Time-dependent mnemonic vulnerability induced by new-learning

Fengtao Shen[1], Yixuan Ku[1,3], Jue Wu[1], Yue Cui[1], Jianqi Li[2], Zhaoxin Wang[1,2] *, Huimin Wang[1,3] *, and Sze Chai Kwok[1,2,3] *

[1] Shanghai Key Laboratory of Brain Functional Genomics, Key Laboratory of Brain Functional Genomics Ministry of Education, School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China

[2] Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai 200062, China

[3] NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, Shanghai 200062, China

*Correspondence: sze-chai.kwok@st-hughs.oxon.org (S.C.K.), wzx425@gmail.com (Z.W.), hwang01@gmail.com (H.W.)

Abstract

Reactivation renders consolidated memory labile again, and the ensuing temporary reconsolidation process is highly susceptible to mnemonic modification. Here, we show that memories in such an unstable state could be reprogrammed by sheer behavioral means, bypassing the need for pharmacological intervention. In two experiments using a "face-location association" paradigm in which participants experienced a "Learning – New-learning – Final-test" programme, we demonstrate that reactivated memory traces were robustly hampered when the new learning was strategically administered within a critical 20-minute time window. Using fMRI, we further advance our theoretical understanding that this lability can be mechanistically explained by the differential activation in the hippocampal-amygdala memory system implicated by the new-learning whereas the mnemonic intrusion caused by newly learned memories is efficaciously reconciled by the left inferior frontal gyrus. Our findings provide important implications for educational and clinical practices in devising effective strategies for memory integration.

Keywords. non-emotional declarative memory, reconsolidation, non-invasive manipulation, hippocampus, amygdala, IFG

1 Introduction

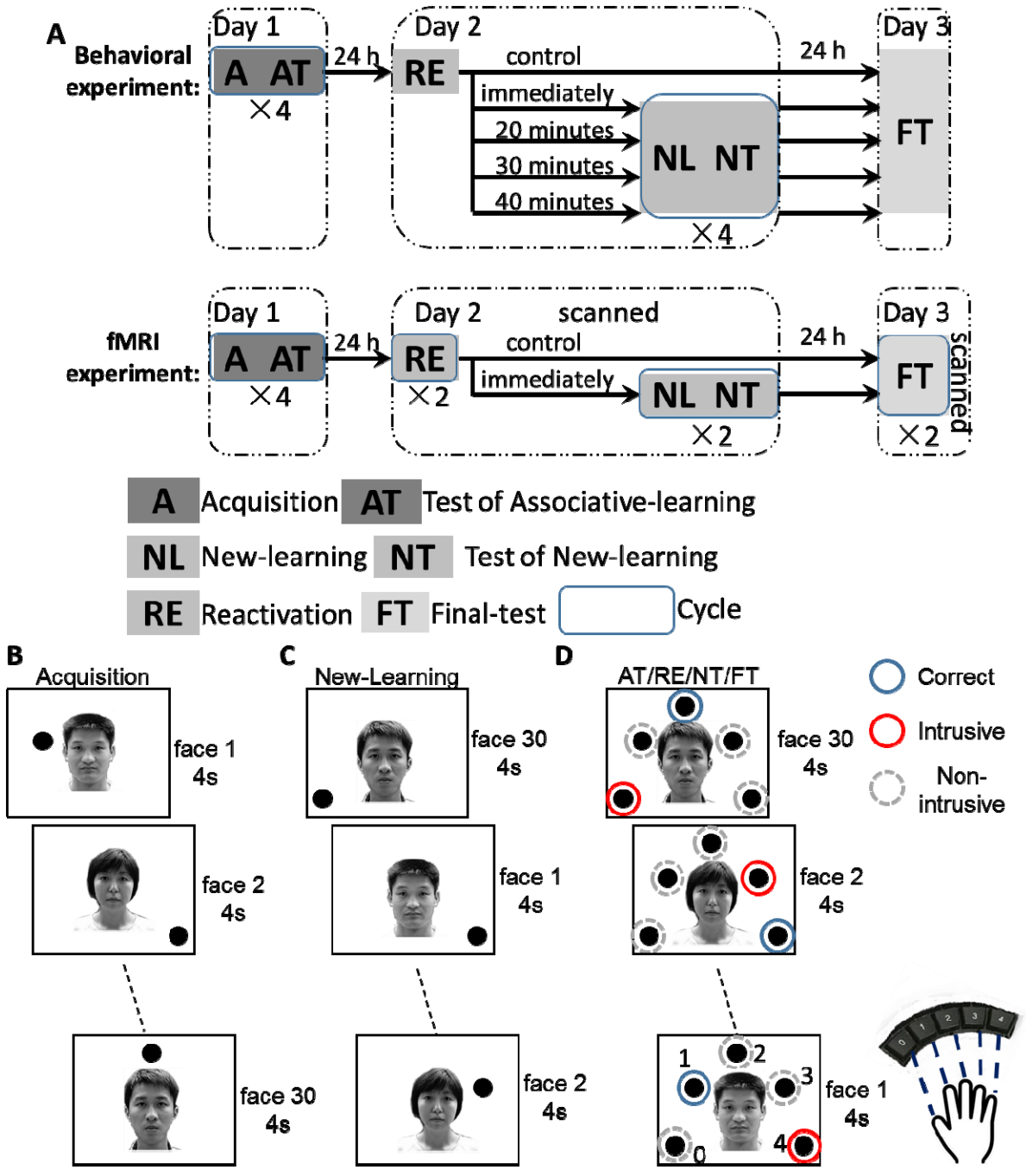
Memory recall is constructive in nature and the mere act of recalling a memory renders it labile and highly susceptible to modification (Nader, Schafe et al. 2000, Lee, Everitt et al. 2004, Lee, Di Ciano et al. 2005, Alberini and LeDoux 2013, Lee, Nader et al. 2017, Scully, Napper et al. 2017). While emotional factors are known to exert effects on reconsolidation of emotional declarative memories (Schwabe and Wolf 2009, Strange, Kroes et al. 2010, Chan and LaPaglia 2013), controversies still surround reconsolidation theories on non-emotional declarative memories. Empirically, post-retrieval manipulations gave rise to inconclusive patterns of results, with some studies showing such manipulation can induce update (Hupbach, Gomez et al. 2007, Hupbach, Gomez et al. 2009, Forcato, Rodriguez et al. 2010), forgetting (Forcato, Burgos et al. 2007), extinction (Nader, Schafe et al. 2000, Schiller, Monfils et al. 2010, Agren, Engman et al. 2012), or enhancement (Coccoz, Maldonado et al. 2011, Coccoz, Sandoval et al. 2013), whereas another set of studies revealing no observable effect (Cammarota, Bevilaqua et al. 2004, Debiec, Doyère et al. 2006, Hupbach, Hardt et al. 2008, Forcato, Argibay et al. 2009, Hupbach, Gomez et al. 2011, Gershman, Schapiro et al. 2013). These previous studies indicate that manipulations after reactivation would induce multiple, and at times conflicting, effects under different conditions (Nader, Schafe et al. 2000, Pedreira, Perez-Cuesta et al. 2002, Walker, Brakefield et al. 2003, Debiec, Doyère et al. 2006, Forcato, Argibay et al. 2009, Sederberg, Gershman et al. 2011, Sevenster, Beckers et al. 2012), it was thus important to characterize these contributory factors. Specifically, reconsolidation is

