

Emotional arousal amplifies competitions across goal-relevant representation:

A neurocomputational framework.

Michiko Sakaki<sup>1 2\*</sup>, Taiji Ueno<sup>3\*</sup>, Allison Ponzio<sup>4</sup>,

Carolyn Harley<sup>5</sup>, Mara Mather<sup>4</sup>

<sup>1</sup>School of Psychology and Clinical Language Sciences, University of Reading, UK.

<sup>2</sup>Research Institute, Kochi University of Technology, Japan.

<sup>3</sup>Faculty of Human Sciences, Takachiho University, Japan.

<sup>4</sup>Davis School of Gerontology, University of Southern California, USA.

<sup>5</sup>Psychology Department, Memorial University of Newfoundland, Canada

\* Equal contribution

Corresponding authors:

Michiko Sakaki: [m.sakaki@reading.ac.uk](mailto:m.sakaki@reading.ac.uk)

Taiji Ueno: [taijiueno7@gmail.com](mailto:taijiueno7@gmail.com)

Accepted in "*Cognition*" (Feb, 2019)

Acknowledgements: The authors declare no conflict of interest. This work was supported by grants from the National Institute on Aging (R01AG025340), the European

Commission (FP7-PEOPLE-2013-CIG), and the Japan Society for the Promotion of Science (JP16H05959; JH16H06406; JP16H02053; JP15K21062; JP18K13374).

### Abstract

Emotional arousal often facilitates memory for some aspects of an event while impairing memory for other aspects of the same event. Across three experiments, we found that emotional arousal amplifies competition among goal-relevant representations, such that arousal impairs memory for multiple goal-relevant representations while enhancing memory for solo goal-relevant information. We also present a computational model to explain the mechanisms by which emotional arousal can modulate memory in opposite ways via the local/synaptic-level noradrenergic system. The model is based on neurophysiological observations that norepinephrine (NE) released under emotional arousal is locally controlled by glutamate levels, resulting in different NE effects across regions, gating either long-term potentiation or long-term depression by activating different adrenergic receptors depending on NE concentration levels. This model successfully replicated behavioral findings from the three experiments. These findings suggest that the NE's local effects, rather than broad effects, are key in determining the effects of emotion on memory.

*Keywords:* norepinephrine, emotion and memory, retrograde amnesia, neural network model

While our memory system does not allow us to remember all the events we experience in life, we typically remember emotional events for a long time. Indeed, decades of research reveals that emotional events are remembered better than neutral events (LaBar & Cabeza, 2006; Talmi, 2013). However, there has been little consensus on how emotion modulates memory in a retrograde manner, with some studies showing retrograde enhancement effects of emotion on memory while others show retrograde impairment effects of emotion on memory (for reviews see Chiu, Dolcos, Gonsalves, & Cohen, 2013; Dolcos, Katsumi, Denkova, & Dolcos, 2017; Dolcos, Katsumi, Weymar, et al., 2017; Mather & Sutherland, 2011). In the current paper, we investigate the mechanisms by which emotion enhances some memories while impairing others.

A predominant view in the literature on emotion and memory is that encountering emotional events induces short-term, phasic arousal reactions and triggers amygdala activity, which facilitates memory for the emotional events (Dolcos, LaBar, & Cabeza, 2004; Kensinger & Corkin, 2004; LaBar & Cabeza, 2006). Consistent with this view, research indicates that norepinephrine (NE) released from the locus coeruleus (LC) under phasic arousal interacts with  $\beta$ -adrenergic receptors in the amygdala to enhance memory for emotional information (for reviews see Markovic, Anderson, & Todd, 2014; McIntyre, McGaugh, & Williams, 2012).

While these accounts primarily focus on how emotional events are remembered well, they also predict that emotion has retrograde impairing effects on memory for neutral preceding events: enhanced memory for emotional information would leave fewer resources available for other information, resulting in impairment for preceding events. Consistent with this prediction, when task-irrelevant emotional stimuli are

presented during working memory tasks, these emotional stimuli are remembered well at the expense of impaired working memory performance for preceding information (Dolcos et al., 2013; Dolcos & McCarthy, 2006; Shafer & Dolcos, 2012). Previous studies extend these findings to long-term memory and demonstrate that exposure to emotional events leads to impaired memory for what has been learned before the emotional events (Hurlemann et al., 2005; Loftus & Burns, 1982; Miu, Heilman, Opre, & Miclea, 2005; Strange, Hurlemann, & Dolan, 2003). These retrograde impairing effects of arousal were diminished by  $\beta$ -adrenergic receptor blockers (Strange et al., 2003), suggesting that NE is involved not only in the memory enhancement effects for emotional information but also in the arousal-induced retrograde amnesia effects.

However, other studies show opposite results: exposure to emotionally arousing events after learning something neutral enhances memory of the previously learned experiences (A. K. Anderson, Wais, & Gabrieli, 2006; Finn & Roediger, 2011; Nielson & Powless, 2007; Nielson, Yee, & Erickson, 2005). These retrograde enhancement effects of arousal are also related to noradrenergic mechanisms (Roosendaal, Castello, Vedana, Barsegyan, & McGaugh, 2008). For example, arousal induced by squeezing a handgrip or vagus nerve stimulation enhances memory for previously learned neutral materials (Clark, Naritoku, Smith, Browning, & Jensen, 1999; Nielson, Radtke, & Jensen, 1996). Critically, the handgrip squeeze procedure increases pupil dilation (an indicator of LC activity) as well as peripheral NE levels (Nielsen & Mather, 2015) and its modulation effect on memory is attenuated by  $\beta$ -adrenergic receptor blockers (Nielson & Jensen, 1994). Likewise, vagus nerve stimulation increased NE levels in the basolateral amygdala (Hassert, Miyashita, & Williams, 2004). Thus, it appears that noradrenergic

mechanisms are involved in the retrograde effects due to emotional arousal, irrespective of whether arousal leads to memory enhancement or retrograde amnesia effects.

In summary, previous studies provide a confusing set of results concerning the retrograde effects of emotional arousal on memory as well as the role of noradrenergic systems in the emotion-memory interaction. Recent theoretical frameworks address this puzzle by positing that stimulus priority serves as a boundary condition between the emotion-induced facilitative vs. impairing effects on memory (Levine & Edelstein, 2009; Mather & Sutherland, 2011). Priority is determined by information's goal-relevance and bottom-up saliency. Recent research also provides evidence consistent with predictions from these theoretical perspectives. According to these studies, arousal induced by encountering emotional events enhances memory for what has happened earlier when the event is goal-relevant and therefore has high priority, but impairs memory for preceding events when the events are goal-irrelevant and have low priority (Knight & Mather, 2009; Lee, Greening, & Mather, 2015; Ponzio & Mather, 2014; Sakaki, Fryer, & Mather, 2014).

How can then arousal have opposing effects at the same time, with enhancing memory representations for high-priority information while suppressing everything else? To address this question, the Glutamate Amplifies Noradrenergic Effects (GANE) framework was recently proposed (Mather, Clewett, Sakaki, & Harley, 2016). GANE posits that glutamate is the brain's primary excitatory neurotransmitter and signals priority, such that neurons that process high-priority information release a larger amount of glutamate than those that process low-priority information. When NE is released

under phasic arousal, NE further stimulates glutamate release via  $\beta$ -adrenergic receptors and glutamate also stimulates NE release. This positive feedback loop between glutamate and NE results in enhanced neural representations for high-priority information under arousal. At the same time, high glutamate signals and high NE signals around high-priority representations lead to stronger GABAergic activity to suppress lower priority representations. In addition, high NE levels near high-priority representations enable to activate low-affinity  $\beta$ -adrenergic receptors to facilitate long-term potentiation (LTP) to enhance memory for high-priority information. In contrast, lower NE levels elsewhere leads to activation of high-affinity  $\alpha$ 1-adrenergic receptors that facilitate long-term depression (LTD). Thus, according to GANE, arousal's opposite effects on memory are explained by a) the selective enhancement of high-priority signals due to the positive feedback loop between glutamate and NE and b) the NE's local action to enable LTP only for high-priority signals.

Importantly, GANE also suggests that there is a boundary condition for the enhancement effects of arousal on memory for high-priority representations. According to GANE, the enhancement effect of arousal on memory for high-priority representations should be limited to situations where there is only one dominant high-priority item; when individuals encounter multiple similarly high-priority items, GANE predicts *the impairing effects* of arousal, rather than enhancement effects. Specifically, when there are many similarly high-priority items, neurons that process high-priority items should release a relatively large amount of glutamate, leading to higher NE levels and higher activity under arousal; but the higher activity level associated with these representations should result in increased GABAergic activity to suppress competing

representations. Thus, these representations would mutually suppress each other and have lower levels of NE that activate only  $\alpha$ 1-adrenergic receptors, resulting in LTD and impaired memory under arousal. In summary, it is expected that arousal enhances memory for solo high-priority representations but impairs memory for multiple competing high-priority representations.

In the current paper, we first tested this prediction by behavioral experiments. To further examine whether our behavioral results are consistent with GANE, we extended a recently developed computational model based on GANE (Lee et al., 2018) and tested whether the same computational model can reproduce the behavioral results.

### **Studies 1 and 2**

To test whether arousal impairs memory for multiple high-priority representations, Studies 1 and 2 used similar procedures. During the experiment, participants first completed a fear conditioning phase, where they learned associations between one tone and electric stimulation (a fear-conditioned tone: CS+) and associations between another tone and lack of stimulation (a neutral-conditioned tone; CS-). This conditioning phase was followed by a learning phase, where participants learned target face images, followed by CS+ or CS- (i.e., arousal manipulation). To address our main goal of the studies, we manipulated the number of target stimuli participants needed to remember. In one condition, participants were shown one face and three objects and asked to remember the face while ignoring the objects (i.e., solo face condition). In the other condition, they were shown four faces and asked to remember all of them (i.e., multiple face condition). The learning phase was followed by a final memory test on all face images. According to GANE, arousal should enhance memory for goal-relevant stimuli

presented alone but should impair memory if those same goal-relevant stimuli were presented with competing equally goal-relevant stimuli. Thus, we expected that CS+ would enhance memory for faces in the solo face condition but impair memory for faces in the multiple face condition.

## Method

**Participants.** Forty-five participants (11 males;  $M_{age} = 20.47$ ,  $SD = 1.62$ ) took part in Study 1 and 50 participants (13 males;  $M_{age} = 21.62$ ,  $SD = 3.31$ ) took part in Study 2 for either course credit or \$15/hour. The sample size of Study 1 was determined by the effect size estimated from a previous study (Sakaki et al., 2014). The sample size of Study 2 was determined based on the effect sizes in the previous study (Sakaki et al., 2014) and Study 1. Participants were not screened for mental health disorders.

