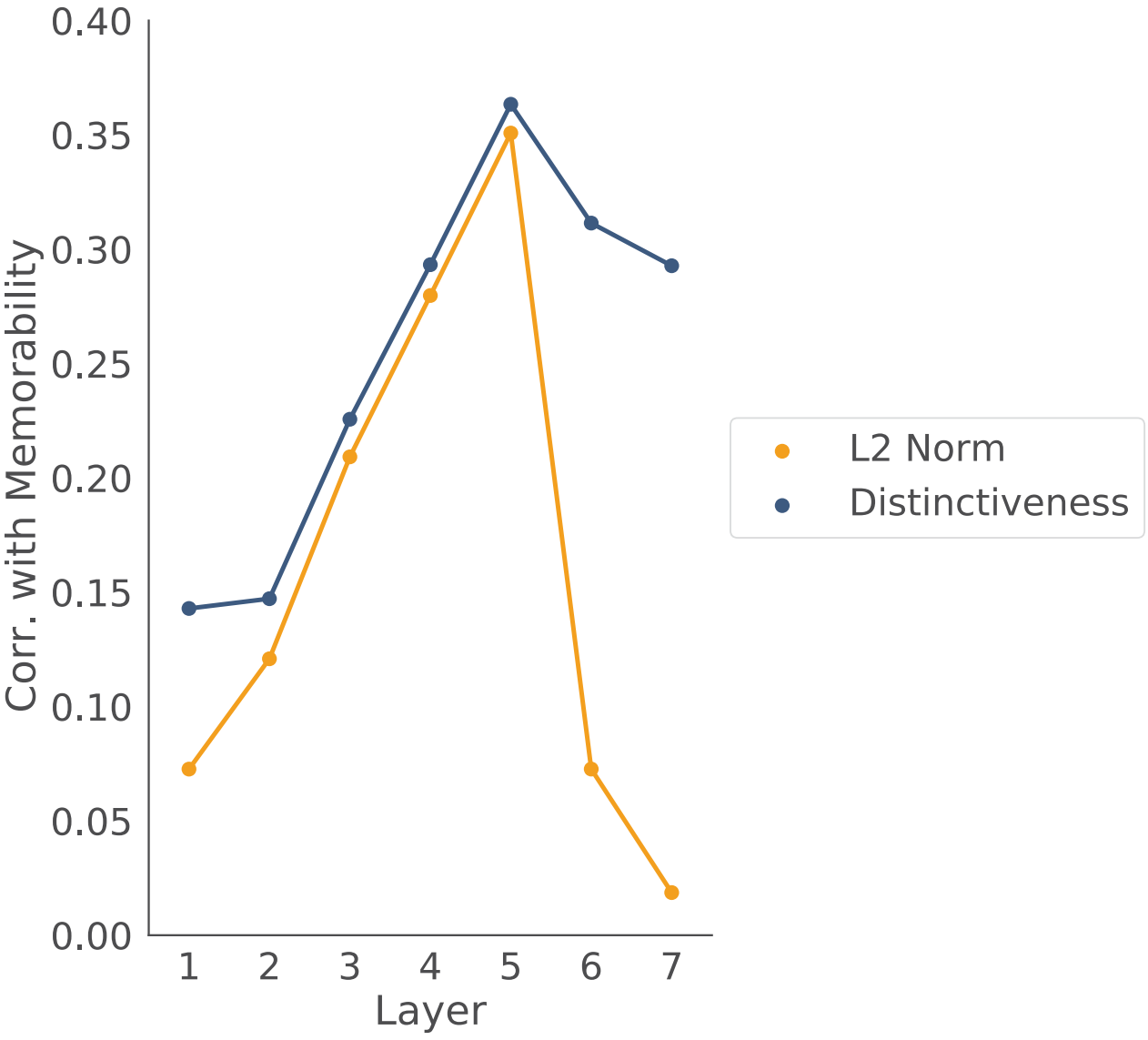

Figure S2: Correlations between memorability and two different measures derived from activation in VGG-16: Distinctiveness (euclidean distance to nearest neighbor) and L2 Norm (L2 norm of the activation vector).