known to be time-sensitive. The presence of this time-dependence in humans has been coarsely derived from studies utilizing either one of the two extreme reactivation-intervention intervals: either too short such that the reconsolidation was still ongoing (e.g., 5 or 10 minutes, (Forcato, Burgos et al. 2007, Forcato, Argibay et al. 2009, Schiller, Monfils et al. 2010, Agren, Engman et al. 2012)), or too long such that the reconsolidation had concluded before the intervention began (e.g., 6 or 10 hours, (Forcato, Burgos et al. 2007, Schiller, Monfils et al. 2010, Agren, Engman et al. 2012)). Here we investigated the detailed temporal characteristics of reconsolidation of declarative memory using gradient-like post-reactivation delays.

In light of the controversies surrounding theories on the reconstructive nature of declarative memories, we evinced that human associative memories can be exquisitely rendered labile by newly-acquired memories within a critical time-window. Using a face-location association learning paradigm, human participants were made to experience acquisition, test of associative-learning, reactivation, new-learning, and final-test across three consecutive days. In a behavioral experiment (Fig. 1A, upper panel), participants encoded 30 face-location associations on Day 1 (day1-Acquisition) and following a 24-hour retention period, they were then divided into 5 groups and asked to recall the associations they had acquired previously on day1 (day2-Reactivation). Importantly, the four different groups of participants received a critical time-dependent new-learning manipulation (i.e., acquiring a new location associated with the original 30 faces) whilst a fifth group acted as a control group and did not receive any new-learning. The day2-New-learning served a critical

interventional purpose, aiming at interfering the originally acquired memories during reconsolidation. On the third day (day3-Final-test), these participants were required to recall again the face-location associations they had learned on day1-Acquisition. We revealed the new-learning that occurred right after reactivation significantly diminished the memory of the originally learned associations in a time-dependent manner.

To elucidate the behavioral effects induced by the new-learning and the neural underpinnings of the reconsolidation processes, we replicated the behavioral experiment with a new group of participants performing a corresponding experiment while their blood-oxygen-level-dependent (BOLD) activity was measured. We probed, at a macro-anatomical level, in which regions might lie the influence of the new-learning on the reconsolidation of non-emotional episodic memory (i.e., how new-learning affected the originally learned memory) and how the intrusive effects thereby induced by the newly-learned associations might manifest neurally. In this fMRI experiment we included only one experimental group, which began their new-learning immediately after reactivation on Day 2 (Fig. 1B-D). We employed fMRI to unravel the mechanisms underlying the processes of integrating new information into consolidated memories during reconsolidation.



104

105 **Figure 1. Paradigm Overview.** (A) Experimental overview for behavioral and fMRI
106 experiments. There were five and two groups in the two experiments, respectively.
107 Each of the experiments, spanning across 3 daily sessions, consisted of four stages:
108 Acquisition, Test of Associative-learning (Day 1), Reactivation, New-learning, Test of

New-learning (Day 2), and Final-test (Day 3). On Day 1, the subjects acquired a set of face-location associations (Acquisition). On Day 2, they were first asked to recall original associations (Reactivation) and were then divided into 4 experimental groups and one control group. After variable delays (i.e., 0', 20', 30', and 40'), they learned another set of associations of linking a new location to each of the original faces (New-learning). The control group did not receive any new learning. Finally, on Day 3, these subjects were asked to recall the originally learned associations which they had acquired on Day 1 (Final-test). The participants in the fMRI experiment were scanned on Days 2 and 3. The cycle "×4" and "×2" denote the numbers of repetition in each of the tests. (B) Original learning (Acquisition) consisted of 30 face-to-location associations. On each trial, a unique face was presented together with a location (out of five possible locations) on the screen for 4 s. The participants were instructed to memorize the associations. Their memories were then tested with Tests of Associative-learning. No feedback was given. (C) Importantly, using the identical procedure, on Day 2, 30 new associations were acquired *de novo* by the participants in the New-learning stage. (D) In Test of Associative-learning (AT), Reactivation (RE), Test of New-learning (NT), and Final-test (FT) stages, on each trial, the participants were required to indicate the correct location matched to each of the faces by pressing a 5-button keypad. In the Final-test, each response was classified into either a Correct response (blue discs), an Intrusive error (red discs), or a Non-intrusive error (grey discs). The colored discs, the face ID numbers and the location numbers (0-4) were not shown in the actual experiment. The order of face-presentation was randomized within and across participants in all stages.

2 Method

The entire study consisted of one behavioral and one fMRI experiment. Each consisted of four experimental sessions across three days. The separation between

days was strictly controlled to be 24 hours (Fig. 1).

2.1 *Subjects*

151 participants took part in the behavioral experiment proper (four experimental groups n = 28 each; control group n = 39) and 46 participants took part in the fMRI experiment (experimental group n = 18; control group n = 28). All of them were recruited from the East China Normal University (17 – 30 years old, mean = 22.05±2.51, SD, 26 males). All had normal or corrected-to-normal vision and reported regular nocturnal sleep and no history of any neurological, psychiatric or endocrine disorder. The participants received monetary compensation for their participation. Written informed consents were obtained from all participants and the study was approved by the University committee on Human Research Protection (UCHRP) at East China Normal University. An additional 60 participants (37 and 23 for the behavioral and fMRI experiments, respectively) were recruited but were not invited to enter the subsequent sessions because their performance accuracy was below 40% in the last round of the Test of Associative-learning on Day 1.

2.2 *Stimuli*

60 grayscale front-facing faces of neutral expression from unfamous volunteers (30 males) were selected from CAS-PEAL-R1 database (http://www.jdl.ac.cn/peal/index.html). These were divided into two sets of 30 faces each (each set consisted of 15 males and 15 females). One set was used in the behavioral and fMRI experiments, in which both day1-Acquisition and day2-New-learning employing the same set of 30 faces (Fig. 1). The second set was

specifically used in the day2-New-learning phase for a control experiment wherein the 30 faces used in the new-learning were different from those used in the original learning session (Supplementary Fig. S1).

2.3 *Behavioral experiments and analysis*

The behavioral experiment investigated how the time interval between memory recall (i.e., reactivation) and interference (i.e., new-learning) affects memory reconsolidation. Four gradient-like time intervals between memory recall and interference were chosen: 0, 20 minutes, 30 minutes, and 40 minutes (four experimental groups). To obtain a reliable baseline for comparison, a control group was included in which no interference was applied (control group).

In a face-location association learning paradigm, the participants first familiarized themselves with the 30 faces on Day 1 (familiarization session) by viewing these faces passively. Each face was presented at the center of the screen for 3 s and separated by a jittered inter-trial interval of 2-4 s (mean = 3s). The whole set of 30 faces was presented three times in a randomized order.

