

# Supporting Information

St. Jacques et al. 10.1073/pnas.1319630110

## SI Materials and Methods

**Novel Encoding Condition Procedure.** During the novel encoding task, participants were shown photographs taken at museums in the Boston area that they had not previously visited. On each trial, participants were shown a photograph of a novel museum stop and asked to rate how much detail, or visual information, was depicted in the photograph on a five-point scale from low to high (Fig. S4). For partial trials (48 trials), fixation immediately followed. For full trials (70 trials), a second photograph of a museum stop immediately followed, which was selected to be similar in content to the first photograph, and participants were instructed to indicate how related (i.e., “could it be taken from the same exhibit?”) the second photograph was to the preceding one, on a five-point scale from low to high. The photograph remained on the screen for 5 s. Trials were separated by a variable fixation (2.5–7.5 s) and distributed exponentially such that shorter inter-trial intervals occurred more frequently than longer. An additional block of the novel encoding condition took place postscanning (70 trials). Photographs from this postscanning task were used to provide studied incorrect photographs in the source memory decision in session 3.

During session 3 in the novel encoding task, participants were shown old photographs [118 trials, i.e., photographs in which a detail rating was made during the novel encoding condition during functional magnetic resonance imaging (fMRI) scanning] and new photographs (70 trials), both taken from museums they had not visited during the study, and were asked to make an old/new decision concerning whether the photograph had been shown during the scanning session (Fig. S4). They were allowed 6 s to make each decision. An associative memory decision followed “yes” responses, in which participants were asked to select the photograph that it had been paired with (i.e., the photograph in which a relatedness rating was made during the novel encoding condition during fMRI scanning). They were shown the correct photograph, a studied but incorrect photograph (i.e., photograph from a postscanning task in which participants made a relatedness rating), or the option to choose “none” if they believed it had been previously shown during a partial trial. The correct and incorrect photographs were matched for similarity in content, usually by choosing scenes from the same museum exhibit. This was followed by a 6-s confidence rating on a five-point scale from low to high.

**Quality Control.** In addition to visually inspecting the MRI data, quantitative parameters were used to determine the quality of each fMRI run. Based on measures developed at the Harvard Center for Brain Science for data from the Siemens 3T 12 channel, the following values were used as exclusion criteria: (i) slice signal to noise ratio < 99, (ii) maximum absolute motion > 2 mm, (iii) number of movements greater than 0.5 mm > 5, or (iv) a combination of poor slice signal to noise ratio (between 99–149) and poor maximum absolute motion (between 1.49–1.99 mm). Problematic runs were excluded from the fMRI analysis. Additionally, participants with two or more problematic fMRI runs were excluded from the fMRI analysis due to an insufficient number of trials for subsequent memory analysis.

**fMRI Analysis.** In the novel encoding condition, we examined neural activity during retrieval that was associated with subsequent true memories. The primary analysis was a paired *t* test that examined the average neural activity across the trial phases for subsequent hits versus subsequent misses.

## SI Results

**Behavioral Results. Reactivation task.** During session 2, related ratings were higher for photographs associated with subsequent hits [mean (*M*) = 3.46, SD = 0.43] versus subsequent misses [*M* = 3.10, SD = 0.51; *t*(32) = 3.74, *P* = 0.001], and for subsequent false alarms (*M* = 3.75, SD = 0.38) versus subsequent correct rejections [*M* = 3.22, SD = 0.42; *t*(32) = 8.69, *P* < 0.0001]. There was also a significant within-participant correlation between relatedness ratings and subsequent true memories (*r* = 0.12, *P* = 0.0001) and between relatedness ratings and subsequent false memories (*r* = 0.19, *P* = 0.0002). Further, there was a significant correlation between reliving and relatedness ratings (*r* = 0.24, *P* < 0.0001) and between reliving ratings and reliving reaction time (RT) (*r* = -0.33, *P* < 0.0001). To examine the unique association between reliving and subsequent memory on a trial-by-trial basis within each individual participant, we conducted partial correlation analyses that controlled for the influence of relatedness and RT. After controlling for relatedness, there was a significant association between reliving and subsequent true memories (*r* = 0.29, *P* < 0.0001) and between reliving and subsequent false memories (*r* = 0.05, *P* = 0.05). After controlling for reliving RT, there was a significant association between reliving and subsequent true memories (*r* = 0.28, *P* < 0.0001) and between reliving and subsequent false memories (*r* = 0.08, *P* = 0.001). Finally, controlling for both relatedness and reliving RT revealed a significant association between reliving and subsequent true memories (*r* = 0.28, *P* < 0.0001) and a trend between reliving and subsequent false memories (*r* = 0.05, *P* = 0.07). In sum, the quality of memory reactivation generally contributed to subsequent true- and false-memory effects irrespective of how quickly the target memory was retrieved or how related novel information was.