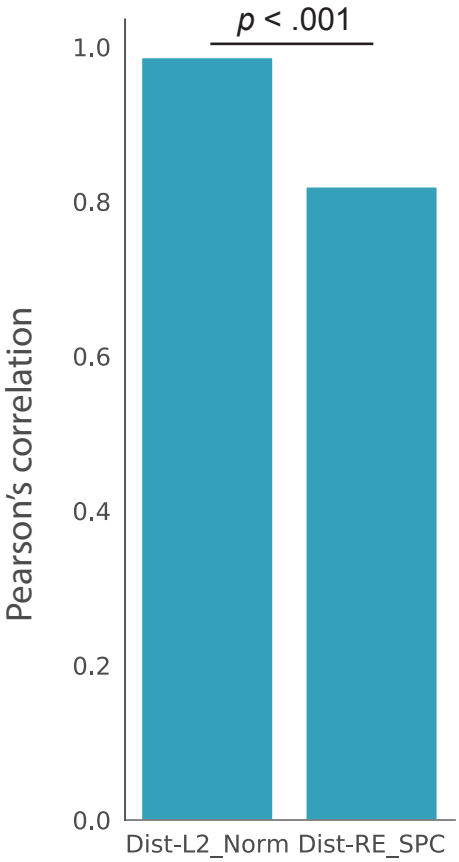

Figure S3: Pearson's correlations between different measures for Layer 5(N_{images} = 2221). Dist(Distinctiveness): euclidean distance to nearest neighbor. L2_Norm: L2 norm of the activation vector. RE_SPC: reconstruction error from the sparse coding model. Statistical significance was assessed with two-tailed Williams' t test (for two correlation values sharing one common variable on the same population; $t(2218) = 73.67$, $p < .001$, Cohen's $q = 1.49$, 95% CI of the difference = [0.15, 0.18]).