Following the familiarization phase, the participants were then asked to memorize 30 face-location associations (day1-Acquisition; A), involving each face being paired with one of five location points on the screen. They were allowed 4 s to learn each pairing (Fig. 1B). Immediately after each acquisition of the 30 face-location pairings, a memory test ensued (Test of Associative-learning, AT, Fig. 1). On each test trial, the face cue and all five location points were presented together, and the participants were asked to indicate within 4 s which location disc was

originally paired with the face in the Acquisition stage by pressing the button corresponding to the target location using an MRI-compatible keypad (see cartoon in Fig. 1D). This Acquisition – Test of Associative-learning procedure was repeated four times with the set of face-location associations presented in a new randomized order in each cycle. The trials were separated by jittered inter-trial intervals of 3-7 s (mean = 5s) and no feedback was given.

On Day 2, the participants were asked to recall their memory of the previously learned face-location associations by identifying the target location that was associated with a given face (day2-Reactivation; RE, Fig. 1). A New-learning procedure was then administered aiming to interfere the processes of memory reconsolidation. The participants were asked to learn to associate the originally-learned faces with a new target location (i.e., learning new face-location associations, Fig. 1C). This New-learning session consisted of four cycles of New-learning (NL) and Test of New-learning (NT).

In order to pinpoint the temporal characteristics of interference on memory reconsolidation, four temporal intervals, namely 0', 20', 30', and 40', between the day2-Reactivation and New-learning were administered separately to the four experimental groups. During these post-reactivation intervals, the participants listened to light music without having to perform any task.

On Day 3, the participants recalled the face-location associations they had acquired on Day 1 (day3-Final-test; FT), identifying the target locations that were associated with given faces from Day 1.

A mixed 5 (between-group factor, four experimental conditions and control condition) × 3 (within-group factor: Day1, Day2 and Day3) analysis of variance (ANOVA) was applied on percentage correct data from the behavioral experiment. Analogously, a mixed 2 (Group Exp. and Ctrl.) × 3 (Day1, Day2 and Day3) ANOVA was applied on the data from the fMRI experiment.

Moreover, to account for inter-subject variability, the within-subjects correct rates were normalized to obtain relative correct rates using the following equations,

$$\text{Relative Correct Rate}_{2\text{-}1}=\frac{\text{Correct Rate}_{\text{Day2-Reactivation}}-\text{Correct Rate}_{\text{Day1-Acquisition}}}{\text{Correct Rate}_{\text{Day1-Acquisition}}}\times 100\%$$

$$\text{Relative Correct Rate}_{3\text{-}2}=\frac{\text{Correct Rate}_{\text{Day3-Final}}-\text{Correct Rate}_{\text{Day2-Reactivation}}}{\text{Correct Rate}_{\text{Day2-Reactivation}}}\times 100\%$$

The within-subjects relative Correct $\text{Rate}_{2\text{-}1}$ reflects the memory decay after Day1-Acquisition before Day2-Reactivation, whereas the relative Correct $\text{Rate}_{3\text{-}2}$ reflects the memory change due to the New-learning intervention.

2.4 *Classification of correct, intrusive and non-intrusive responses*

During the Final-test session, the participants were instructed to respond to the target location as they learned in the acquisition on Day 1. Since the experimental groups experienced new-learning on Day 2, there were three categories of responses in the day3-Final-test. If the response was correctly matched with acquisition, it was a correct hit. If it was incorrectly matched with the location they acquired in the new learning on Day 2, it was classified as an intrusive error. Responses made to the other three locations would be non-intrusive errors (Fig. 1D). We compared the difference

between the correct and the intrusive proportions among the groups. If the correct rate/intrusive ratio was not significantly different between the experimental groups and the control group, then we would infer that new learning did not cause any significant effect. By contrast, if there were significant differences in the correct rate/intrusion ratio between the groups, we would conclude that the new learning might have disrupted the original-memory more severely in the experimental group(s).

2.5 *Control experiment: Effectiveness of content-similarity in memory intervention*

In declarative memories, content similarity shared between the acquisition and new-learning material is a key factor for effective intervention as only similar new materials were found to induce memory update, disruption or enhancement via reconsolidation (Forcato, Burgos et al. 2007, Hupbach, Gomez et al. 2007, Coccoz, Maldonado et al. 2011, Forcato, Rodriguez et al. 2011). We hypothesized that material used in the post-reactivation intervention has to be similar enough to those used in the acquisition to cause any discernible effect on the reconsolidation processes. To test this prediction, we ran an additional control experiment in which we utilized new and unencountered faces as the post-reactivation new-learning material (i.e., new faces to be paired up with the original locations).

2.6 *MRI acquisition and preprocessing*

Participants were scanned in a 3T MRI scanner (Trio Tim, Siemens) with a quadrature volume head coil at the Shanghai Key Laboratory of Magnetic Resonance. Thirty-three slices of functional MR images were acquired using a gradient EPI

sequence (EPI volumes per run = 192, FOV = 210×210 mm^2, matrix = 64×64, in-plane resolution = 3.75×3.75 mm^2, thickness = 4 mm, without gap, repetition time = 2 s, echo time = 30 ms, flip angle = 90°), covering the entire brain. A high-resolution structural image for each participant wasp also acquired using 3D MRI sequences for anatomical co-registration and normalization (FOV = 256×256 mm^2, matrix = 256×256, slice thickness=1 mm, without gap, repetition time = 2530 ms, echo time = 2.34 ms, flip angle = 7°).

SPM8 (Wellcome Department of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm/) was used for data processing. For each participant, the functional images were realigned to correct for head movements. The structural image was co-registered with the mean EPI image, then segmented and generated normalized parameters to MNI space. Functional images were then normalized to the MNI space using these parameters, re-sampled to 2 mm isotropic voxel size and then spatially smoothed using an isotropic Gaussian kernel of 8 mm FWHM (full-width half-maximum). High-pass temporal filtering with a cut-off of 128 s was performed to remove low-frequency drifts.

2.7 *fMRI data analysis*

The fMRI experiment examined the neural correlates underlying the several aspects elicited by the new-learning interference on memory reconsolidation. The same face-location association learning paradigm as in the behavioral experiment was adopted. We implemented two Reactivation sessions on Day 2 and also two Final-test sessions on Day 3 to ensure a decent volume of data to be collected for the fMRI

267 analyses. Across the two sessions, we collected data for 60 test trials (i.e., two
268 repetitions of the complete set of the 30 face-location pairs). Informed by the
269 behavioral experiment that the memory reconsolidation processes were susceptible
270 after a 0'-delay, we accordingly targeted at the 0' condition here. We included one
271 control group, which received no new-learning after reactivation, for comparison.
272 Trials were separated by jittered inter-trial intervals of 3-7 s (mean = 5s) and 18 4-s
273 blank trials were included as baseline measurement. Each of the New-learning runs
274 lasted for 10 min and each of the Reactivation/Final-test runs lasted for 6 min (Fig.
275 1C-D). The fMRI data for the day2-New-learning were not included for analysis.