**Stimuli.** All images were grayscale without any background objects or scenes. Targets were images of 160 celebrity faces (80 males and 80 females) obtained from the Internet (see Supplemental materials for a list of celebrities used in the study). For each of the 160 celebrities, two different images were included: these two images depicted the same person at a similar age, but were different in a range of characteristics, including orientation, hair styles, gaze directions, and mouth shapes (e.g., open mouth vs. close mouth). None of them had clear emotional expressions. One from each pair was presented as a memorandum during initial presentation and the other was used in a working memory test.<sup>1</sup> Images that were used as memoranda were counterbalanced across participants.

The 160 celebrities were pseudo-randomly assigned to one of the five conditions (CS+/solo, CS-/solo, CS+/multiple, CS-/multiple, and foils in a final memory test) so that

each condition had an equal number of female and male faces. The assignment of each celebrity to the five conditions was counterbalanced across participants. Among the 32 celebrities assigned to each of the four learning conditions (CS+/solo, CS-/solo, CS+/multiple, and CS-/multiple), eight celebrities were randomly chosen and used as foils in the working memory test. The remaining 24 celebrities were randomly assigned to one of the three working memory (WM) test item type conditions (see Procedures for details) for each participant. An additional 12 celebrity faces were used in booster trials, where participants received an unconditioned stimulus (US) after the CS+ (CS+ with shock trials).

An additional 144 celebrity faces (72 males and 72 females) were used as distractors in the multiple face condition. Each target face assigned to the multiple face condition was pseudo-randomly paired with three other faces so that each trial included two females and two males; this allowed us to ensure that the priority of each face due to its sex was similar across all four faces presented in each trial. We did not control other features (e.g., age, race).

In the solo face condition, we used 144 objects as distractors. They were obtained from previous research (Brady, Konkle, Alvarez, & Oliva, 2008; Kensinger, Garoff-Eaton, & Schacter, 2006; Sakaki et al., 2014) as well as the Internet. They were chosen from a wide range of categories, such as animals, kitchen utensils, tools, instruments, clothes, plants, food, stationary, vehicle, furniture pieces, buildings and so on. Each target face assigned to the solo condition were randomly paired with three objects. We also used an additional 15 objects and 15 celebrity faces as distractors in booster trials.

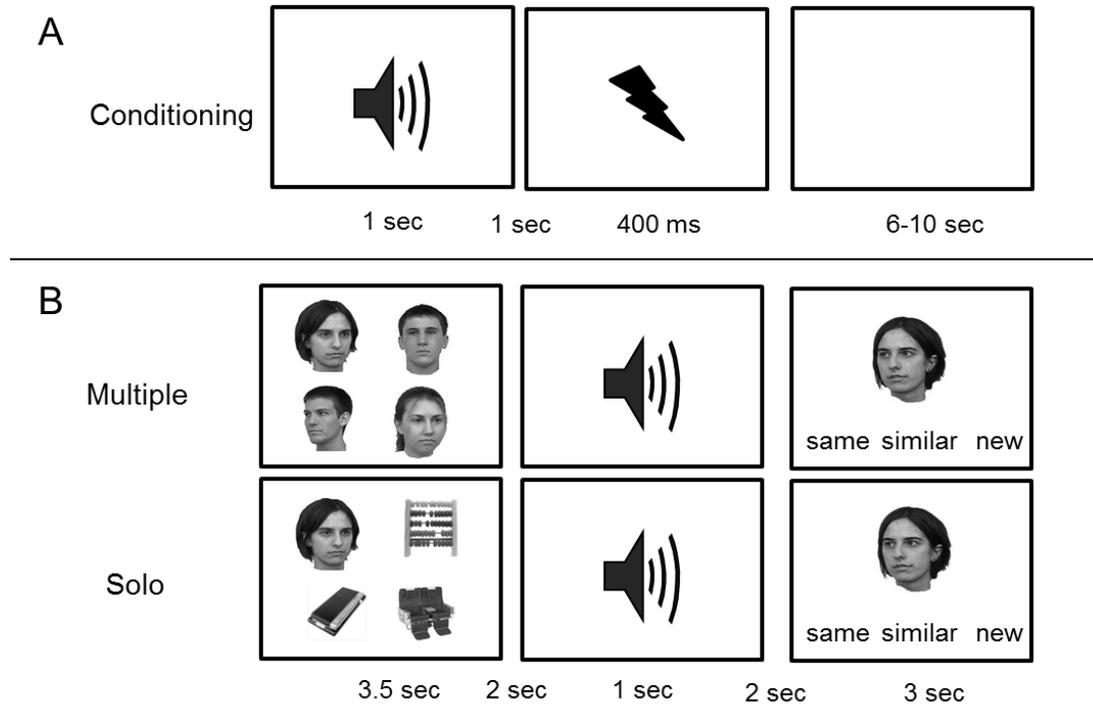
**Procedures.** After giving consent, participants completed questionnaires on their mood (Watson, Clark, & Tellegen, 1988), depression (Radloff, 1977) and anxiety (Spielberger, Gorsuch, & Lushene, 1970) and the Wechsler Test of Adult Reading (Wechsler, 2001); the results from these measures will not be reported in the paper.

Participants next completed the fear-conditioning phase whose procedure was developed based on previous studies (Lee, Baek, Lu, & Mather, 2014; Lee et al., 2015; Lee, Sakaki, Cheng, Velasco, & Mather, 2014). Conditioned stimuli (CS) were two neutral tones that were different in pitch; which tone was used as CS+ was counterbalanced across participants. The US was electric stimulation. The fear conditioning task included 36 trials in a randomized order (12 trials for CS+ with shock, 12 trials for CS+ without shock and 12 trials for CS-). On each trial (Figure 1A), participants were presented with CS+ or CS- for 1 sec. To ensure that participants paid attention to tones, they were asked to press a key to indicate whether the tone was high- or low- pitched. After a 1-sec blank screen, participants received a shock for 0.4 sec in the CS+ with shock trials. In contrast, participants did not receive a shock in other trials. Each trial ended with a jittered fixation cross (6, 8, or 10 sec). Prior to the task, participants were informed which tone was predictive of shock but were not informed the probability of shock. To confirm that participants had acquired fear-conditioned responses to CS+, we monitored participants' skin conductance reactions (SCR) during the fear-conditioning task.

After completing the conditioning phase, participants completed a questionnaire on their mood (Watson et al., 1988; which will not be reported in the paper), followed by a learning phase developed based on previous research (Figure 1B; Dolcos &

McCarthy, 2006). On each trial, participants saw four images for 3.5 sec: one celebrity face with three objects in the solo face condition and four celebrity faces in the multiple face condition. In the solo face condition, participants were asked to remember the face and ignore three objects. In contrast, in the multiple face condition, they were asked to remember all four faces. After a 2-sec blank screen, participants were presented with CS- or CS+ for 1 sec, followed by a WM test image. The WM test image was identical to the one of the memoranda on one-third of the trials (same condition), was different but depicted the same person as one of the memoranda on another one-third of the trials (similar condition), and depicted a new person not shown as memoranda on the remaining trials (new condition). Participants were told to select the “similar” option when the test image was different in any feature from those in the memoranda but depicted the same person. Participants indicated whether the image was “same,” “similar” or “new” by key press within 3 sec; if they failed to press a key within 3 sec, the image automatically disappeared.

The learning phase included eight trials in each condition from a 2 (arousal: CS+, CS-) x 2 (study set type: solo, multiple) x 3 (WM test item type: same, similar, new) design. During these 96 trials, participants did not receive shock even when they were presented with CS+. To prevent extinction, we included an additional nine booster trials, where participants received a 0.4-sec shock 1 sec after CS+. Data from these booster trials were not included in our analyses. The total 105 trials were divided into 3 blocks, each of which included 3 booster trials and 32 main trials in a randomized order.



**Figure 1.** Procedures in experiments. (A) In the conditioning phase, participants learned associations between a neutral tone and electric stimulation (CS+) and associations between another tone and lack of stimulation (CS-). (B) During the learning phase in Studies 1 and 2, participants learned four faces in the multiple condition or one face in the solo condition, followed by CS+ or CS-. Face images were taken from the TarrLab face database for illustration purpose only (stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition, Carnegie Mellon University, <http://www.tarrlab.org>).

The learning phase was followed by a demographic questionnaire and a surprise old-new recognition test for faces. This final memory test included 96 old faces (one face image from each encoding trial) and 32 new faces (foils). For each face, participants first indicated whether each image was old or new without any time limit. In Study 1, participants further indicated their confidence with a 4-point scale (1 = definitely new, 2 = maybe new, 3 = maybe old, and 4 = definitely old). The confidence rating was not included in Study 2. In this final memory test, for old stimuli, we used images used

as memoranda during the learning phase (rather than those presented during the WM test). The final memory test did not include objects in either Study 1 or Study 2.

**Electric shock and skin conductance.** Electric stimulation was delivered to the third and fourth fingers of the left hand as US via a shock stimulator (E13-22; Coulbourn Instruments, Allentown, PA) which was connected to a grounded RF filter. Prior to the experiment, we determined the intensity of “highly unpleasant but not painful” electric stimulation for each participant (Study 1:  $M = 1.31$  mA;  $SD = 0.37$ ; Study 2:  $M = 1.52$  mA;  $SD = 0.23$ ). The level determined was used throughout the experiment as the US. SCR data were recorded at 1k Hz sampling rates through the MP-150 system (BIOPAC system, Goleta, CA).

SCR data epochs were extracted from a time window between 0 and 8 sec after CS tone onset, and baseline-corrected between 0 and 1 sec. The peak SCR amplitude was taken between 1 and 8 sec from the trial-by-trial average SCR epoch as a function of CS tone. To examine the effects of conditioning, rather than shock itself, our SCR analysis focused on SCR between the CS+ without shock condition and the CS- condition. Due to technical problems, SCR data from one participant were not recorded; thus data from this participant were not included in the SCR analysis.