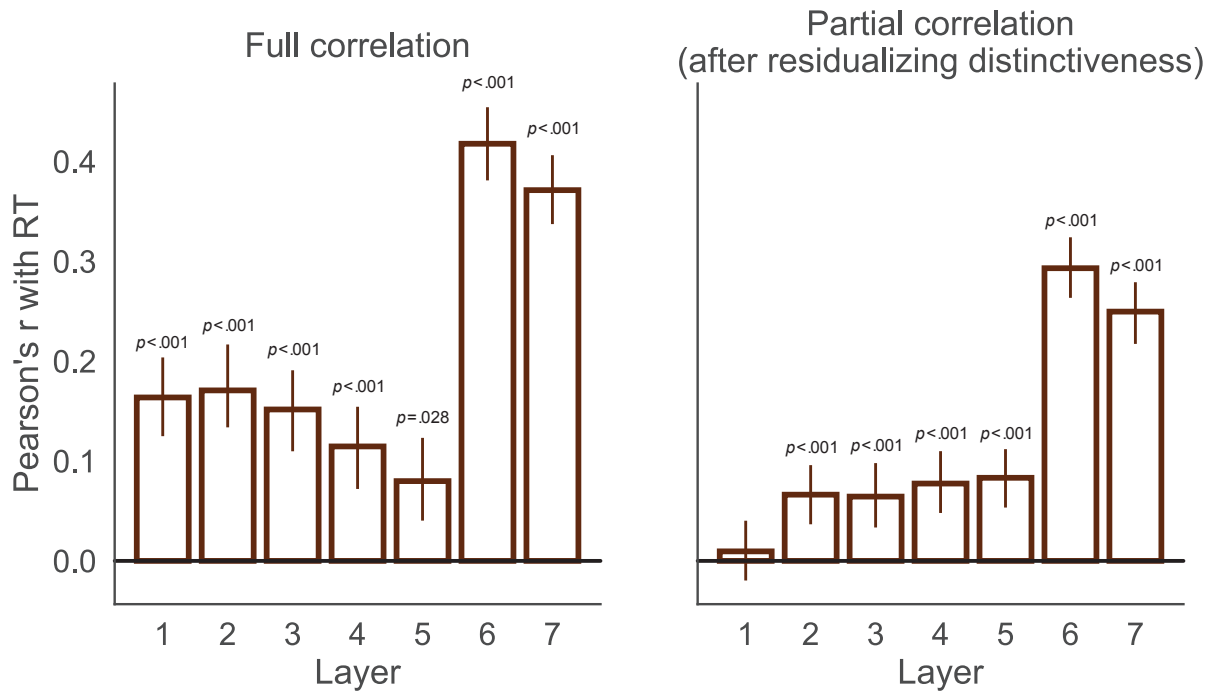

Figure S4: Pearson's r between memorability and reconstruction error (normalized with L2 Norm of the activation vector; Layer 1: $r = .16$ [.12, .20], $p < .001$; Layer 2: $r = .17$ [.13, .22], $p < .001$; Layer 3: $r = .15$ [.11, .19], $p < .001$; Layer 4: $r = .11$ [.07, .15], $p < .001$; Layer 5: $r = .08$ [.04, .12], $p = .028$; Layer 6: $r = .42$ [.38, .45], $p < .001$; Layer 7: $r = .37$ [.34, .41], $p < .001$) and partial Pearson's r after accounting for distinctiveness (Layer 1: $r = .01$ [-.02, .04], $p > .999$; Layer 2: $r = .07$ [.04, .10], $p < .001$; Layer 3: $r = .06$ [.03, .10], $p < .001$; Layer 4: $r = .08$ [.05, .11], $p < .001$; Layer 5: $r = .08$ [.05, .11], $p < .001$; Layer 6: $r = .29$ [.26, .32], $p < .001$; Layer 7: $r = .25$ [.22, .28], $p < .001$). Statistical comparisons are made using direct bootstrap hypothesis testing based on a two-tailed test threshold. P values have been corrected for multiple comparisons with Bonferroni correction ($\alpha = 0.05/14$). Error bars represent 95% confidence intervals from 1000 bootstrapping iterations. $N_{images} = 2221$.

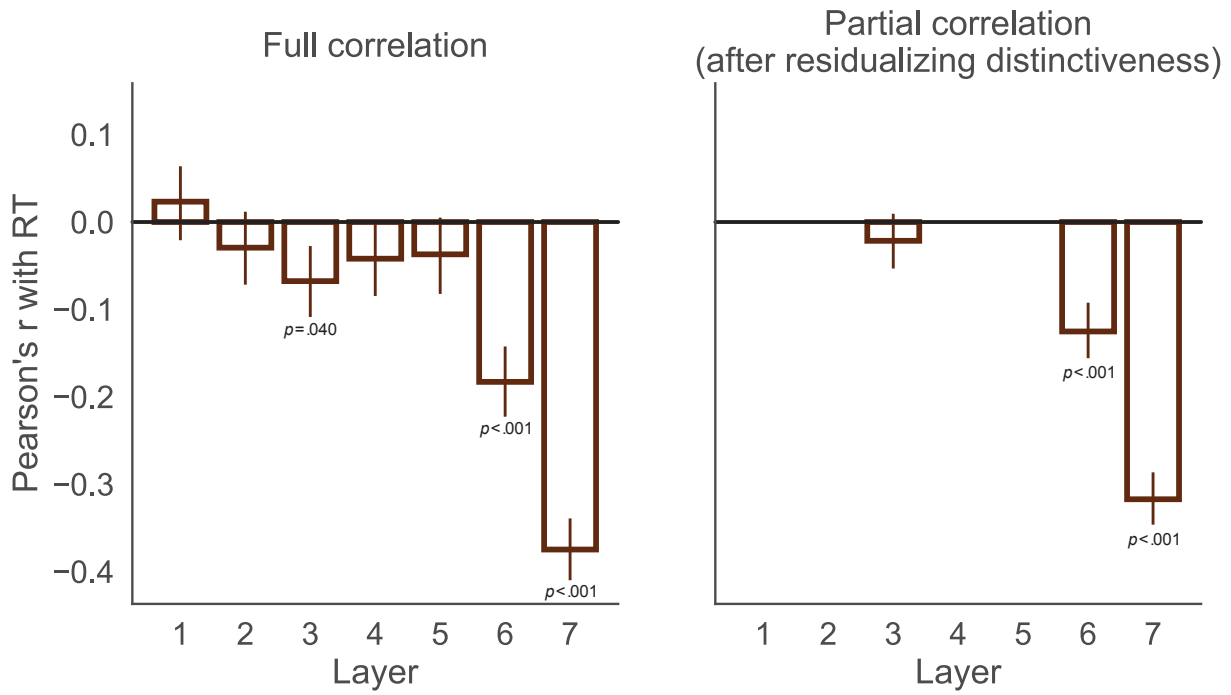

Figure S5: Pearson's r between response times during retrieval and reconstruction error (normalized with L2 Norm of the activation vector; Layer 1: r = .02 [-.02, .06], $p > .999$; Layer 2: r = -.03 [-.07, .01], $p > .999$; Layer 3: r = -.07 [-.11, -.03], $p = .040$; Layer 4: r = -.04 [-.08, .00], $p = .660$; Layer 5: r = -.04 [-.08, -.01], $p = .840$; Layer 6: r = -.18 [-.22, -.14], $p < .001$; Layer 7: r = -.37 [.41, -.34], $p < .001$) and partial Pearson's r after accounting for distinctiveness(Layer 3: r = -.02 [-.05, -.01], $p > .999$; Layer 6: r = -.13 [-.16, -.09], $p < .001$; Layer 7: r = -.32 [-.35, -.29], $p < .001$). Note that partial correlation was only performed if the full correlation turned out to be statistically significant. Therefore, in the right plot, there were no partial correlation results for Layers 1, 2, 4 and 5. Error bars represent 95% confidence intervals from 1000 bootstrapping iterations. Statistical comparisons are made using direct bootstrap hypothesis testing based on a two-tailed test threshold. P values have been corrected for multiple comparisons with Bonferroni correction ($\alpha = 0.05/10$). $N_{images} = 2221$.