276 Two sets of analyses (Day×Group model and Intrusion model) were carried out
277 using a general linear model (**Error! Reference source not found.**). Statistical
278 inference was based on a random effects approach, which comprised first-level
279 analyses estimating contrasts of interest for each subject and second-level analyses for
280 statistical inference at the group level with non-sphericity correction. For both models,
281 in the first-level, each of the 60 test trials was modelled with a canonical
282 hemodynamic response function time-locked to the trial onset as an event-related
283 response with that trial's duration (mean duration = 2466 ms). The design matrix
284 included six head motion regressors to remove the residual effects of head motion.
285 The blank trials were not modelled. The estimated parameters values were used for
286 the second-level group analysis.

287 The first model (Day×Group model) sought to identify brain areas that activated
288 during reactivation and final-test. This allowed us to calculate the interaction effect

between the two factors for finding any evidence of episodic memory reconsolidation. In the first-level analysis, the model included five regressors: $R_{(Day2,Exp)}$, $R_{(Day3,Exp)}$, $R_{(Day2,Ctrl)}$, $R_{(Day3,Ctrl)}$, Misses, reflecting the responses of the experimental and control groups in day2-Reactivation and day3-Final-test. For the group-level analysis, the single-subjects contrast images for the 2 experimental conditions (i.e., "Day2/Day3" trials, averaged across the two fMRI-runs) for each of the two groups were entered into a mixed design ANOVA with "Day" as the within-subject variable and "Group" as the between-groups variable. The random effects analysis consisted of an ANOVA assessing the significance of Delta *T*-covariate at the group level. The statistical threshold was set to *P*-FWE=0.05, whole brain corrected at peak level (cluster size estimated at *P*-unc. = 0.005). With our *a prior* prediction, we performed small volume correction (SVC) using a functional mask derived from subsequent memory effects as the volume of interest (covering the hippocampus and the amygdala, (Kim 2011)).

The second model (Intrusion model) concerned responses during the final-test, specifically investigating how new learning affected the original memory trace during the reconsolidation process. The first-level model included three regressors obtained from the day3-Final-test, reflecting the three types of the responses (correct, intrusive or non-intrusive). Six motion regressors were also included. For the group-level analysis, the single-subjects contrast images for the 3 experimental conditions (i.e., "correct/intrusive/non-intrusive" trials, averaged across the two fMRI-runs) were entered into an ANOVA. The statistical threshold was set to *P*-FWE=0.05, whole brain corrected at peak level (cluster size estimated at *P*-unc. = 0.005). The random

effects analysis consisted of a one-sample t-test assessing the significance of Delta *T*-covariate at the group level. Specifically, the "Intrusive > Non-intrusive" contrast revealed a cluster in the left inferior frontal gyrus. We accordingly extracted the beta estimates of the left IFG from each subject using Marsbar and correlated these beta estimates with the proportion of correct responses and the proportion of intrusive errors separately.

3 Results

3.1 *Behavioral results: Main experiment*

We revealed compelling evidence in support of the existence of reconsolidation. In the behavioral experiment, the Day × Group repeated measures ANOVA on percentage correct showed a strong "Day × Group" interaction effect ($F_{(8, 292)} = 7.26$, $P < 0.001$, Fig. 2A, Supplementary Table S1). We then ran two separate ANOVAs and found the group differences were only in the day3-Final-test, ($F_{(4, 146)} = 5.11$, $P = 0.002$, Fig. 2A, Supplementary Table S1) but not in day2-Reactivation ($F_{(4,146)} = 0.39$, $P = 0.81$). In order to account for individual variances, we normalized the percentage correct data and re-ran ANOVAs on these synthetic, more sensitive indices. A 2 (correct $rate_{2\text{-}1}$; correct $rate_{3\text{-}2}$) x 5 (Group) repeated measures ANOVA equally showed a strong interaction between the factors ($F_{(4,146)} = 6.00$, $P < 0.001$). Two separate ANOVAs showed that the interaction was driven by a main effect in relative correct $rate_{3\text{-}2}$ between Days 2 and 3, confirming that the significant between-group differences were specifically caused by new-learning (relative correct $rate_{3\text{-}2}$: $F_{(4, 146)} = 9.75$, $P < 0.001$, Fig. 2B right, Supplementary Table S2) but not before

new-learning (relative correct rate$_{2\text{-}1}$: $F_{(4, 146)} = 0.39$, $P = 0.81$, Fig. 2B left).

Motivated by previous findings on the time-dependence of post-retrieval manipulations (Forcato, Burgos et al. 2007, Schiller, Monfils et al. 2010, Agren, Engman et al. 2012, Chan and LaPaglia 2013), we then tested the hypothesis that there would be a critical time-window for the observable post-reactivation reconsolidation. As expected, the difference in the relative correct rates between Day 2 and Day 3 for Group 0' and 20' were significantly lower that other three groups (all Ps < 0.05, LSD multi-comparison, Fig. 2B), indicating the influence of new-learning was indeed highly time-dependent.

It has been reported that new-learning could produce an intrusive effect to our memories by replacing the original memories in a specific retrograde manner (Hupbach, Gomez et al. 2007, Hupbach, Gomez et al. 2009, Schiller, Monfils et al. 2010). In view of this, we tested for the intrusive effect in the current context. On each trial, there were five location points; each of which could be a potential choice. Operationally, for the experimental groups, at day3-Final-test, responses made to the target location would be a hit, responses made to the newly-learned location would be an intrusive error, whereas responses made to any of the other three locations would be a non-intrusive error (see Methods). The intrusive proportion of Group 20' was significantly higher than other three groups (all Ps < 0.05, Fig. 2C, Supplementary Fig. S2, Supplementary Table S3), whereas these intrusive errors in the other three groups did not differ. Interestingly, in Group 20', the intrusive proportion did not differ from the correct rate, while in other three groups the correct rates were significantly higher

than the intrusive proportions (Supplementary Fig. S2, Supplementary Table S3), implying the new-learning might have caused differential effects on Group 0' and 20'.

We further analyzed these intrusive effects in all experimental groups. Interestingly, the quantity of intrusive errors in the 20' condition is significantly higher than those in the other conditions ($F_{(3,108)} = 5.08$, $P = 0.003$; LSD multi-comparison: $P_{(20'>0')} = 0.035$*, $P_{(20'>30')} = 0.001$**, $P_{(20'>40')} = 0.001$** vs. $P_{(0'>30')} = 0.21$, $P_{(0'>40')} = 0.23$, $P_{(30'>40')} = 0.96$), indicating the intrusive effects induced by new-learning following different post-reactivation delays are differential.