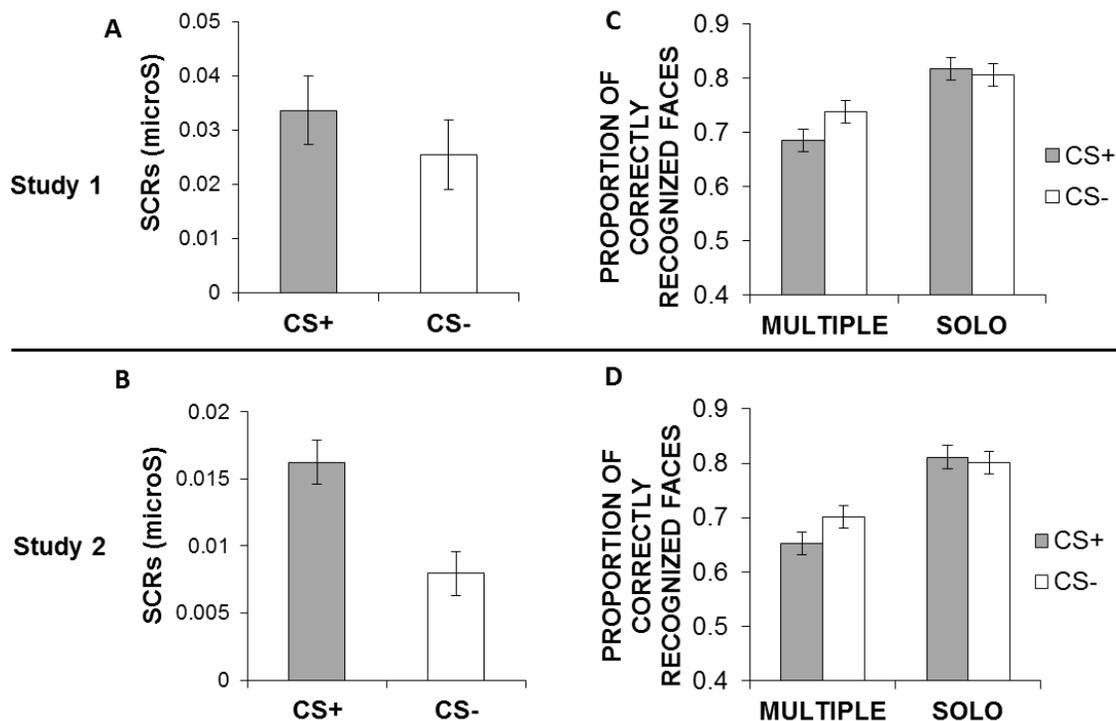
## Results

**Fear conditioning.** Participants showed high accuracy in discriminating the high- vs. low- pitched tones during the conditioning phase with no significant differences across the conditions ( $ps > .10$ ; Table 1). In addition, participants had greater SCR to CS+ than to CS- (Figure 2A-B),  $t(43) = 3.19$ ,  $p = .0027$ ,  $d = .48$ ,  $t(49) = 4.72$ ,  $p < .0001$ ,  $d = .67$ , for Studies 1 and 2, respectively, indicating that they acquired arousal

responses to CS+.

**Table 1. Accuracy of participants' response in the tone discrimination task during the fear conditioning phase.**

	Studies		
	1	2	3
CS+ with shock	0.99	1	0.96
CS+ without shock	0.99	1	0.97
CS-	0.99	0.99	0.96



**Figure 2.** Results from Studies 1 and 2. (A-B) CS+ induced higher SCRs than did CS- during the conditioning phase. (C-D) During the final memory test, there were selective impairment effects of emotional arousal on memory in the multiple condition. Error bars represent standard errors of means.

**Memory test.** We next tested the main prediction of the study that arousal should enhance memory for solo goal-relevant items but impair memory for multiple goal-relevant items. To examine the effects of these factors with initial working memory performance equated, we focused on faces from trials where participants correctly answered the WM test. Two participants in Study 1 and three participants in Study 2 were at chance for foils ( $M < .55$ ) in the final memory test; data from these participants were excluded when analyzing results in the final memory test. The remaining participants were significantly better than chance in responding to foils (Study 1:  $M = .79$ , Study 2:  $M = .84$ ),  $t(42) = 16.31$ ,  $p < .0001$ ,  $d = 2.49$ ,  $t(46) = 21.70$ ,  $p < .01$ ,  $d = 3.16$ , for Studies 1 and 2 respectively.

Two 2 (arousal: CS-, CS+) x 2 (study set type: solo, multiple) x 3 (WM test item type: same, similar, new) ANOVAs were performed on the proportion of correctly recognized faces in the final memory test in Studies 1 and 2 respectively. In both studies, we found main effects of study set type,  $F(1, 42) = 33.60$ ,  $p < .0001$ ,  $\eta^2 = .45$ ,  $F(1, 46) = 51.32$ ,  $p < .0001$ ,  $\eta^2 = .55$ , and of WM test item type,  $F(2, 83) = 31.65$ ,  $p < .0001$ ,  $\eta^2 = .42$ ,  $F(2, 91) = 25.54$ ,  $p < .0001$ ,  $\eta^2 = .40$ , and an interaction between WM test item type and study set type,  $F(2, 83) = 8.89$ ,  $p = .0003$ ,  $\eta^2 = .16$ ,  $F(2, 91) = 9.05$ ,  $p = .0003$ ,  $\eta^2 = .16$ , such that having multiple faces significantly impaired memory performance with test items that were similar (Study 1:  $M_{\text{solo}} = .81$ ,  $M_{\text{mult}} = .70$ ; Study 2:  $M_{\text{solo}} = .84$ ,  $M_{\text{mult}} = .72$ ) or new (Study 1:  $M_{\text{solo}} = .77$ ,  $M_{\text{mult}} = .60$ ; Study 2:  $M_{\text{solo}} = .76$ ,  $M_{\text{mult}} = .53$ ),  $F_s(1, 83) = 16.20$ ,  $38.61$ ,  $p_s < .0001$ ,  $d_s = .49$ ,  $1.10$ ,  $F_s(1, 91) = 14.97$ ,  $52.88$ ,  $p_s < .0001$ ,  $d_s = .48$ ,  $1.04$ , but not with those that were the same as the target ( $p$

= .60,  $p > .20$ ; Study 1:  $M_{\text{solo}} = .85$ ,  $M_{\text{mult}} = .83$ ; Study 2:  $M_{\text{solo}} = .77$ ,  $M_{\text{mult}} = .81$ ). This interaction is not surprising given that even in the multiple face condition, participants had a second chance to relearn the images during the WM test in the same condition, which could help their performance in the final test. More importantly, in both studies, we found a significant interaction between arousal and study set type (Figures 2C-D),  $F(1, 42) = 4.50$ ,  $p = .04$ ,  $\eta^2 = .08$ ,  $F(1, 46) = 4.66$ ,  $p = .04$ ,  $\eta^2 = .12$ . Consistent with our prediction, in the multiple face condition, CS+ significantly impaired memory for faces compared with CS-,  $F(1, 42) = 5.97$ ,  $p = .02$ ,  $d = .32$ ,  $F(1, 46) = 6.38$ ,  $p = .02$ ,  $d = .32$ . In contrast, there were no significant effects of arousal in the solo face condition in either study (Study 1:  $p = .48$ ,  $d = .11$ ; Study 2:  $p = .60$ ,  $d = .09$ ).

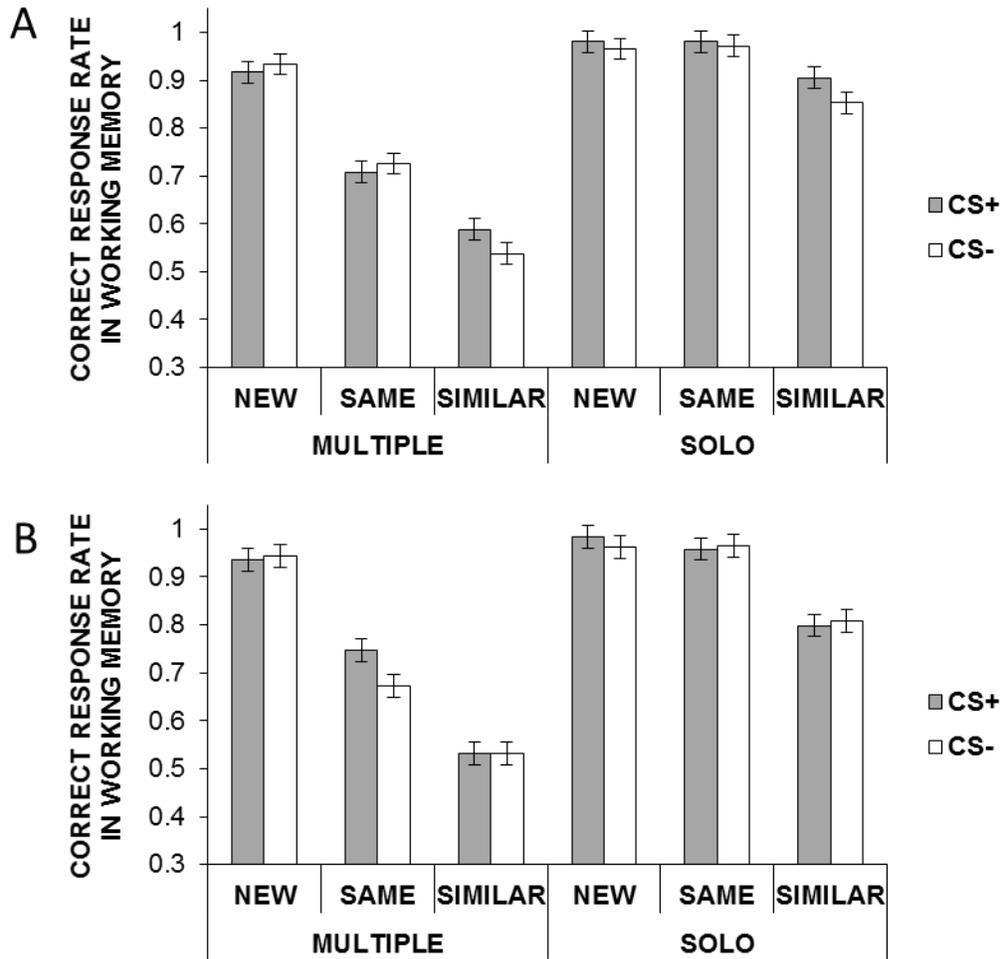
A similar 2 (arousal) x 2 (study set type) x 3 (WM test item type) ANOVA on the confidence ratings in Study 1 confirmed the effects of study set type,  $F(1, 42) = 39.22$ ,  $p < .0001$ ,  $\eta^2 = .48$ , and of WM test item type,  $F(2, 83) = 31.12$ ,  $p < .0001$ ,  $\eta^2 = .43$ , and an interaction between them,  $F(2, 83) = 18.35$ ,  $p < .0001$ ,  $\eta^2 = .29$ . But neither the interaction between arousal and study set type ( $p = .11$ ), nor any other effects involving arousal were significant ( $ps > .19$ ).