To examine RT in session 2, we conducted a 2 (subsequent memory stimuli: reactivated target, reactivated lure) × 2 (subsequent response: “yes,” “no”) ANOVA separately on the reliving and relatedness RT in phase 1 and 2, respectively. For reliving RT, there was also a significant main effect of subsequent response [*F*(1, 32) = 15.94, *P* = 0.0003,  $\eta_p^2$  (partial eta squared) = 0.33] with faster reliving RT for yes (*M* = 2.63 s, SD = 0.44 s) versus no (*M* = 2.76 s, SD = 0.46 s) responses. However, this was qualified by a significant interaction [*F*(1, 32) = 17.10, *P* = 0.0002,  $\eta_p^2$  = 0.35]. Post hoc analyses revealed that reliving RT was faster for memories associated with subsequent hits (*M* = 2.60 s, SD = 0.44 s) versus misses (*M* = 2.83 s, SD = 0.46 s) [*t*(32) = -5.78, *P* = 0.000002] but there was no difference in RT associated with subsequent false alarms (*M* = 2.66 s, SD = 0.44 s) versus correct rejections (*M* = 2.69 s, SD = 0.45 s). There was no significant main effect of the subsequent memory condition for reliving RT. For relatedness RT, there was a significant main effect of subsequent response [*F*(1, 32) = 12.24, *P* = 0.001,  $\eta_p^2$  = 0.28] with faster RT for yes (*M* = 2.30 s, SD = 0.40 s) versus no (*M* = 2.46 s, SD = 0.51 s) responses. There were no other significant effects.

To examine RT in session 3, we conducted a 2 (stimuli: target, lure) × 2 (response: yes, no) × 2 (condition: reactivated, baseline) ANOVA separately on RT for recognition memory and confidence ratings (note: degrees of freedom for RT and confidence rating reflect the exclusion of five participants with zero baseline false alarms). For recognition memory RT, there was a main effect of response [*F*(1, 27) = 32.12, *P* = 0.000005,  $\eta_p^2$  = 0.54], which was reflected by faster RT for recognition memory for yes (*M* = 2.94 s, SD = 0.56 s) versus no (*M* = 3.28 s, SD = 0.60 s) responses.

However, there was also a significant stimulus  $\times$  response interaction [ $F(1, 27) = 19.41, P = 0.0002, \eta_p^2 = 0.42$ ]. Post hoc tests indicated that this was reflected by faster RTs for target yes ( $M = 2.29$  s,  $SD = 0.42$  s) versus no ( $M = 3.36$  s,  $SD = 0.61$  s) responses [ $t(32) = -7.56, P < 0.0001$ ] compared with lure yes ( $M = 3.10$  s,  $SD = 0.65$  s) versus no ( $M = 3.20$  s,  $SD = 0.59$  s) responses [ $t(32) = -1.18, P = 0.28$ ]. There was a significant response  $\times$  condition interaction [ $F(1, 27) = 55.79, P < 0.00001, \eta_p^2 = 0.67$ ] which was reflected by faster RTs for reactivated yes ( $M = 2.78$  s,  $SD = 0.42$  s) versus no ( $M = 3.42$  s,  $SD = 0.57$  s) responses [ $t(32) = -9.74, P < 0.0001$ ] compared with baseline yes ( $M = 3.11$  s,  $SD = 0.64$  s) versus no ( $M = 3.14$  s,  $SD = 0.63$  s) responses [ $t(32) = -0.24, P = 0.81$ ]. However, these effects were qualified by a significant three-way interaction [ $F(1, 27) = 11.56, P = 0.002, \eta_p^2 = 0.30$ ]. Follow-up analyses indicated that there was a significant stimuli  $\times$  response interaction in RT for recognition memory in the reactivated condition [ $F(1, 32) = 120.87, P < 0.0001, \eta_p^2 = 0.79$ ] but not in the baseline condition [ $F(1, 27) = 1.19, P = 0.29, \eta_p^2 = 0.04$ ]; RTs were faster for reactivated target yes ( $M = 2.51$  s,  $SD = 0.35$  s) versus no ( $M = 3.54$  s,  $SD = 0.60$  s) responses compared with reactivated lure yes ( $M = 3.05$  s,  $SD = 0.49$  s) versus no ( $M = 2.76$  s,  $SD = 0.46$  s) responses.