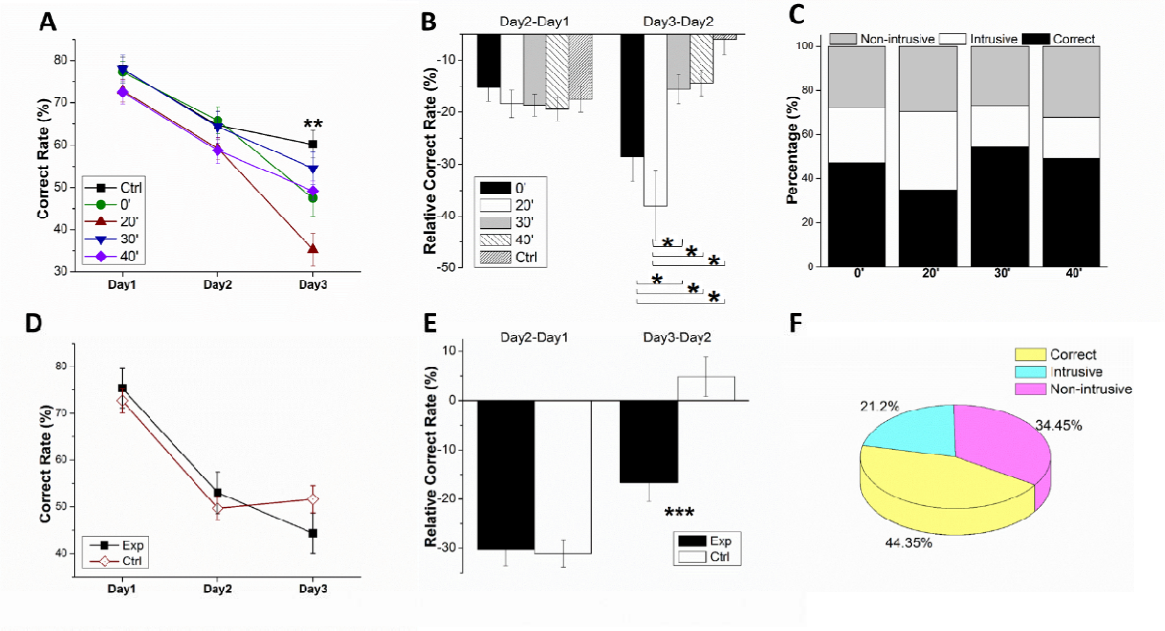


Figure 2. Behavioral results of both experiments. (A) Memory performance plotted as a function of days in the behavioral experiment. Memory of the face-to-location associations diminished in all five groups gradually across days but a main effect of Group was found on Day 3 at the Final-test. (B) The reduction in memory in the behavioral experiment was plotted for relative correct $rate_{2-1}$ and relative correct

rate$_{3-2}$, respectively. There was no difference in Group for relative correct rate$_{2-1}$ but there was a significant interaction for the relative correct rate$_{3-2}$. Post-hoc tests confirmed that the memory for the Group 0' and 20' decreased far more drastically than Group 30', 40' and the control group. (C) The intrusive proportion of Group 20' was significantly larger than other groups. (D) Behavioral result in the fMRI experiment was consistent with that of the behavioral experiment. Both Group Exp. and Ctrl. performed similarly on Day 2. But the performance of the Group Exp., who had received post-reactivation New-learning on Day 2, diminished far more severely than Group Ctrl. at the Final-test. (E) Using a relative measure, in the fMRI experiment, there was no Group difference in the relative correct rate$_{2-1}$, but Group Exp. was significantly more impaired than Group Ctrl. in the relative correct rate$_{3-2}$. (F) The intrusive proportion of Group Exp. in the fMRI experiment (21.2%) was similar as Group 0' in the behavioral experiment (cf. leftmost bar in Fig. 2C). Error bar denotes standard error of the means. * P<0.05, ** P<0.01, *** P<0.001.

3.2 *Control experiment results: Effective manipulation requires high content-similarity between acquisition and intervention*

In this control experiment, the new-face-learning caused no effect on reconsolidation. We ran a 3 × 2 repeated measures ANOVA (Day×Group) on percentage correct and found neither a group main effect nor an interaction effect (group main effect: $F_{(1,2)} = 1.49$, $p = 0.24$; interaction: $F_{(2, 2)} = 1.29$, $p = 0.29$; Supplementary Fig. S1). We conducted the post-hoc tests regardless and confirmed there were no group differences in day2-Reactivation ($t_{(1, 20)} = 0.47$, $p = 0.65$) or day3-Final-test ($t_{(1, 20)} = -0.22$, $p = 0.82$), nor in the relative correct rate$_{3-2}$ between Days 2 and 3 ($t_{(1, 20)} = -1.36$, $p = 0.19$). These indicate that new-learning using "new faces" was ineffective in causing interference in the memory traces during

396 reconsolidation.

397 **3.3 *fMRI experiment results***

398 We have thus far established in the behavioral experiment that new-learning
399 following reactivation did intrude into the already encoded, yet labile memories, and
400 produce overt changes in terms of memory behavior. We then tap into the rather
401 complicated and unresolved mechanisms of reconsolidation by means of functional
402 imaging. We replicated these behavioral patterns in the fMRI experiment with a new
403 group of participants. A 2×2 repeated measures ANOVA (Day×Group) showed an
404 interaction effect ($F_{(2, 88)} = 6.86$, $P = 0.002$, Fig. 2D). The performance for the
405 experimental group was significantly lower than that of control group in the relative
406 correct $rate_{3\text{-}2}$ ($t_{(44)} = -3.65$, $P < 0.001$, Fig. 2E right) but not in the relative correct
407 $rates_{2\text{-}1}$ ($t_{(44)} = 0.18$, $P = 0.860$, Fig. 2E left).

408 To look into the neural correlates, we ran a "Day×Group" model to test for the
409 interaction between Day and Group to look for the effects of new-learning on original
410 memory. Specifically, the interaction term ($R_{Day2,Exp}$-$R_{Day3,Exp}$) vs. ($R_{Day2,Ctrl}$-$R_{Day3,Ctrl}$)
411 revealed activation of left hippocampus and right amygdala (Fig. 3). Both regions
412 yielded significant activation (hippocampus: peak *P*-svc = 0.049; amygdala: peak
413 *P*-svc = 0.037) with small volume correction (SVC) (volume-of-interest obtained
414 from a subsequent memory effects contrast: remembered vs. forgotten)(Kim 2011).
415 Notably, the amygdala has been known to be related to emotional processes especially
416 by those that are involved in fear and threat memory reconsolidation (Agren, Engman
417 et al. 2012, Schiller, Kanen et al. 2013). However, in the present setting, considering

418 our paradigm did not contain any emotional factors, the right amygdala was
419 implicated regardless.
420
421
422
423 Table 1. Summary of all 1st-/2nd-level analyses and contrasts for both models.