**Working memory test.** While our main focus is on final memory performance, according to GANE, a similar interaction between arousal and the number of goal-relevant items can occur even for participants' ability to briefly maintain presented stimuli due to the positive feedback loop between glutamate and NE. To address this issue, each response was coded as a correct response when the participants made "same" responses for exact same faces as the one shown as memoranda, "similar" responses for similar images that depicted the same person as memoranda and "new"

responses for images of individuals that were not shown as memoranda.

This correct response rate during the WM task was analyzed by separate 3 (arousal) x 2 (study set type) x 3 (WM test item type) ANOVAs for each study (Figures 3A-B).  $F$ -,  $t$ - and effect size values are reported below for Studies 1 and 2 in order for each effect. Across both studies, there were significant main effects of study set type,  $F(1, 44) = 212.21, p < .00001, \eta^2 = .86$ ,  $F(1, 49) = 238.51, p < .00001, \eta^2 = .86$ , and of WM test item type,  $F(2, 88) = 79.51, p < .00001, \eta^2 = .64$ ,  $F(2, 98) = 78.90, p < .00001, \eta^2 = .62$ . We also found a significant interaction between study set type and WM test item type,  $F(2, 88) = 35.44, p < .00001, \eta^2 = .64$ ,  $F(2, 98) = 45.16, p < .00001, \eta^2 = .46$ . But the interaction between study set type and arousal was not significant ( $p = .41$ ).

In addition to these similar results across studies, in Study 1, we found a significant interaction between arousal and WM test item type,  $F(2, 88) = 3.90, p = .02, \eta^2 = .08$ . Subsequent analyses revealed that arousal induced by CS+ enhanced performance in the similar condition ( $M_{cs+} = .75$  vs.  $M_{cs-} = .70$ ),  $F(1, 88) = 8.90, p = .004, d = .37$ , but not in the other two conditions ( $ps > .60$ ). These results suggest that emotional arousal induced by CS+ enhances the ability to discriminate similar inputs, thus facilitating pattern separation. We also obtained a pattern separation index (Yassa et al., 2011): [ $p$ (“similar”|similar items) –  $p$ (“similar”|old items)] and confirmed better pattern separation index score in the CS+ than in the CS- condition in Study 1,  $F(1, 44) = 4.13, p = .048, \eta^2 = .09$ , but not in Study 2 ( $p = .82$ ).



**Figure 3.** Results from the working memory test in Studies 1 (A) and 2 (B). Error bars represent standard errors of means.

**Discussion**

According to GANE, arousal should impair memory for multiple goal-relevant items. Consistent with this prediction, in Studies 1 and 2, emotional arousal induced by CS+ impaired memory for faces when there were four high-priority faces that needed to be remembered. However, contrary to our prediction, CS+ did not significantly enhance memory for solo high-priority faces. Thus, our results from these two experiments are partially consistent with the prediction of GANE.

Neither Study 1 nor Study 2 showed a significant interaction between study set type and arousal during the WM test, suggesting that the effects of arousal proposed in GANE might be stronger after a delay due to the role of NE on the long-term potentiation (LTP) and the long-term depression (LTD; Salgado, Köhr, & Treviño, 2012). In addition, in Study 1, we found a significant interaction between arousal and WM test item type during WM test, such that arousal enhanced performance in discriminating similar inputs. Recent research suggests that emotional arousal enhances pattern separation performance (Leal, Tighe, Jones, & Yassa, 2014; Segal, Stark, Kattan, Stark, & Yassa, 2012). Our results support this notion that emotional arousal facilitates one's ability to discriminate similar inputs, although the valence-by-WM test item type interaction was not significant in Study 2 ( $p = .27$ ) nor the effects of arousal on the pattern separation index ( $p = .82$ ).

### **Study 3**

Results from Studies 1 and 2 are consistent with the prediction from GANE that arousal impairs memory for multiple goal-relevant information. However, the arousal-induced memory enhancement effects in the solo condition were not significant. In Study 3, we increased priority of faces in the solo face condition to increase the effect size of this enhancement effect. To increase the priority of faces in the solo face condition, we a) added a preceding spatial cue to indicate the location of goal-relevant images as done in a prior study (Lee et al., 2015), b) presented four objects in the multiple condition, rather than faces, to make faces in the solo face condition stand out within the experiment context, and c) highlighted goal-relevant images by a red box. We also tested participants' memory for objects to see the effects of emotional arousal on

those representations.

## Method

**Participants.** Forty participants (15 males;  $M_{age} = 21.05$ ,  $SD = 2.12$ ) took part in the experiment for either course credit or \$15/hour. The sample size was determined based on the effect size in another previous study (Lee et al., 2015) which used similar procedures (i.e., presenting spatial cues to indicate the location of subsequent high priority stimuli). As in Studies 1-2, participants were not screened for mental health disorders.

**Stimuli. Faces.** Images of 80 celebrities (40 females and 40 males) were chosen from those used in Studies 1 and 2. For each celebrity, we included two different images: one of them was used in the initial presentation as a memorandum and the other was used in a working memory test (the image used as a memorandum was counterbalanced across participants).

We created two stimulus sets (each included 20 females and 20 males). One of them was used in the CS+ condition, the other was used in the CS- condition and the assignment was counterbalanced across participants. Among the 40 faces assigned to each condition, 24 were used as targets in the learning phase, another eight were used as foils in the WM task and the remaining eight were used as foils in the final memory test; the assignment of each face to targets vs. foils was determined pseudo-randomly for each participant so that all conditions had the same number of females and males. Each of the 24 faces chosen as target were further pseudo-randomly assigned to one of the three WM test item type conditions for each participant while matching the number of male and female images across the conditions. Another eight faces (4 males and 4

females) were used in booster trials.

**Objects.** We used 208 pairs of photo objects that shared the same verbal label but had different perceptual features (e.g., color, shape and orientation). They were obtained from previous research (Brady et al., 2008; Kensinger et al., 2006; Sakaki et al., 2014) and the Internet. As we did for faces, from each image pair, one of them was used in the initial presentation as a memorandum and the other was used in a working memory test; which image was used as a memorandum was counterbalanced across participants.

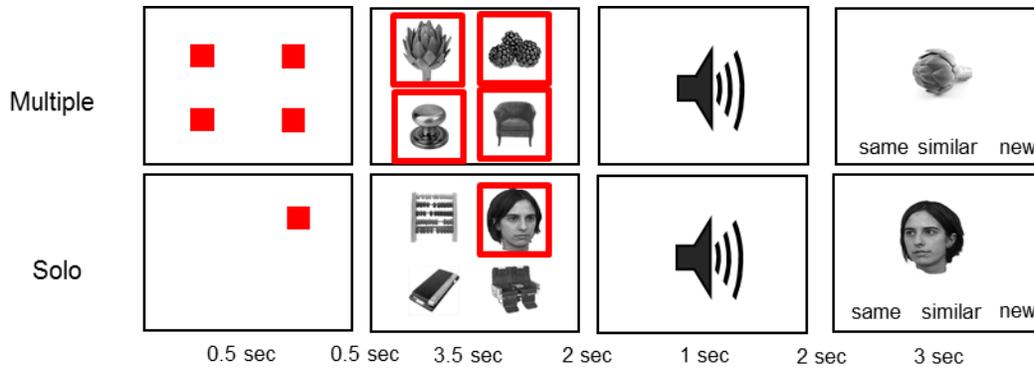
The objects were split into two stimulus sets (one for the CS+ and the other for the CS- condition; the assignment was counterbalanced across participants). The 104 objects in each set were randomly assigned to one of the following conditions for each participant: 24 objects were used as targets in the multiple object condition, 24 objects were used as distractors in the multiple condition, 24 were used as distractors in the solo condition, 8 foils in the WM test and 24 foils in the last memory test. Those assigned to targets were further randomly assigned to one of the three WM test item type conditions. In addition, we used an additional 192 objects as two more distractors presented in each trial and 8 object pairs in booster trials.

**Procedures and design.** The procedures were based on Studies 1 and 2. After filling out the same set of questionnaires as Studies 1 and 2, participants completed the fear-conditioning procedure; the average shock intensity determined for each participant was similar to that of Studies 1 and 2 ( $M = 1.45$  mA;  $SD = 0.52$ ). Participants next completed the questionnaire on their mood (Watson et al., 1988), followed by the learning phase. The learning phase was similar to that in Studies 1 and 2 (Figure 4),

except that a) participants were shown four objects rather than four faces in the multiple condition (multiple object condition) and b) each trial started with spatial cues to indicate the location of high-priority images: trials in the multiple object condition started with four red boxes for 500 ms, while trials in the solo face condition started with a red box for 500 ms which was shown at the location of a subsequent face.

The learning phase included 96 trials: 8 for each of the 2 (arousal: CS+ vs. CS-) x 2 (study set type: multiple vs. solo) x 3 (WM test item type: same, similar vs. new) conditions. We also included an additional 12 booster trials, where participants received stimulation following CS+ tone. These 108 trials were divided into three blocks, each of which included 32 main trials and 4 booster trials in a randomized order.

After participants completed the learning phase, they filled out a demographic questionnaire and completed a recognition test. The recognition test included 48 faces (one from each solo trial), 48 objects (one from each solo trial), 32 objects that were tested during the WM test (one from each multiple trial with the same and similar WM test), 64 un-tested objects from the multiple condition (one from each multiple trial with the same and similar WM test and two from each multiple trial with the new WM test), 16 new faces and 48 new objects. Images were presented in a randomized order, irrespective of whether they were faces or objects. For each image, participants were asked to indicate whether the image was old or new by key press.



**Figure 4.** Procedures in the learning phase in Study 3. Participants learned four objects in the multiple condition and one face in the solo condition, while faces in the solo condition were preceded by cues and highlighted to enhance their priority. Face images were taken from the TarrLab face database for illustration purpose only (stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition, Carnegie Mellon University, <http://www.tarrlab.org>).