To examine the confidence rating RT in session 3, we conducted a 2 (stimuli: target, lure)  $\times$  2 (response: yes, no)  $\times$  2 (condition: reactivated, baseline) ANOVA. There was a main effect of response [ $F(1, 27) = 24.94, P < 0.0001, \eta_p^2 = 0.48$ ] reflected by faster RTs for confidence ratings on yes ( $M = 1.13$  s,  $SD = 0.38$  s) versus no ( $M = 1.25$  s,  $SD = 0.46$  s) responses. There was also a trend for a main effect of stimuli [ $F(1, 27) = 3.94, P = 0.06, \eta_p^2 = 0.13$ ] with faster RTs for targets ( $M = 1.15$  s,  $SD = 0.48$  s) than lures ( $M = 1.21$  s,  $SD = 0.39$  s). However, these effects were qualified by a significant response  $\times$  stimuli interaction [ $F(1, 27) = 25.58, P < 0.0001, \eta_p^2 = 0.49$ ]. Post hoc analyses indicated that this was reflected by faster RTs for confidence ratings for yes ( $M = 0.96$  s,  $SD = 0.34$  s) versus no ( $M = 1.34$  s,  $SD = 0.53$  s) responses to targets [ $t(32) = -7.25, P < 0.0001$ ] but slower RTs for yes ( $M = 1.30$  s,  $SD = 0.40$  s) versus no ( $M = 1.14$  s,  $SD = 0.33$  s) responses to lures [ $t(32) = 2.85, P = 0.008$ ]. There were no other significant effects.

To examine differences in confidence for recognition memory decisions made in session 3, we conducted a 2 (stimuli: target, lure)  $\times$  2 (response: yes, no)  $\times$  2 (condition: reactivated, baseline) ANOVA. There was a main effect of response [ $F(1, 27) = 46.81, P < 0.0001, \eta_p^2 = 0.63$ ] reflected by higher confidence ratings for yes ( $M = 3.67, SD = 0.55$ ) versus no ( $M = 3.07, SD = 0.53$ ) responses. There was a significant response  $\times$  condition interaction [ $F(1, 27) = 10.53, P = 0.003, \eta_p^2 = 0.28$ ], which was reflected by higher confidence for reactivated yes ( $M = 3.77, SD = 0.42$ ) versus “no” ( $M = 2.95, SD = 0.45$ ) responses [ $t(32) = 10.57, P < 0.0001$ ] than for baseline yes ( $M = 3.58, SD = 0.60$ ) versus no ( $M = 3.20, SD = 0.60$ ) responses [ $t(32) = 4.01, P = 0.0003$ ]. There was also a significant response  $\times$  stimuli interaction [ $F(1, 27) = 126.10, P < 0.0001, \eta_p^2 = 0.82$ ], indicating higher confidence for yes ( $M = 4.04, SD = 0.39$ ) versus no ( $M = 2.80, SD = 0.60$ ) responses to targets [ $t(32) = 13.70, P < 0.001$ ], and equal confidence for yes ( $M = 3.31, SD = 0.71$ ) and no ( $M = 3.37, SD = 0.46$ ) responses made to lures [ $t(32) = -0.64, P = 0.52$ ]. There were no other significant effects.

Accuracy ( $d'$ ) and response bias ( $c$ ) were calculated according to detection theory (1) for each condition. Consistent with data showing that reactivation increase both true and false memories relative to baseline, there were no differences in accuracy between the reactivated ( $d' = 1.19, SD = 0.36$ ) and baseline ( $d' =$

1.33,  $SD = 0.67$ ) conditions [ $t(32) = -1.29, P = 0.21$ ]. However, response bias differed in the two conditions [ $t(32) = -8.51, P < 0.0001$ ] such that there was a negative response bias in the reactivated condition (i.e., tendency to respond “yes”;  $c = -0.21, SD = 0.32$ ), but a positive response bias in the baseline condition (i.e., tendency to respond “no”;  $c = 0.26, SD = 0.38$ ).