Models	**First-level**	**Second-level**	
	Regressors	*Contrasts*	*Search Volume*
Day×Group	$R_{Day2,Exp}$, $R_{Day3,Exp}$, $R_{Day2,Ctrl}$, $R_{Day3,Ctrl}$, Misses	($R_{Day2,Exp}$-$R_{Day3,Exp}$) > ($R_{Day2,Ctrl}$-$R_{Day3,Ctrl}$)	SVC (hippocampus and amygdala)
Intrusion	Correct, Intrusive, Non-intrusive (including Misses)	Intrusive > Non-intrusive	Whole-brain; SVC (left IFG)

424 R refers to trials in which subjects made a response on day2-Reactivation and
425 day3-Final-test, irrespective of being correct or incorrect; Misses refer to trials of no
426 response. Corrects, Intrusive and Non-intrusive errors classification are illustrated in
427 Fig. 1D.
428

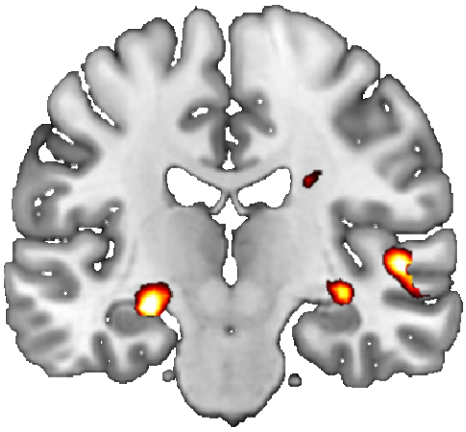


Figure 3. Neural correlates associated with the impact of new-learning on reconsolidation. Hippocampal and amygdala are differentially activated at the Final-test following New-learning administered during reconsolidation on Day 2, as given by the interaction term: $(R_{Day2,\ Exp}\text{-}R_{Day3,\ Exp}) > (R_{Day2,\ Ctrl}\text{-}R_{Day3,\ Ctrl})$; *P*-svc < 0.05.

In a separate model (Intrusion model), the contrast "Intrusive Non-intrusive errors" revealed activation of the left inferior frontal gyrus (IFG, Fig. 4). This indicated the post-reactivation new-learning was associated with activation in the inferior frontal area, which has long been implicated in resolving interference between competing mnemonic representations of the originally learned and newly acquired associations (Badre, Poldrack et al. 2005). For consistency, we also performed SVC for the IFG using a functional mask defined in a previous study (mid-ventrolateral PFC, post-retrieval selection) (Badre, Poldrack et al. 2005) and confirmed the results (peak *P*-svc = 0.010).

To elucidate the functional significance of the inferior frontal activation in

relation to the behavioral results, we extracted the beta estimates of BOLD signal from the left IFG cluster and correlated these individual beta estimates with subjects' percentage correct and intrusive proportion respectively. The individual beta estimates showed a significant positive correlation with the percentage correct rates (r = 0.48, *P* = 0.045, Fig. 4 left panel), whereas the beta estimates showed a negative correlation with the subjects' intrusive proportion (r = -0.45, *P* = 0.060, Fig. 4 right panel). We interpret these pattern of results as that the IFG is involved in mediating the recollection bias towards the originally learned information. The more strongly the inferior frontal cortex is activated, the successful the participant would be in discriminating the respective memory traces associated with the original acquisition and new-learning, whereas a weaker inferior frontal involvement signifying a lower ability in dealing with the competition between mnemonic representations of the initially-learned and newly-acquired associations.

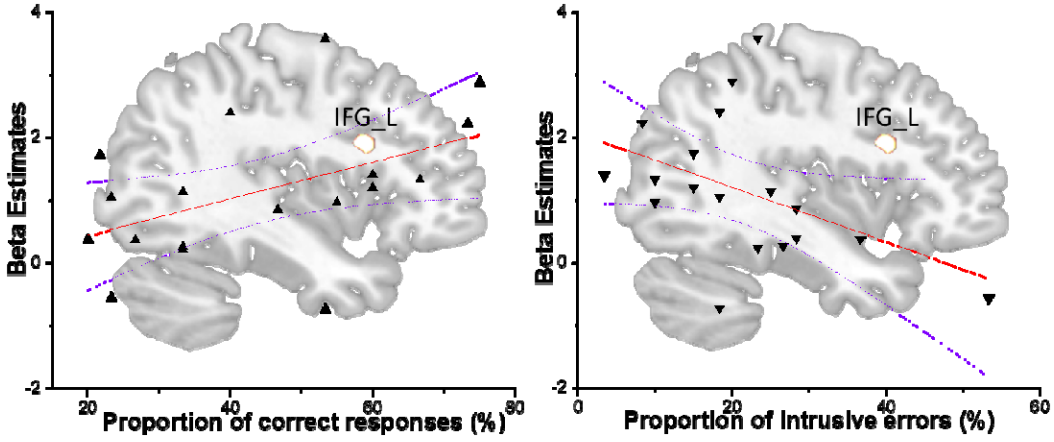


Figure 4. Engagement of IFG by new-learning intrusion. Left IFG activation

462 measured on the Final-test reflects individuals' variability in guarding against memory
463 intrusion imposed by New-learning during reconsolidation on Day 2. The left inferior
464 frontal gyrus is more activated by intrusive errors than by non-intrusive errors during
465 Final-test ($P < 0.05$). This difference in neural activation mediated the behavioral
466 performance. Activation in the left IFG across participants is correlated positively
467 with the percentage correct ($P = 0.045$, left), but is correlated in a negative trend with
468 the number of intrusive errors ($P = 0.060$, right). Such IFG activation is however not
469 correlated with the number of non-intrusive errors ($P > 0.5$, not shown). This result
470 shows that post-reactivation new-learning manipulates memory by affecting
471 reconsolidation on day 2, with the intrusion-effects being observed on day 3 in the
472 Final-test. Triangles on the scatterplots represent individual subjects. The central line
473 is the best linear fit with 90% confidence interval.
474

475 **4 Discussion**

476 In light of the previous studies which failed to observe the reconsolidation
477 process in humans and non-human animals (Cammarota, Bevilaqua et al. 2004,
478 Debiec, Doyère et al. 2006, Forcato, Argibay et al. 2009), we deduced several factors
479 which might be instrumental for the reconsolidation processes at play. In declarative
480 memories, content similarity shared between the acquisition and new-learning
481 material is a key factor for effective intervention as only similar new materials were
482 found to induce memory update, disruption or enhancement via reconsolidation
483 (Forcato, Burgos et al. 2007, Hupbach, Gomez et al. 2007, Coccoz, Maldonado et al.
484 2011, Forcato, Rodriguez et al. 2011). Based on the results of the control experiment,
485 we ascertain that the new-learning was most effective in affecting reconsolidation
486 when "same faces" were employed. We thus assert that reconsolidation could be

disrupted by post-reactivation new-learning *if and only if* the new material was similar enough to those involved in the acquisition, establishing content similarity in the associative memory traces between acquisition and new-learning to be a determinant factor. If the new-face-learning was distinct from the reactivated memory traces then these new-face-learning might have induced a different set of consolidation processes independently of the targeted reactivation.