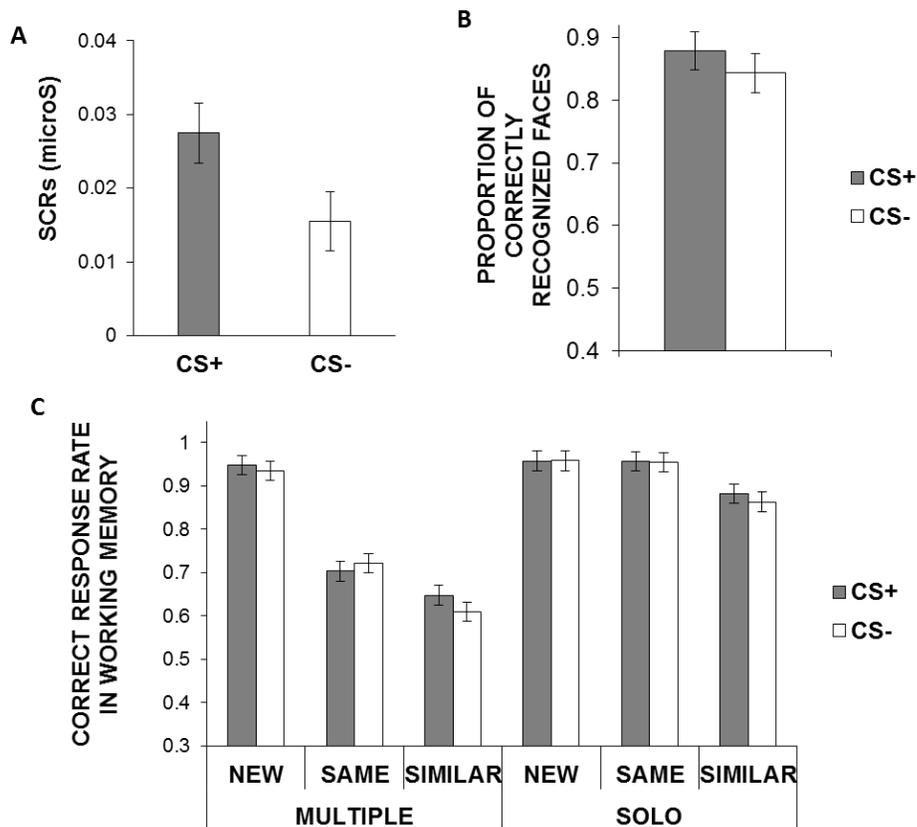
## Results and Discussion

**Fear conditioning.** Participants showed high accuracy in discriminating the two tones during the fear-conditioning phase (Table 1) with no significant differences across the conditions ( $p = .95$ ). They also showed higher SCR in the CS+ without shock condition than in the CS- condition (Figure 5A),  $t(39) = 3.58$ ,  $p = 0.0009$ ,  $d = .57$ .

**Final memory for faces.** We next examined performance during the final memory test to test our main prediction. One participant was at chance when responding to foils ( $M < .55$ ); data from this participant were not included in analyses reported in this section. The remaining participants were significantly better than chance in responding to foils ( $M = .79$ ),  $t(42) = 24.07$ ,  $p < .0001$ ,  $d = 3.85$ .

To address whether arousal enhances memory for faces in the solo condition, a 2 (arousal) x 3 (WM test item type) ANOVA was performed on the proportion of correctly identified faces. This ANOVA revealed a significant effect of arousal,  $F(1, 38) = 4.15$ ,  $p = .049$ ,  $d = 0.22$ , with no significant effects of WM test ( $p = .29$ ) and no

significant interactions ( $p = .98$ ). Participants showed better memory for faces in the CS+ condition than in the CS- condition (Figure 5B), suggesting that emotional arousal enhances memory for solo high-priority items when they have sufficient priority.



**Figure 5.** Results from Study 3. (A) CS+ induced higher SCRs than did CS- during the conditioning phase. (B) During the final memory test, emotion enhanced memory for solo high-priority faces. (C) Performance in the working memory test. Error bars represent standard errors of means.

**Final memory for objects.** We next tested the effects of emotion on memory for objects. First, we examined memory performance for objects that were tested in the WM test in the multiple condition by a 2 (arousal: CS+ vs. CS-) x 2 (WM test item type: same vs. similar) ANOVA. Note that objects in the solo condition were never tested during the WM test and therefore not included in this analysis. This ANOVA revealed a significant

effect of WM test item type,  $F(1, 38) = 8.01$ ,  $p = .007$ ,  $\eta^2 = .17$ , reflecting better memory for the same ( $M = .91$ ) than the similar condition ( $M = .83$ ). But neither the main effect of arousal ( $p = .69$ ) nor the interaction ( $p = .87$ ) was significant. Participants remembered objects equally well irrespective of arousal both in the same ( $M_{cs+} = .92$ ,  $M_{cs-} = .91$ ) and in the similar condition ( $M_{cs+} = .84$ ,  $M_{cs-} = .83$ ).

We also examined memory performance for objects that were not tested in the WM test by another 2 (arousal: CS+ vs. CS-) X 2 (study set type: solo vs. multiple) ANOVA. There was a significant effect of study set type,  $F(1, 38) = 225.74$ ,  $p < .0001$ ,  $\eta^2 = .85$ , suggesting better memory in the multiple condition ( $M = .61$ ) than in the solo condition ( $M = .14$ ). This is not surprising as participants did not expect a memory test for objects from the solo condition. More importantly, neither the main effect of arousal ( $p = .94$ ), nor the arousal-by-study set type interaction ( $p = .88$ ) was significant. Once again, participants showed similar performance irrespective of arousal both in the solo ( $M_{cs+} = .15$ ,  $M_{cs-} = .14$ ) and multiple conditions ( $M_{cs+} = .61$ ,  $M_{cs-} = .61$ ). In summary, memory for objects did not show any effects of arousal (see General Discussion for further discussion regarding this result).

**Working memory.** Finally, a 2 (arousal) x 2 (study set type) x 3 (WM test item type) ANOVA was performed on the correct response rate during the WM test (Figure 5C) to examine the effects of arousal on participants' ability to briefly maintain presented stimuli. This ANOVA revealed significant main effects of study set type,  $F(1, 39) = 132.09$ ,  $p < .0001$ ,  $\eta^2 = .82$ , and of WM test item type,  $F(2, 78) = 68.49$ ,  $p < .0001$ ,  $\eta^2 = .66$ . There was also a significant interaction between them,  $F(2, 78) = 29.45$ ,  $p < .0001$ ,  $\eta^2 = .39$ . But neither the arousal-by-study set type interaction ( $p =$

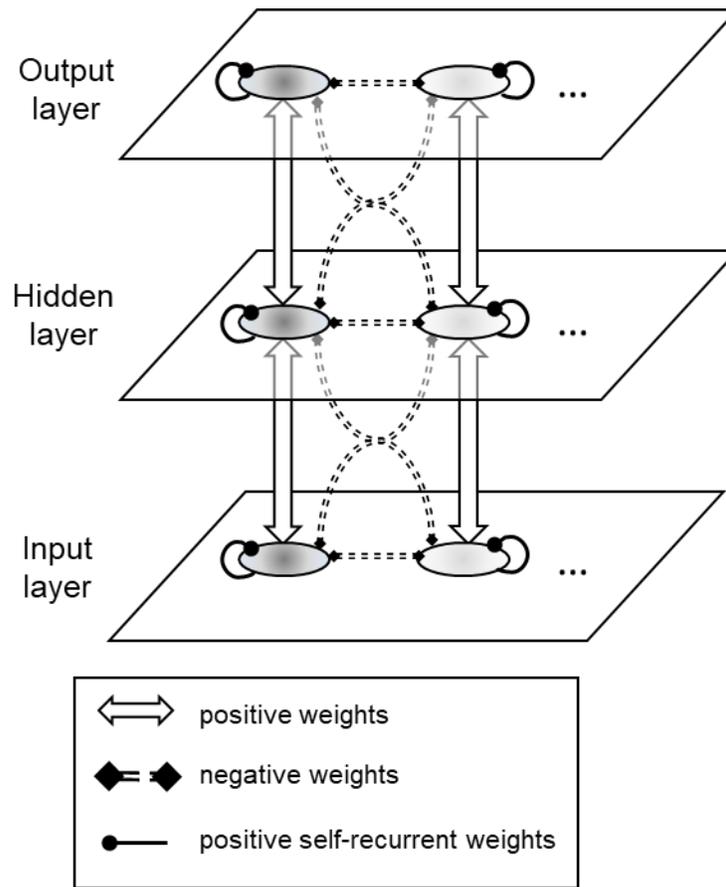
.86) nor any other effects including arousal was significant ( $ps > .30$ ). We also calculated the pattern separation index but it did not show significant effects of arousal ( $p = .31$ ).

### **General Discussion across Studies 1 through 3**

Across Studies 1 through 3, participants' performance in the final memory test was consistent with GANE. In Studies 1 and 2, we found that emotional arousal impaired memory for multiple high-priority faces. In Study 3, we found that emotional arousal enhanced memory for solo high-priority faces. These results suggest that emotional arousal amplifies competitions across goal-relevant items and impairs memory for goal-relevant items when there are multiple high-priority items. However, contrary to our prediction, the enhancement effects of arousal on high-priority stimuli were not significant in Studies 1 and 2, where high-priority faces were not cued in advance. Thus, the results were more nuanced than our initial expectation.

### **Simulation**

To test whether our behavioral results are consistent with GANE, we extended a recently developed computational model based on GANE (Lee et al., 2018) which concerns the effects of arousal on perception. Given that the current study concerns the effects of arousal on memory, we added the Hebbian learning mechanisms and a few more assumptions to explain the effects of NE on memory processing.



**Figure 6.** A schematic representation of our neural network model. The figure displays only two of the input and output units for simplicity.

In brief, the model has an input, hidden and output layers (Figure 6). Each unit in the input layer represents one stimulus in a localist manner. Likewise, each unit in the output layer represents one stimulus. The model included 80 units in each of the three layers. The input units influenced their corresponding output units via an intermediary hidden layer. All units were fully connected with each other within each layer as well as with units in the adjacent layer but with different strengths; connection strengths were

determined based on back-propagation algorithm so that activation of each input unit would activate one unit in the hidden layer and this hidden unit in turn activated the corresponding output unit<sup>2</sup> (see Appendix 1 for further details of the pre-training procedures). The model is based on standard neural network models but has a few critical parameters to represent GANE's assumptions (see Appendix 1 for further details): a) excitatory effects of NE on activation via glutamate release to enable the positive feedback loop between NE and glutamate ( $Glutamate_{i,t}$  in Equation 8), b) inhibitory effects of NE on competing representations via GABAergic signals ( $GABA_{k,t}$  in Equation 8), c) a learning\_rate for LTP based on  $\beta$  receptors (LTP\_learning\_rate in Equation 9) and d) a learning\_rate for LTD based on  $\alpha 1$ -adrenergic receptors (LTD\_learning\_rate in Equation 9).