**Novel encoding task.** During session 2, related ratings were higher for photographs associated with subsequent hits ( $M = 4.13, SD = 0.38$ ) versus subsequent misses ( $M = 3.70, SD = 0.46$ ) [ $t(32) = 8.81, P < 0.0001$ ], but there was no difference in detail ratings made for hits ( $M = 3.05, SD = 0.34$ ) versus misses ( $M = 3.00, SD = 0.37$ ). There was also a significant within-participant correlation between the detail and relatedness ratings ( $r = 0.11, P < 0.0001$ ). RT was faster when making the detail rating [ $t(32) = -2.70, P = 0.01$ ] and relatedness rating [ $t(32) = -4.39, P = 0.0001$ ] for subsequent hits (detail RT:  $M = 2.35, SD = 0.39$ ; relatedness RT:  $M = 1.97, SD = 0.30$ ) versus misses (detail RT:  $M = 2.42, SD = 0.40$ ; relatedness RT:  $M = 2.09, SD = 0.46$ ).

We analyzed recognition-memory performance in session 3 (hits:  $M = 0.50, SD = 0.14$ ; false alarms:  $M = 0.19, SD = 0.10$ ; percent correct:  $M = 0.62, SD = 0.08$ ;  $d' = 0.96, SD = 0.43$ ;  $c = 0.49, SD = 0.36$ ; associative memory:  $M = 0.45, SD = 0.10$ ). To examine differences in RT for recognition memory decisions, we conducted a 2 (stimuli: target, lure)  $\times$  2 (response: yes, no) ANOVA. There was a significant stimuli  $\times$  response interaction [ $F(1, 32) = 11.42, P = 0.002$ ], which was reflected by faster RT for correct responses (hits, correct rejections;  $M = 2.05, SD = 0.43$ ) than for incorrect responses (misses, false alarms;  $M = 2.17, SD = 0.52$ ). There were no significant main effects on RT for recognition memory decisions. During the associative memory decision, paired  $t$  tests indicated that RT was faster for correct ( $M = 2.80, SD = 0.50$ ) versus incorrect ( $M = 2.71, SD = 0.50$ ) associative memory decisions [ $t(32) = -2.04, P = 0.05$ ]. There were no other significant differences.

#### fMRI Results: Subsequent Memory Effects That Differ for True and False Memories.

There were also some differences in neural recruitment associated with subsequent true- and subsequent false-memory effects (Table S2). First, compared with subsequent false-memory effects, subsequent true-memory effects were associated with greater recruitment of the left dorsolateral prefrontal cortex (PFC) and left superior parietal cortex regions during the target versus lure presentation (Fig. S2 A and C). Second, the left occipital cortex was recruited to a greater extent overall for true versus false memories, whereas bilateral temporal cortices were recruited more for false versus true memories (Fig. S2 B and C). However, neural recruitment that differed for true versus false subsequent memory was unrelated to the quality of reactivation (Fig. S2D). Thus, the pattern of neural recruitment observed here may reflect divergent processing of the target and lure photographs that differentially contributes to true versus false subsequent memory, rather than differences related to memory reactivation. One interpretation is that subsequent true memories are supported by increased engagement of frontoparietal control regions during presentation of the target along with greater attention to visual information in the photographs, consistent with evidence that true and false memories can sometimes be distinguished by neural recruitment in occipital cortex (2, 3). In contrast, subsequent false memories involved less engagement of frontoparietal control regions and greater involvement of bilateral temporal cortices, which some studies have linked to conceptual processes that contribute to the formation of false memories (4).

- Macmillan NA, Creelman CD (2005) *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, Mahwah, NJ).
- Slotnick SD, Schacter DL (2004) A sensory signature that distinguishes true from false memories. *Nat Neurosci* 7(6):664–672.

- Okado Y, Stark C (2003) Neural processing associated with true and false memory retrieval. *Cogn Affect Behav Neurosci* 3(4):323–334.
- Dennis NA, Kim H, Cabeza R (2007) Effects of aging on true and false memory formation: An fMRI study. *Neuropsychologia* 45(14):3157–3166.