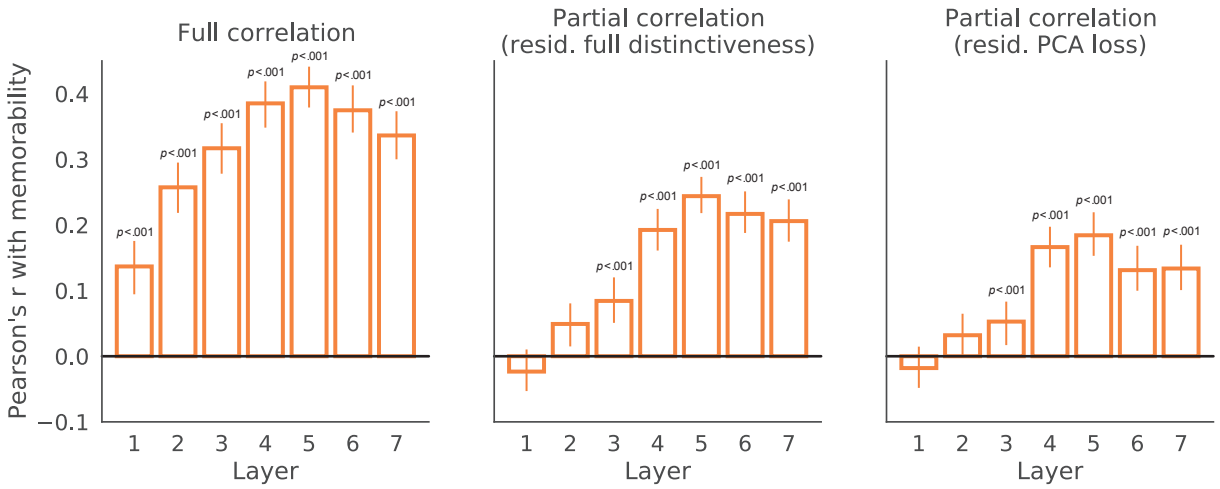

Figure S6: Pearson's r between memorability and reconstruction error (obtained from training with 1000 PCs; Layer 1: $r = .14$ [.09, .18], $p < .001$; Layer 2: $r = .26$ [.22, .30], $p < .001$; Layer 3: $r = .32$ [.28, .36], $p < .001$; Layer 4: $r = .39$ [.35, .42], $p < .001$; Layer 5: $r = .41$ [.38, .44], $p < .001$; Layer 6: $r = .37$ [.34, .41], $p < .001$; Layer 7: $r = .34$ [.30, .37], $p < .001$), partial Pearson's r after accounting for distinctiveness (Layer 1: r = -.02 [-.05, .01], $p > .999$; Layer 2: $r = .05$ [.01, .08], $p = .210$; Layer 3: $r = .08$ [.05, .12], $p < .001$; Layer 4: $r = .19$ [.16, .22], $p < .001$; Layer 5: $r = .24$ [.22, .27], $p < .001$; Layer 6: $r = .22$ [.19, .25], $p < .001$; Layer 7: $r = .21$ [.17, .24], $p < .001$) and partial Pearson's r after accounting for reconstruction loss due to reducing to 1000 PCs (Layer 1: r = -.02 [-.05, .01], $p > .999$; Layer 2: $r = .03$ [.00, .06], $p = .966$; Layer 3: $r = .05$ [.02, .08], $p < .001$; Layer 4: $r = .17$ [.14, .20], $p < .001$; Layer 5: $r = .18$ [.15, .22], $p < .001$; Layer 6: $r = .13$ [.10, .17], $p < .001$; Layer 7: $r = .13$ [.10, .17], $p < .001$). For each of the 7 layers, the top 1000 principal components were learned in an independent set of 8000 images randomly sampled from the Microsoft COCO dataset[68]. We then applied the transformation to the activations evoked by the target images in the Isola dataset and thus reduced the dimensionality of each image's activation to 1000. The 1000 PCs were used for training the sparse coding models and obtaining the reconstruction errors. We additionally evaluated whether the observed relationship between the reconstruction error derived from reconstructing PCs and memorability could be confounded by the PCA dimensionality reduction procedure by deriving a PCA loss measure per image. PCA loss was defined as the euclidean distance between the original DCNN activation vector (with all units in a given layer) and the reconstructed vector by inverting the transformation matrix and applying it to the reduced 1000-PC for each image. Error bars represent 95% confidence intervals from 1000 bootstrapping iterations. Statistical comparisons are made using direct bootstrap hypothesis testing based on a two-tailed test threshold. P values have been corrected for multiple comparisons with Bonferroni correction ($\alpha = 0.05/21$).$N_{images} = 2221$.