We ran a simulation experiment and tested whether our computational model produces the same pattern as the psychological experiments. The procedures in the simulation mimicked those in the psychological experiments. Specifically, we had the model learn either multiple high-priority items (multiple condition) or one high-priority item (solo condition) followed by the manipulation of arousal. If arousal amplifies the competition across high-priority items, arousal should enhance memory in the solo condition but impair memory in the multiple condition.

## Method

**Design.** There were 80 input units in each layer of our model, which allowed us to have 80 items in simulation (one input unit represented one item in a localist manner). The 80 items were assigned to one of the four conditions: 2 (arousal: arousal, no-arousal) X 2 (study set type: solo, multiple).

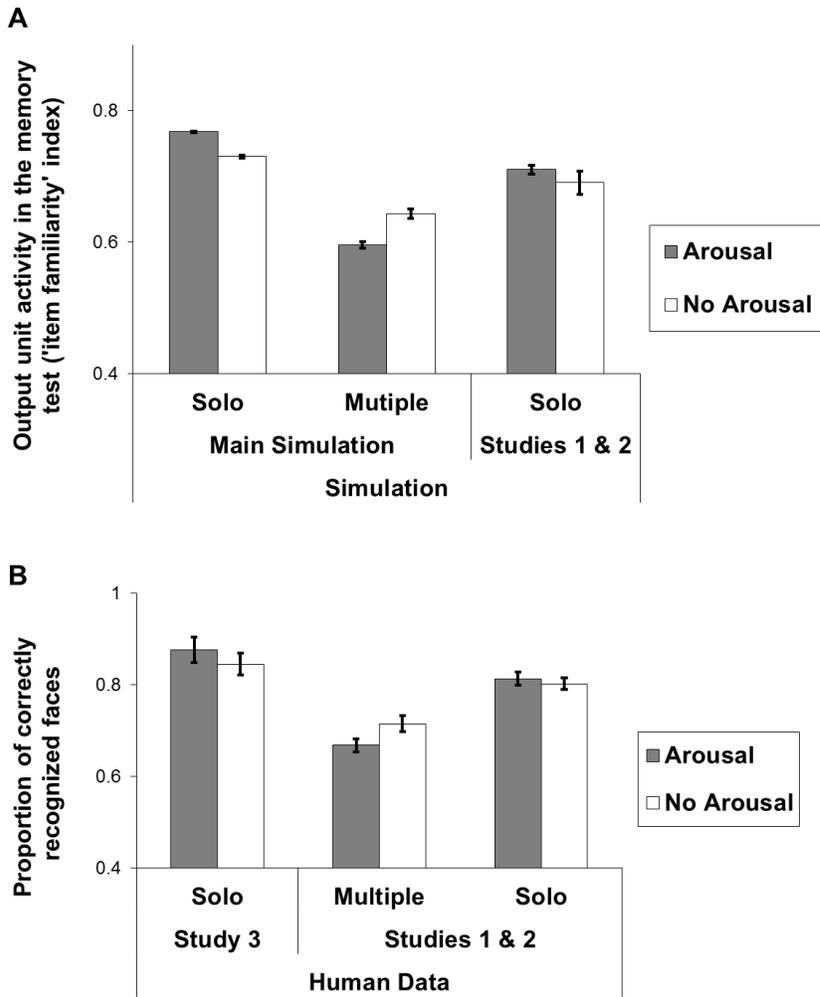
**Simulation procedure.** First, the model completed a learning phase, where it was presented with one high-priority stimulus and three to-be-ignored stimuli (solo condition) or four high-priority stimuli (multiple condition). The learning phase included 20 trials (five trials from each condition) in a randomized order and each trial included 50 time steps. To manipulate post-encoding arousal, we included two stages in each trial of the learning phase: a) an encoding stage (1-20 time steps), where the model encodes high-priority items and b) a maintenance stage (21-50 time steps), where the model no longer receives external inputs but holds the information whilst updating its status at every time step.

The procedures used in the encoding stage were different between the solo vs. multiple conditions. In the solo condition, the model received the external input value of 1.0 (hard-clamping) for one item (i.e., a high-priority item; hard-clamping), but received the external input value of 0.5 for the other three items (low-priority items; hard-clamping). In contrast, in the multiple condition, we used different procedures. One possible assumption for participants' behaviors in the multiple condition is that they always processed all four items in parallel and paid equal attention to all of them. However, it is also possible that participants engaged in serial processing, such that they paid attention to a certain item at one point but moved their attention to another item in a next time point, followed by an attention to another item in a following time point. We incorporated such participants' serial behavior (see Appendix 2 for the results from the model based on the parallel processing during the encoding stage). Specifically, at every time step, one of the four input units was randomly selected and received the external input value of 1.0, whereas the remaining three received the value

of 0.5 at that time step; each of the four units had the equal chance to be selected at every time step and this random selection was repeated for 20 time steps in trials of the multiple condition.

This encoding stage was followed by the maintenance stage, where we used the same procedure across the condition. In the maintenance stage, all external input values were set to zero, while the model updated its internal status (i.e., activations in the hidden and output units). To manipulate arousal, we applied the arousal manipulation (Equation 6) after the 30<sup>th</sup> time step in the arousal condition (31-50 time steps). In contrast, Equation (7) was applied in the non-arousal condition, as well as 1-30<sup>th</sup> timesteps in the arousal condition. After the 50<sup>th</sup> time step, the weights were updated through the learning algorithms (Equations 9-10).

After the learning phase of all the 80 items, we tested the models' memory for each item. The learning algorithms applied during the learning phase allowed the network to activate the output units of the studied items more strongly than before the learning phase. In the test phase, each item was presented one by one with external input value of 1.0 and the network updated its status for 20 time steps (i.e., the same time steps as the encoding phase during the learning phase). The resultant output value in the target output unit was measured and was taken as a measure of memory performance (an index of item's 'familiarity'). Similar approaches have been widely used in past simulation studies (e.g., Norman, Newman, & Detre, 2007; Norman & O'Reilly, 2003).

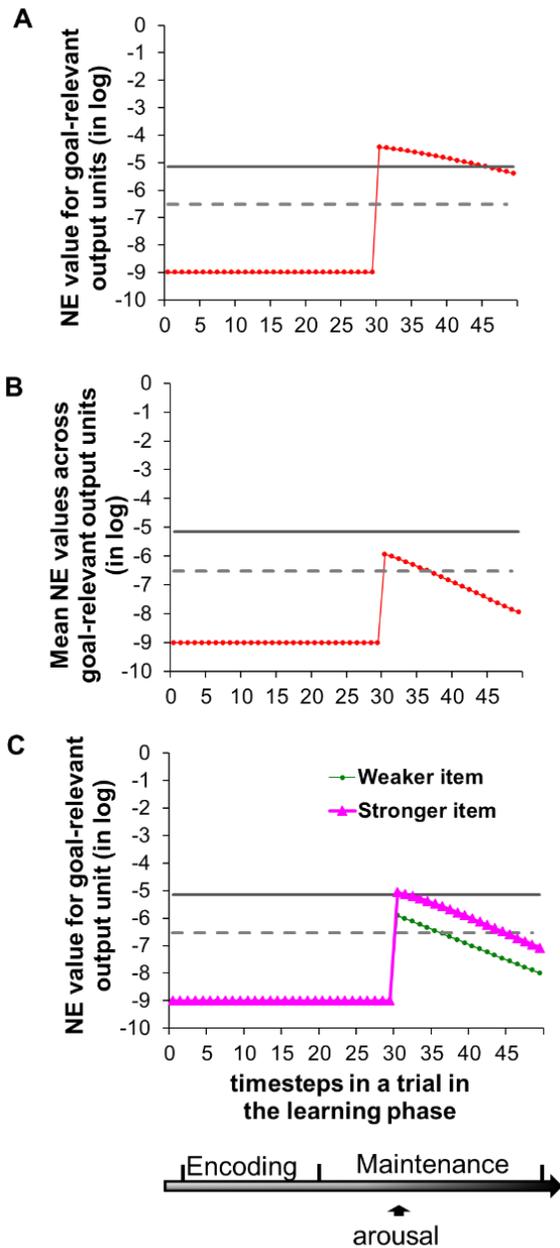


**Figure 7.** Simulation results for the arousal-induced competition. (A) The model showed that arousal enhances memory in the solo condition but impairs memory in the multiple condition. The model also reproduced the weaker enhancement effects of arousal observed in Studies 1 and 2. Data were averaged across 20 simulations. Error bars represent standard errors. (B) The results from our three studies were summarized to make it easier to compare the results from our simulation and psychological data.

**Results**

The ‘familiarity’ value was enhanced by arousal in the solo condition, but impaired by arousal in the multiple condition (Figure 7A left panel). Thus, the model showed the expected interaction between arousal and the number of high-priority items.

To understand the effects of arousal during learning, we next examined NE levels during the learning phase and found that NE levels play a critical role in the opposing effects of arousal between the solo vs. multiple conditions. In the solo condition (Figure 8A), the NE level reached the threshold to activate  $\beta$ -receptors under arousal and these units did not receive strong inhibitory signals as there were no other items that showed strong activity, which helped maintain activation of  $\beta$ -receptors and induce LTP. In contrast, in the multiple condition, when we examined the average NE levels across four goal-relevant items, the average NE level was below the threshold to activate low-affinity  $\beta$ -receptors (Figure 8B). However, a closer inspection of the model revealed that local NE levels differed across four high-priority items; some items were initially activated more strongly than other items<sup>3</sup> and the local NE levels reached the threshold to activate  $\beta$ -receptors for these stronger items (Figure 8C). This resulted in increased GABAergic inhibitory signals from these items, leading to increased suppression of other weaker goal-relevant items. Moreover, the stronger items also mutually suppressed each other and as a result, local NE levels quickly decreased even for these items. This resulted in the activation of high-affinity  $\alpha$ 1-receptors to induce inhibitory LTD for all four items.



**Figure 8.** Local NE values in the output unit during the learning phase in (A) the solo condition, (B) the multiple condition (here we present the averaged NE values over four items) and (C) the multiple condition (separately for strongly activated items and weakly activated items). To illustrate the threshold to activate  $\beta$  receptors (a solid line) and the threshold to activate  $\alpha_1$  receptors (a gray dashed line), raw NE values were log-transformed after converting them into nMol and shown in the figures. Data from one of the 20 simulations.

## **Discussion**

We found that the model's 'familiarity' index was impaired by arousal when there were multiple goal-relevant items, but enhanced by arousal when there was only one goal-relevant item. These results suggest that arousal amplifies competition among high-priority representations and can impair memory for high-priority information when there are multiple high-priority items but facilitate memory for solo high-priority information.