**Table S2. Trial phase (target, lure) × memory condition (true, false) ANOVA on subsequent memories**

Region	BA	MNI coordinates			t	Z	Voxels
		x	y	z			
Main effect of memory condition							
True memories > false memories							
Occipital cortex	17	2	-88	0	3.80	2.66	37
False memories > true memories							
Lateral temporal cortex	21/22	-52	-2	-10	4.13	3.96	34
	21/22	58	-2	-10	3.81	3.67	100
Main effect of trial phase							
Target > lure							
Dorsal anterior cingulate cortex	32	-6	16	48	3.74	3.61	56
Lure > target					No significant voxels		
Memory condition × trial phase							
True memories (target > lure) > false memories (target > lure)							
Dorsolateral PFC	9/45	-46	24	26	3.46	3.35	34
Dorsal anterior cingulate cortex	32/8	-6	20	46	4.08	3.92	109
Premotor cortex	6/8	-32	12	50	3.96	3.81	115
Posterior inferior parietal cortex	39/19	-32	-78	34	3.79	3.65	35
Anterior inferior parietal cortex	40/7	-36	-56	48	3.42	3.32	15
Precuneus	7	-14	-70	44	3.83	3.69	73
Retrosplenial cortex	30	-16	-58	22	3.44	3.34	22
	30/29	22	-60	24	4.22	4.04	85
Posterior parahippocampal cortex	36	-28	-30	-14	3.85	3.71	49
False memories (target > lure) > true memories (target > lure)							
					No significant voxels		

MNI, Montreal Neurological Institute. BA, Brodmann's area (approximate).

**Table S3. Subsequent true memory for novel encoding**

Region	BA	MNI coordinates			t	Z	Voxels
		x	y	z			
Medial PFC	9	-10	52	36	3.79	3.34	11
Anterior hippocampus		-14	-10	-14	4.32	3.70	58
Fusiform cortex	37	38	-32	-16	4.13	3.57	81

MNI, Montreal Neurological Institute; BA, Brodmann's area (approximate).

**Table S4. Trial phase (target, lure) × memory (hit, false alarm) ANOVA on parametric modulation by reliving**

Region	BA	MNI coordinates			t	Z	Voxels
		x	y	z			
Main effect of memory condition							
Hit > false alarm							No significant voxels
False alarm > hit							
Dorsolateral PFC	9	60	2	26	3.68	3.56	63
Ventrolateral PFC	44	-54	2	22	3.29	3.19	13
	44/45	40	36	4	3.88	3.73	36
Motor cortex	4	-62	-12	30	3.51	3.40	15
Auditory cortex	41	44	-26	14	3.62	3.50	23
Temporal pole	38	-36	12	-20	3.88	3.74	12
Main effect of trial phase							
Target > lure							
Rostral medial PFC	10	-8	64	18	3.90	3.75	26
Ventromedial PFC	11	-2	44	-14	5.07	4.77	513
Ventrolateral PFC	47	-50	40	-4	4.49	4.27	274
	47/11	-24	22	-16	4.07	3.90	47
Lateral orbitofrontal cortex	11	-38	38	-16	3.48	3.37	10
Anterior premotor cortex	8	-24	32	48	5.46	5.09	852
Lateral temporal cortex	21	-66	-32	-6	4.58	4.35	90
	21	-54	-20	-26	4.20	4.02	223
	21	60	2	-16	3.85	3.70	60
Posterior inferior parietal cortex	39	-46	-70	34	4.60	4.37	978
Posterior cingulate	31	-12	-50	34	4.60	4.37	955
Amygdala		-22	0	-16	3.48	3.37	11
		24	-4	-22	3.68	3.55	53
Anterior hippocampus		32	-18	-22	3.48	3.37	12
Posterior hippocampus		-28	-32	-10	3.55	3.43	17
Ventral striatum		8	0	-6	4.30	4.11	55
Cerebellum		18	-76	-30	3.74	3.60	25
Lure > target							
Ventrolateral PFC	47	32	20	-2	3.58	3.47	44
Somatosensory cortex	1/2/3	-48	-18	50	4.47	4.25	168
Posterior parietal cortex	40	-32	-32	44	3.60	3.48	14
Memory condition × trial phase							
Hit (target > lure) > false alarm (target > lure)							
Rostral medial PFC	9	-10	52	8	4.32	4.12	629
Ventral anterior cingulate cortex	32/9	-10	42	26	3.68	3.56	22
Anterior premotor cortex	8	-22	14	44	3.78	3.64	80
Frontal eye field	8	-16	28	48	3.5	3.39	25
Posterior cingulate	31/7	-2	-64	32	3.96	3.81	102
False alarm (target > lure) > hit (target > lure)							
Ventromedial PFC	11	-2	44	-16	3.41	3.31	23
Ventrolateral PFC	47	-52	42	-4	3.75	3.61	11
Lateral temporal cortex	21	-52	-8	-10	3.49	3.38	20
	20	-54	-20	-26	3.67	3.55	15
Anterior hippocampus		32	-18	-20	3.61	3.49	16
Ventral striatum		8	0	-6	4	3.84	19
Midbrain		8	-28	-22	3.66	3.53	18
Hit (lure > target) > false alarm (lure > target)							No significant voxels
False alarm (lure > target) > hit (lure > target)							No significant voxels