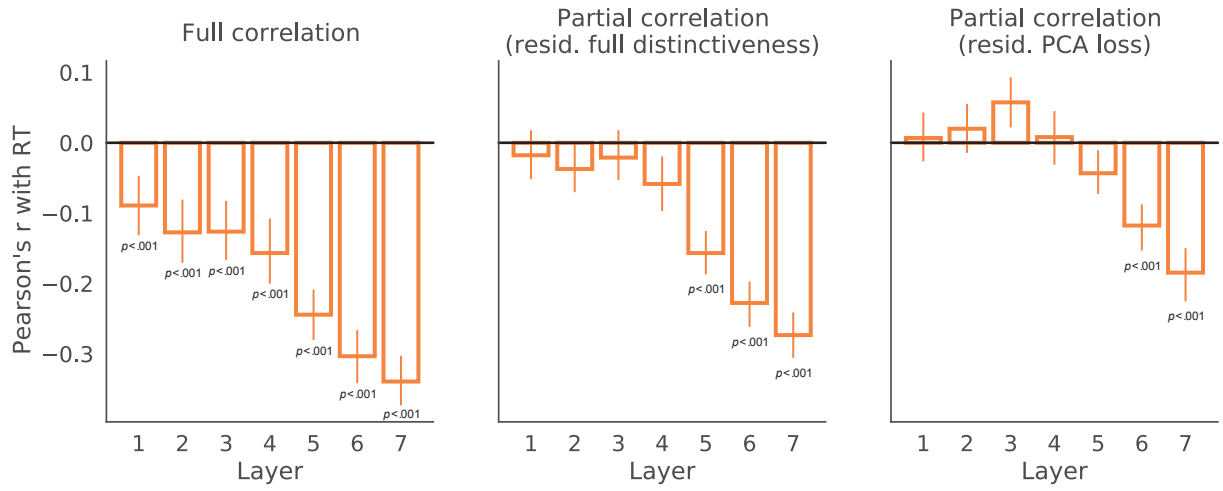

Figure S7: Pearson's r between response times during retrieval and reconstruction error (obtained from training with 1000 PCs; Layer 1: r = -.09 [-.13, -.05], $p < .001$; Layer 2: r = -.13 [-.17, -.08], $p < .001$; Layer 3: r = -.13 [-.17, -.08], $p < .001$; Layer 4: r = -.16 [-.20, -.11], $p < .001$; Layer 5: r = -.24 [.28, .21], $p < .001$; Layer 6: r = -.30 [-.34, -.27], $p < .001$; Layer 7: r = -.34 [-.37, -.30], $p < .001$), partial Pearson's r after accounting for distinctiveness(Layer 1: r = -.02 [-.05, .02], $p > .999$; Layer 2: r = -.04 [-.07, -.00], $p = .798$; Layer 3: r = -.02 [-.05, .02], $p > .999$; Layer 4: r = -.06 [-.10, -.02], $p = .084$; Layer 5: r = -.16 [-.19, -.13], $p < .001$; Layer 6: r = -.23 [-.26, -.20], $p < .001$; Layer 7: r = -.27 [-.31, -.24], $p < .001$) and partial Pearson's r after accounting for reconstruction loss due to reducing to 1000 PCs (Layer 1: r = .01 [-.03, .04], $p > .999$; Layer 2: r = .02 [-.01, .06], $p > .999$; Layer 3: r = .06 [.02, .09], p = .147; Layer 4: r = .01 [-.03, .04], $p > .999$; Layer 5: r = -.04 [-.07, -.01], $p = .861$; Layer 6: r = -.12 [-.15, -.09], $p < .001$; Layer 7: r = -.18 [-.23, -.15], $p < .001$). See the caption for Figure S6 for details on the PCA training and application procedures. Error bars represent 95% confidence intervals from 1000 bootstrapping iterations. Statistical comparisons are made using direct bootstrap hypothesis testing based on a two-tailed test threshold. P values have been corrected for multiple comparisons with Bonferroni correction ($\alpha = 0.05/21$). Only those that survived multiple comparisons were labelled with asterisk(s) on the figure. ***: $p < .001$. $N_{images} = 2221$.