## **Simulation of Studies 1 and 2**

As mentioned above, the enhancing effects of arousal in the solo condition were not significant in Studies 1 and 2. Thus, one remaining question is whether the model can reproduce the weaker effects of arousal observed in the solo condition in the first two studies compared with Study 3. Although there were several differences in the procedures between the first two studies and Study 3, one critical difference is that high-priority items were cued in advance in Study 3 but not in the first two studies. In Study 3, faces were preceded by a red spatial cue, which helped participants focus on faces without looking at other images, which was particularly helpful in the solo condition. In contrast, in Studies 1 and 2, participants did not know the location of faces in advance and needed to look at other images briefly to find out high-priority images.

To test whether this face identification process is responsible for arousal's weaker enhancement effects in the first two studies, we ran another simulation where we included an identification stage at the beginning of the trials: a) an identification stage (1-10 time steps), where the model receives the same inputs for all four items to identify high-priority items irrespective of the conditions, b) the encoding stage (11-20

time steps), where the model receives stronger input for the single high-priority item, and c) the maintenance stage (21-50 time steps), where the model no longer receives external inputs but holds the information whilst updating its status at every time step. During the identification stage, irrespective of whether it was the solo vs. multiple condition, we used the same procedure as the encoding stage of the multiple condition in the main simulation. All the other procedures were the same as the main simulation.

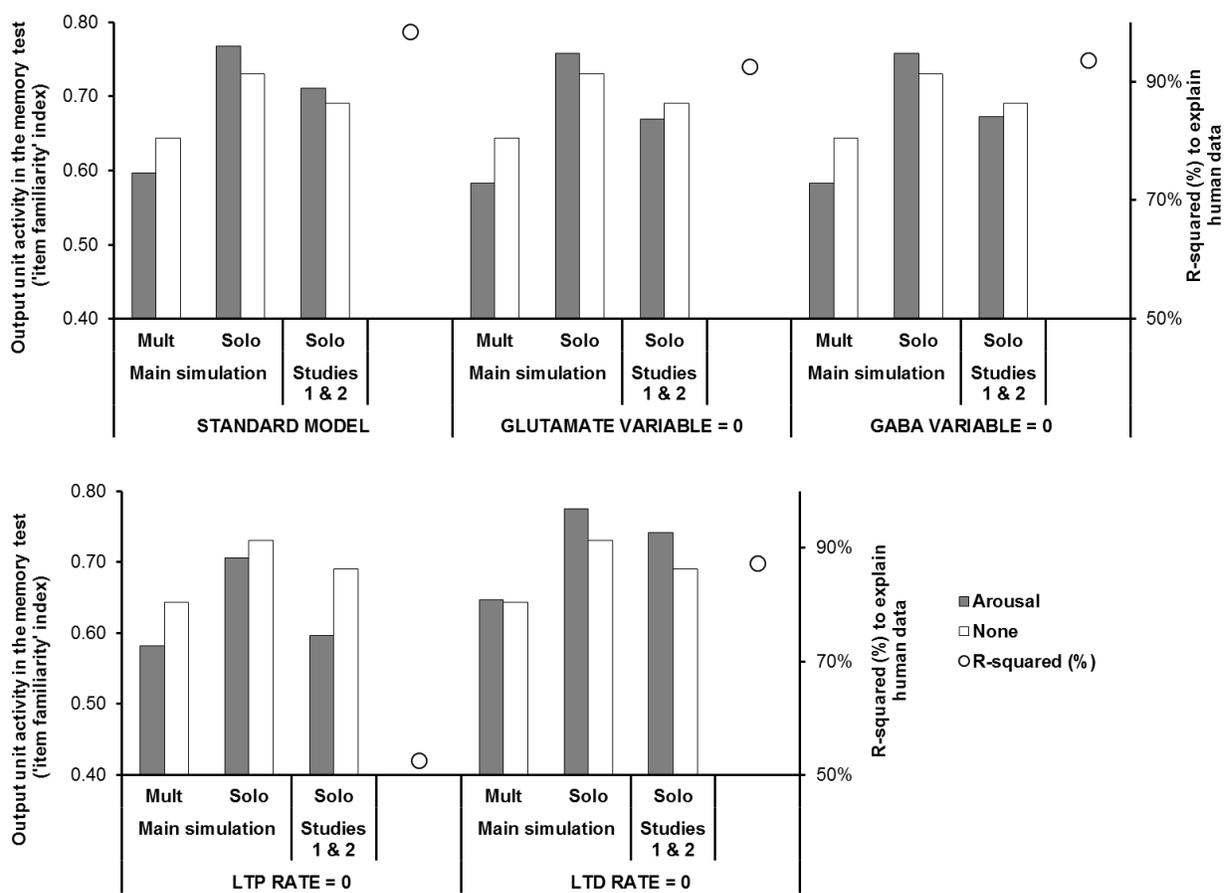
In this simulation, we found only a small enhancement effect of arousal in the solo condition (Figure 7A; right panel) as observed in Studies 1 and 2 (Figure 7B). Thus, our GANE-based model showed similar patterns to our behavioral results. The model's 'familiarity' index was impaired by arousal when there were multiple goal-relevant items, but enhanced by arousal when there was only one goal-relevant item. In addition, when we changed the procedures to make them similar to those in Studies 1 and 2, the model demonstrated a smaller enhancement effect of arousal in the solo condition.

### **Model evaluation**

Taken together, all the behavioral data through Studies 1 to 3 have been simulated by our model in a qualitative way. Assessing a quantitative fit was challenging because individual observations from the simulation and the human data are not directly comparable. Nevertheless, it was possible to assess the fit between the two by predicting the averaged human data presented in Studies 1 to 3 (Figure 7B) by the model data presented in Figure 7A. A linear regression showed a highly significant relationship between the model and the human data, with more than 98.40% of the available variance in the humans' performance accurately predicted by the values

derived from the simulation. These results suggest that while our behavioral results are not as clear as we expected, these results are consistent with GANE.

We also tested whether the model was able to offer a good qualitative/quantitative fit to the behavioral data if any of the critical parameter values was set to zero. This approach is helpful to evaluate whether the complexity of our model is necessary or not. Although the complexity of our model is due to the complexity of NE's neurophysiological effects and we used the same model as the one used in the previous study (Lee et al., 2018), it is still crucial to evaluate the necessity of these mechanisms to replicate the behavioral pattern. Figure 9 shows the outcomes of a series of these control simulations where we removed parameters crucial for GANE. These parameters include one for the excitatory effects of NE on activation via glutamate release ( $Glutamate_{i,t}$  in Equation 8), one for the inhibitory effects of NE on activation via GABAergic signals ( $GABA_{k,t}$  in Equation 8) and NE-based modulations on learning\_rate for LTP and LTD (see Equation 9). The results indicated that removing any of these key parameters led to either a low quantitative fit (i.e., lower prediction accuracy shown in lower  $R^2$  values) or a failure to simulate the enhancement effect on the solo condition or impairment effect on the multiple condition in Studies 1 and 2 (i.e., low qualitative fit). Taken together, our results suggest that all the parameters in the model were important.



**Figure 9.** The model's behavior and its fit with behavioral data when changing key parameters for GANE. Data were averaged across 20 simulations.

## General Discussion

There is ample evidence that neuromodulators affect a range of our behaviors and cognitive processes (Aston-Jones & Cohen, 2005; K. C. Berridge & Robinson, 1998; Bouret & Sara, 2005; Lisman, Grace, & Duzel, 2011; Sara, 2009). The literature of emotion and memory is not an exception. Previous studies have revealed a role for NE, a neuromodulator released from LC when encountering emotional events, in explaining the effects of emotion on memory (for reviews see Markovic et al., 2014; McIntyre et al., 2012). Since noradrenergic innervation is widespread across the brain (C. W. Berridge & Waterhouse, 2003), past studies often take the perspective of the broad and diffuse effects of NE by positing that NE has similar effects across regions or focus on the NE action in the amygdala. However, these theories do not clearly explain how emotional arousal sometimes enhances and sometimes impairs memory in a retrograde manner via the same noradrenergic mechanisms.

A recently proposed model, GANE, addresses this issue and focuses on the local and synaptic-level NE mechanisms (Mather et al., 2016). According to GANE, high-priority representations are enhanced by arousal via local synaptic NE regulation mechanisms. In the current study, we tested whether GANE is a plausible model to explain the effects of arousal on memory by instantiating it into a computational model. According to GANE, there is a boundary condition for the emotional arousal's enhancing effects on memory for high-priority information: While emotional arousal should enhance memory for solo high-priority information, it should amplify competition among representations and impair memory for high-priority information when there are multiple high-priority items. Across three behavioral studies, we found that the number of high-

priority items plays role in determining the effects of arousal on memory as expected by GANE: arousal enhances memory for goal-relevant information when there is only one goal-relevant item but impairs memory for goal-relevant information when there are multiple goal-relevant items. We also developed our computational model based on previous physiological studies and found that our model reproduces the results from our psychological experiments. Together, the current results suggest the importance of the neurobiological mechanisms posited in GANE to explain the retrograde effects of arousal on memory.

One puzzling result from the current study is the lack of the effects of emotional arousal on memory for objects in Study 3. In the current study, we included objects from a wide range of categories and therefore objects were more distinctive than faces. But even if objects were distinctive, arousal presumably should still impair memory for objects in the multiple object condition in Study 3 based on performance for faces in the multiple face condition in Studies 1 and 2. But arousal did not significantly affect memory for objects.

These differences in the effects of arousal on inhibition due to competition between items may be due to the differences in how distinctive faces were from each other versus how distinctive objects were from each other. In the present model, each processing unit in the hidden layer sends negative inputs to other units via GABAergic mechanisms and the size of these lateral inhibitory signals are set to the same value across all the interacting units within the same layer for the sake of simplicity. But lateral inhibition is typically observed only in neighboring neurons (Markram et al., 2004) and the brain is organized for each semantic category (e.g., Downing, Chan, Peelen, Dodds,

& Kanwisher, 2006). Thus, it is possible that processing units for similar items are connected with each other with highly negative weights that allow stronger inhibitory signals, whereas processing units for distinctive items from different categories are connected with each other with smaller negative weights that allow weaker inhibitory signals. Critically, according to our model, the arousal-induced impairment effects are largely due to the GABAergic mechanisms (see Equation 8 in Appendix 1). Thus, consistent with the behavioral findings, our model also suggests that arousal's impairment effects on memory are weak when high-priority items are distinctive from each other or when low-priority items are distinctive from high-priority items. Indeed, when we reduced the value of  $GABA_{i,t}$  parameter from 0.15 (standard model) to 0.10, the model also showed a smaller effect of impairment in the multiple condition ( $M_{\text{Arousal}} = .636$ ,  $M_{\text{NoArousal}} = .643$ ) than the standard impairment effect in the main simulation ( $M_{\text{Arousal}} = .596$ ,  $M_{\text{NoArousal}} = .643$ ). In summary, our model reproduced behavioral results from the three experiments with the same set of assumptions and the same set of parameters as the one from a previous study (Lee et al., 2018) -- except for the additional parameters that are relevant to learning algorithm.