MNI, Montreal Neurological Institute; BA, Brodmann's area (approximate).

**Table S5. Trial phase (target, lure) × memory (hit, false alarm) ANOVA on parametric modulation by relatedness**

Region	BA	MNI coordinates			t	Z	Voxels
		x	y	z			
<b>Main effect of memory condition</b>							
Hit > false alarm		No significant voxels					
False alarm > hit		No significant voxels					
Dorsolateral PFC	45	-56	20	14	3.64	3.51	46
Ventrolateral PFC	47	-34	34	-8	4.39	4.18	134
	47	56	32	0	3.81	3.66	71
Anterior cingulate cortex	32	-12	46	2	4.71	4.45	157
	32	-12	48	22	3.60	3.47	26
	32	-16	34	26	3.92	3.77	33
Rostral medial PFC	9	8	56	16	3.91	3.91	89
	10	-10	62	24	3.55	3.43	12
Primary motor cortex	4	-64	-4	20	4.37	4.16	57
Precuneus	7	14	-80	48	3.71	3.58	26
Lateral temporal cortex	20	44	-16	-28	3.64	3.51	10
Occipitotemporal cortex	47	56	-66	8	3.49	3.38	40
<b>Main effect of trial phase</b>							
Target > lure		No significant voxels					
Rostral PFC	10	-22	42	14	3.90	3.75	34
Dorsolateral PFC	9/46	-40	36	12	4.24	4.05	445
	9	42	16	22	3.87	3.72	83
Premotor cortex	6	-32	-12	42	5.08	4.77	152
Anterior premotor cortex	8/6	0	14	54	4.08	3.90	112
Frontal eye field	8	-48	4	40	3.71	3.57	21
Precuneus	7	-14	-68	40	3.71	3.58	67
Retrosplenial cortex	30/29	20	-60	24	4.39	4.18	135
	30	-20	-64	16	3.75	3.62	24
Midbrain		4	-22	-22	4.36	4.16	55
Cerebellum		36	-52	-32	4.01	3.85	122
Lure > target		No significant voxels					
<b>Memory condition × trial phase</b>							
Hit (target > lure) > false alarm (target > lure)		No significant voxels					
False alarm (target > lure) > hit (target > lure)		No significant voxels					
Rostral PFC	10	-22	42	14	3.69	3.56	10
Dorsolateral PFC	46	-40	36	12	3.95	3.79	43
Ventrolateral PFC	44	40	14	20	4.09	3.92	109
Retrosplenial cortex	30/29	20	-60	24	4.23	4.04	125
Anterior hippocampus		32	-20	-18	3.88	3.73	43
Cerebellum		28	-34	-30	3.85	3.70	25
Midbrain		4	-22	-22	3.72	3.59	29
Hit (lure > target) > false alarm (lure > target)		No significant voxels					
Posterior insular cortex	13	-38	-20	16	3.80	3.66	45
Nucleus accumbens		14	-10	-10	3.69	3.56	33
False alarm (lure > target) > hit (lure > target)		No significant voxels					

MNI, Montreal Neurological Institute; BA, Brodmann's area (approximate).