In the rodents, any intervention disrupting memory reconsolidation is only effective when it is administered shortly after reactivation (Nader, Schafe et al. 2000, Debiec, LeDoux et al. 2002, Pedreira, Perez-Cuesta et al. 2002, Debiec and Ledoux 2004), suggesting that reconsolidation is a highly time-dependent phenomenon. In the humans, there has not been a consensus on the precise interval for this mnemonic fragility (Forcato, Burgos et al. 2007, Schiller, Monfils et al. 2010, Agren, Engman et al. 2012). Our current study incorporated a range of gradient-like post-reactivation delays. The New-learning administered within 20 minutes caused retrograde amnesia, whereas delays longer than that elicited no effect. Our results thus provide a qualifier on defining the critical time-window for post-reactivation manipulation to be effective for inducing forgetting: immediately after reactivation when memory is being updated. When the interval was long and beyond the susceptible period, the reactivated memories would become stable again and immune to any new-learning, thus no effect would be observed. This conclusion is further verified by the analyses of the intrusive effect reported in Fig. 2C, which illustrate that the differential intrusive effects induced by new-learning following different post-reactivation delays.

509 Our fMRI findings demonstrate how the memory systems might have acted
510 interactively in declarative memory reconsolidation. It is known that memory
511 reactivation will render consolidated memory (hippocampus-independent) to be
512 hippocampus-dependent again (Debiec, LeDoux et al. 2002, Kelly, Laroche et al.
513 2003, Lee, Everitt et al. 2004). Our fMRI results reveal that memory processes during
514 reconsolidation are hippocampus-dependent, strengthening the view that the
515 hippocampal and amygdala involvement change with the passage of time during
516 reconsolidation (Agren, Engman et al. 2012, Schwabe, Nader et al. 2012). When the
517 post-reactivation manipulation requiring the hippocampus (and amygdala) to process
518 new but similar information during active reconsolidation, the originally acquired
519 memories would be affected by disruption or intrusion.

520 In contrast to previous studies (Nader, Schafe et al. 2000, Debiec and Ledoux
521 2004, Lee, Di Ciano et al. 2005), the amygdala activation was presently observed in
522 the absence of emotional input or incentive factors (neutral faces □ location
523 association). We proposed two possible explanations for this: First, the faces encoded
524 by the participants might inherently carry emotional valence and collaterally engaged
525 the amygdala. However, an alternative, more nascent, account is that the amygdala
526 has a seat during declarative memories reconsolidation, irrespective of emotion
527 aspects, acting in concert with the hippocampus. We are in favor of the latter account
528 especially our results align with some recent causal evidence that the human
529 amygdala possesses a general capacity to endogenously initiate memory prioritization
530 processes of declarative memories without eliciting any subjective emotional response

(Inman, Manns et al. 2018), establishing the amygdala as an overarching operator of downstream memory processes.

The activation in the left inferior frontal gyrus was differentially increased by intrusive events, suggesting that left IFG is involved in discriminating the originally learned and newly-learned memories and deciding which memories should be reactivated according to the cue (Zhang, Feng et al. 2004, Badre, Poldrack et al. 2005, Moss, Abdallah et al. 2005, St Jacques, Olm et al. 2013). Due to the high similarity between the originally learned and newly learned memories, the participants have to recollect the episodes in greater detail to overcome the competition and meet the goal in recalling the relevant, correct memories among competitive sources. In line with the view that the left ventral PFC mediates post-retrieval selection during source recollection and decision (Badre, Poldrack et al. 2005, Badre and Wagner 2007), our findings of increased left IFG activation characterize this region as a target area for manipulating memory retrieval especially during reconsolidation. The individual difference in left IFG activation among participants further serves as an indicator of individual's ability in reconciling the mnemonic intrusion during memory reconsolidation.

5 Conclusion

Overall, we reveal three neuro-behavioral features in declarative memory reconsolidation in humans. The results provided insights into the mechanisms of episodic memory reconsolidation, suggesting that reactivation can indeed effectively trigger reconsolidation with several qualifiers. First, new-learning is effective only

553 when sharing common components with initial learning (acquisition). Second, we
554 establish the existence of a critical time-window for reconsolidation, defining it to be
555 20 minutes. Third, we show the involvement of the hippocampus and amygdala in
556 integrating newly-formed memories during reconsolidation, and with the IFG
557 resolving the mnemonic competition caused by the intrusion by newly-formed
558 memories. From a translational perspective, the present findings support the
559 possibility that non-invasive manipulation may one day make drug therapy obsolete
560 and carry important implications for educational and clinical practices in devising
561 learning strategies.

562

563 **Supplementary information** containing 2 figures and 3 tables is included.

564

Author Contributions

All authors contributed to the study design. F. S., J. W., and Y. C. conducted the behavioral experiment. F. S. and J. L. performed the fMRI experiment. F. S., Y. K., Z. W., H. W. and S. C. K. analyzed the data. F. S., Z. W., H. W. and S. C. K. wrote and approved the final version of manuscript.

Acknowledgements

We thank Qing Cai for her advice on MRI data collection and analysis.

Funding

This research was supported by National Key Fundamental Research (973) Program of China Grant 2013CB329501 (Y.K.), the Natural Science Foundation of China 31271134 & 30970968 (H.W.), Shanghai Municipal Education Commission Innovative Project 10ZZ35 (H.W.), Large Instruments Open Foundation at ECNU (H.W.), Ministry of Education of PRC Humanities and Social Sciences Research Grant 16YJC190006 (S.C.K.), STCSM Shanghai Pujiang Program 16PJ1402800 (S.C.K.), and STCSM Natural Science Foundation of Shanghai 16ZR1410200 (S.C.K.).

Declaration of Conflicting Interest

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

587

588 **Open Practices Statement**

589 The data that support the findings of this study are available from the corresponding

590 author on request.

591

592

593 REFERENCES

594

595 Agren, T., J. Engman, A. Frick, J. Bjorkstrand, E. M. Larsson, T. Furmark and M.
596 Fredrikson (2012). "Disruption of reconsolidation erases a fear memory trace in the
597 human amygdala." Science **337**(6101): 1550-1552.
598 Alberini, C. M. and J. E. LeDoux (2013). "Memory reconsolidation." Current Biology
599 **23**(17): R746-R750.
600 Badre, D., R. A. Poldrack, E. J. Pare-Blagoev, R. Z. Insler and A. D. Wagner (2005).
601 "Dissociable controlled retrieval and generalized selection mechanisms in
602 ventrolateral prefrontal cortex." Neuron **47**(6): 907-918.
603 Badre, D. and A. D. Wagner (2007). "Left ventrolateral prefrontal cortex and the
604 cognitive control of memory." Neuropsychologia **45**(13): 2883-2901.
605 Cammarota, M., L. R. Bevilaqua, J. H. Medina and I. Izquierdo (2004). "Retrieval
606 does not induce reconsolidation of inhibitory avoidance memory." Learning &
607 Memory **11**(5): 572-578.
608 Chan, J. C. and J. A. LaPaglia (2013). "Impairing existing declarative memory in
609 humans by disrupting reconsolidation." Proc Natl Acad Sci U S A **110**(23):
610 9309-9313.
611 Coccoz, V., H. Maldonado and A. Delorenzi (2011). "The enhancement of
612 reconsolidation with a naturalistic mild stressor improves the expression of a
613 declarative memory in humans." Neuroscience **185**: 61-72.
614 Coccoz, V., A. V. Sandoval, J. Stehberg and A. Delorenzi (2013). "The temporal
615 dynamics of enhancing a human declarative memory during reconsolidation."
616 Neuroscience **246**: 397-408.
617 Debiec, J., V. Doyère, K. Nader and J. E. LeDoux (2006). "Directly reactivated, but
618 not indirectly reactivated, memories undergo reconsolidation in the amygdala." Proc
619 Natl Acad Sci U S A **103**(9): 3428-3433.
620 Debiec, J. and J. E. Ledoux (2004). "Disruption of reconsolidation but not
621 consolidation of auditory fear conditioning by noradrenergic blockade in the
622 amygdala." Neuroscience **129**(2): 267-272.
623 Debiec, J., J. E. LeDoux and K. Nader (2002). "Cellular and systems reconsolidation
624 in the hippocampus." Neuron **36**(3): 527-538.
625 Forcato, C., P. F. Argibay, M. E. Pedreira and H. Maldonado (2009). "Human
626 reconsolidation does not always occur when a memory is retrieved: the relevance of
627 the reminder structure." Neurobiol Learn Mem **91**(1): 50-57.
628 Forcato, C., V. L. Burgos, P. F. Argibay, V. A. Molina, M. E. Pedreira and H.
629 Maldonado (2007). "Reconsolidation of declarative memory in humans." Learn Mem
630 **14**(4): 295-303.
631 Forcato, C., M. L. Rodriguez and M. E. Pedreira (2011). "Repeated
632 labilization-reconsolidation processes strengthen declarative memory in humans."
633 PLoS One **6**(8): e23305.
634 Forcato, C., M. L. Rodriguez, M. E. Pedreira and H. Maldonado (2010).

"Reconsolidation in humans opens up declarative memory to the entrance of new information." Neurobiol Learn Mem **93**(1): 77-84.

Gershman, S. J., A. C. Schapiro, A. Hupbach and K. A. Norman (2013). "Neural context reinstatement predicts memory misattribution." J Neurosci **33**(20): 8590-8595.

Hupbach, A., R. Gomez, O. Hardt and L. Nadel (2007). "Reconsolidation of episodic memories: a subtle reminder triggers integration of new information." Learn Mem **14**(1-2): 47-53.

Hupbach, A., R. Gomez and L. Nadel (2009). "Episodic memory reconsolidation: Updating or source confusion?" Memory **17**(5): 502-510.

Hupbach, A., R. Gomez and L. Nadel (2011). "Episodic memory updating: the role of context familiarity." Psychon Bull Rev **18**(4): 787-797.

Hupbach, A., O. Hardt, R. Gomez and L. Nadel (2008). "The dynamics of memory: context-dependent updating." Learn Mem **15**(8): 574-579.

Inman, C. S., J. R. Manns, K. R. Bijanki, D. I. Bass, S. Hamann, D. L. Drane, R. E. Fasano, C. K. Kovach, R. E. Gross and J. T. Willie (2018). "Direct electrical stimulation of the amygdala enhances declarative memory in humans." Proceedings of the National Academy of Sciences: 201714058.

Kelly, A., S. Laroche and S. Davis (2003). "Activation of mitogen-activated protein kinase/extracellular signal-regulated kinase in hippocampal circuitry is required for consolidation and reconsolidation of recognition memory." J Neurosci **23**(12): 5354-5360.

Kim, H. (2011). "Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies." Neuroimage **54**(3): 2446-2461.

Lee, J. L., P. Di Ciano, K. L. Thomas and B. J. Everitt (2005). "Disrupting reconsolidation of drug memories reduces cocaine-seeking behavior." Neuron **47**(6): 795-801.

Lee, J. L., B. J. Everitt and K. L. Thomas (2004). "Independent cellular processes for hippocampal memory consolidation and reconsolidation." Science **304**(5672): 839-843.

Lee, J. L., K. Nader and D. Schiller (2017). "An update on memory reconsolidation updating." Trends in cognitive sciences **21**(7): 531-545.

Moss, H. E., S. Abdallah, P. Fletcher, P. Bright, L. Pilgrim, K. Acres and L. K. Tyler (2005). "Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus." Cereb Cortex **15**(11): 1723-1735.

Nader, K., G. E. Schafe and J. E. Le Doux (2000). "Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval." Nature **406**(6797): 722-726.

Pedreira, M. E., L. M. Perez-Cuesta and H. Maldonado (2002). "Reactivation and reconsolidation of long-term memory in the crab Chasmagnathus: protein synthesis requirement and mediation by NMDA-type glutamatergic receptors." J Neurosci **22**(18): 8305-8311.

Schiller, D., J. W. Kanen, J. E. LeDoux, M.-H. Monfils and E. A. Phelps (2013). "Extinction during reconsolidation of threat memory diminishes prefrontal cortex

679 involvement." Proceedings of the National Academy of Sciences **110**(50):
680 20040-20045.
681 Schiller, D., M. H. Monfils, C. M. Raio, D. C. Johnson, J. E. Ledoux and E. A. Phelps
682 (2010). "Preventing the return of fear in humans using reconsolidation update
683 mechanisms." Nature **463**(7277): 49-53.
684 Schwabe, L., K. Nader, O. T. Wolf, T. Beaudry and J. C. Pruessner (2012). "Neural
685 signature of reconsolidation impairments by propranolol in humans." Biological
686 Psychiatry **71**(4): 380-386.
687 Schwabe, L. and O. T. Wolf (2009). "New episodic learning interferes with the
688 reconsolidation of autobiographical memories." PLoS One **4**(10): e7519.
689 Scully, I. D., L. E. Napper and A. Hupbach (2017). "Does reactivation trigger episodic
690 memory change? A meta-analysis." Neurobiol Learn Mem **142**(Pt A): 99-107.
691 Sederberg, P. B., S. J. Gershman, S. M. Polyn and K. A. Norman (2011). "Human
692 memory reconsolidation can be explained using the temporal context model."
693 Psychon Bull Rev **18**(3): 455-468.
694 Sevenster, D., T. Beckers and M. Kindt (2012). "Retrieval per se is not sufficient to
695 trigger reconsolidation of human fear memory." Neurobiol Learn Mem **97**(3):
696 338-345.
697 St Jacques, P. L., C. Olm and D. L. Schacter (2013). "Neural mechanisms of
698 reactivation-induced updating that enhance and distort memory." Proc Natl Acad Sci
699 U S A **110**(49): 19671-19678.
700 Strange, B. A., M. C. Kroes, J. E. Fan and R. J. Dolan (2010). "Emotion causes
701 targeted forgetting of established memories." Front Behav Neurosci **4**: 175.
702 Walker, M. P., T. Brakefield, J. A. Hobson and R. Stickgold (2003). "Dissociable
703 stages of human memory consolidation and reconsolidation." Nature **425**(6958):
704 616-620.
705 Zhang, J. X., C. M. Feng, P. T. Fox, J. H. Gao and L. H. Tan (2004). "Is left inferior
706 frontal gyrus a general mechanism for selection?" Neuroimage **23**(2): 596-603.
707