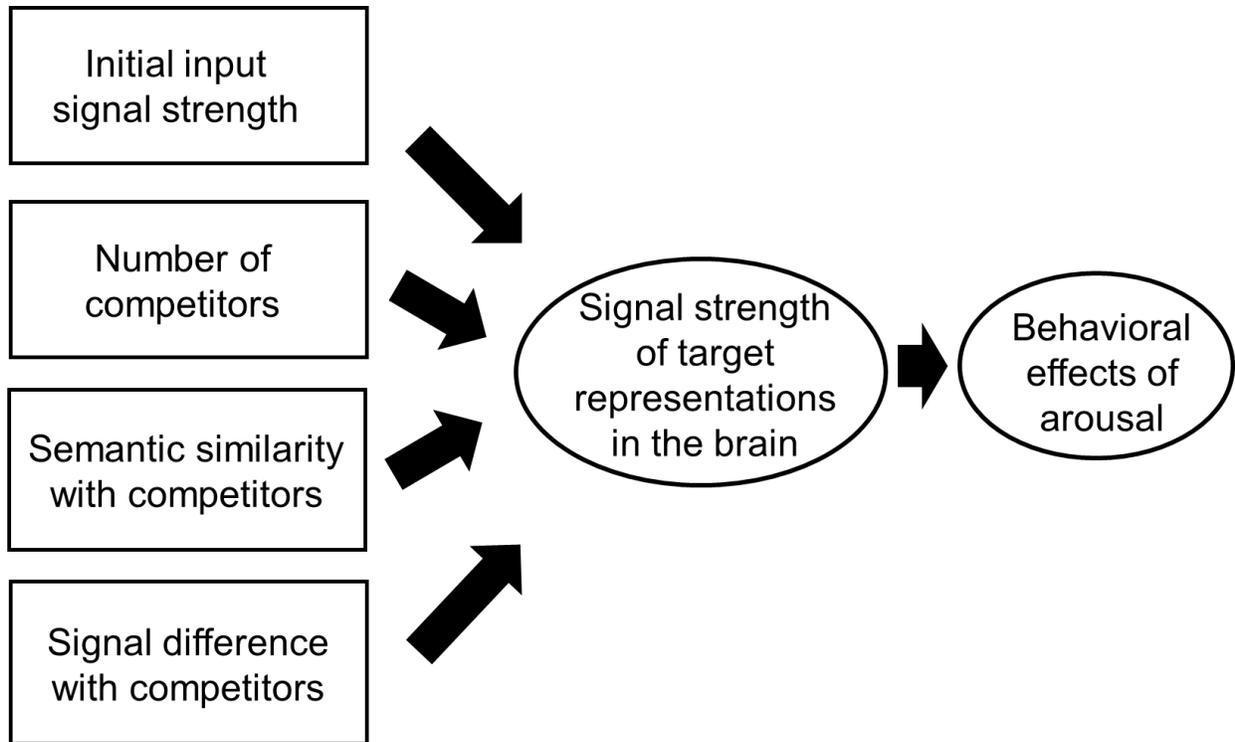
It should be noted that while across three studies, we observed results consistent with the model and GANE, none of the three studies showed a significant double dissociation. As we just described, in Study 3, we found emotion-induced enhancement effects for solo high-priority solo items but did not find a significant effect for emotion-induced impairment for multiple high-priority objects (though this lack of significant effect was consistent with our model). Likewise, in Studies 1 and 2, we observed the emotion-induced impairment effect for multiple high-priority items but we did not find a significant

emotion-induced facilitation effect for solo high-priority items (although once again the lack of significant effects was reproduced by our model). Previous studies also sometimes demonstrated that arousal impairs memory for low-priority items but does not necessarily facilitate memory for high-priority items (Clewett, Sakaki, Nielsen, Petzinger, & Mather, 2017; Ponzio & Mather, 2014). Other research also documented that arousal enhances memory for high-priority items but does not necessarily impair memory for low-priority items (Lee et al., 2015).

These results suggest that there are thresholds for items' signal strength to be enhanced by arousal or impaired by arousal. In other words, arousal would not enhance memory for items that are *relatively* high in priority in the context but do not exceed the threshold for the enhancement effects. Likewise, arousal would not impair memory for items that are *relatively* low in priority in the context but still have sufficiently high levels of signal strength beyond the threshold to induce the impairing effects of arousal. These thresholds may be related to the biological threshold to activate low-affinity  $\beta$  receptors that induce LTP and the threshold to activate high-affinity  $\alpha_1$  receptors that induce LTD.

It is unclear how these biological thresholds translate to behavioral manipulations. However, according to our model and the current findings, there are several factors that should interact with initial input strength in determining the strength of item representations in the brain and thus affecting the effects of arousal (Figure 10). The first factor is the number of competitors. When there is no competitor, the target item representations would not receive any inhibitory effects from competitors and thus maintain strong signals. But as the number of competitors increases, they compete with the high-priority target via the GABAergic mechanisms, resulting in the lower signal

strength for the high-priority target which would allow only for the impairing effects of arousal. The second factor is semantic distinctiveness of items as described above, such that high-priority items are more likely to be suppressed when other items are semantically similar than they are semantically distinctive. The other factors, which are not completely separable from the first factor, are the signal strength of competitors and the difference in input signal strength between high-priority targets and low-priority competitors. Specifically, when the signal strength of low-priority competitors is low and there is a large difference in the signal strength between high- and low- priority representations, high-priority representations would be able to release a larger amount of glutamate and NE to help them enhance signal strength, without receiving much inhibitory signals from low-priority representations, leading to the facilitation effects of arousal. But when there is only a small difference in the signal strength between high- and low- priority representations and when the signal strength of low-priority representations is high, the high-priority representations would receive strong inhibitory signals, thus leading to the impairing effects of arousal. The spacial cues we used in Study 3 presumably helped to enhance the signal difference in high- vs. low- priority representations by ensuring that participants allocated most of their attention to a single high-priority item from the beginning of a trial. Importantly, these factors should not act in an isolation but interact with one another (for example, even when low-priority representations have high signal strength, its inhibitory effects on high-priority representations are the function of semantic similarity between them). Future research is required to fully test the interaction across these factors.



**Figure 10.** The potential factors that determine signal strength of high-priority target representations and affect whether the item shows facilitative or impairing effects of arousal.

In addition, it is possible that there are additional parameters/mechanisms that play roles in determining how arousal affects memory beyond what we posit in our model. For example, one recent study shows that the enhancement effects of arousal for high-priority items were related to prolonged levels of salivary alpha amylase (Clewett, Sakaki, Nielsen, et al., 2017) which is known to be correlated to tonic pupil size and tonic activity of the LC-NE system (C. J. Anderson, Colombo, & Unruh, 2013). Thus, the arousal-induced enhancement effects might be due to the interaction between the phasic NE effects modelled in this study and the tonic NE effects. Future research needs to fully understand how tonic and phasic NE activity interact in modulating the effects of emotion on memory.

The notion that NE is important for emotion-memory interactions is not new. Previous studies also suggest that NE plays major roles in the emotional modulation of memory (Hurlemann et al., 2005; Markovic et al., 2014; McIntyre et al., 2012; Roozendaal et al., 2008; Strange et al., 2003). However, it has been unclear how the LC-NE system sometimes allows for retrograde enhancement effects of emotion on memory (Roozendaal et al., 2008) but sometimes leads to retrograde impairment effects of emotion on memory (Hurlemann et al., 2005; Strange et al., 2003). The current research suggests that local NE levels are key in determining whether emotional arousal leads to memory impairment or enhancement in a retrograde manner.

In another line of research, researchers have examined how encountering emotional stimuli interrupts ongoing executive control tasks. For example, when task-irrelevant emotional distractors capture attention, these emotional distractors are remembered well in long-term memory, but they diminish activity in the dorsal executive brain regions, resulting in the disrupted working memory performance for other neutral stimuli (Dolcos, Diaz-Granados, Wang, & McCarthy, 2008; Dolcos et al., 2013; Dolcos & McCarthy, 2006; Shafer & Dolcos, 2012). These impairing effects in the working memory have been explained by positing that the ventral affective system which is responsible for processing of emotional stimuli competes with the dorsal executive system, leading to poor performance in tasks that require the dorsal executive system (Bush, Luu, & Posner, 2000; Dolcos, Iordan, & Dolcos, 2011). However, emotional responses are often associated with both ventral and dorsal prefrontal cortical regions (e.g., Phan, Wager, Taylor, & Liberzon, 2002), raising the question of whether these two systems are always competing with each other. In contrast, according to our model,

emotional arousal triggers the local self-regulation mechanisms due to the interaction between NE and glutamate, which results in impairing effects on signals from whatever stimuli that have lower priority. Critically, emotional stimuli tend to have higher priority than other neutral stimuli due to individuals' goals (e.g., emotional stimuli may activate one's goal to maintain their survival) and/or higher bottom-up saliency (e.g., blood is red and salient than other stimuli). Thus, our model can provide a parsimonious account (without positing separate neural systems) about how and why emotional stimuli often disrupt ongoing executive function performance for other non-emotional materials while emotional stimuli themselves are remembered well.

Other computational models have also focused on the effects of the LC-NE system on cognitive processing (Eldar, Cohen, & Niv, 2013; Eldar, Niv, & Cohen, 2016; Nieuwenhuis, Gilzenrat, Holmes, & Cohen, 2005; Usher, Cohen, Servan-Schreiber, Rajkowski, & Aston-Jones, 1999). Based on the observation that NE increases signal-to-noise ratios (SNR) in neurons in sensory cortices (Foote, Freedman, & Oliver, 1975; Sara, 2009), the previous models increase the magnitude of the gain parameter in Equation (1) under arousal. Thus, both our model and these previous models predict that arousal leads to higher SNR. However, our model is different from the previous models in two important ways.

First, our model focuses on the local NE regulation mechanisms via its interaction with glutamate and GABA. In contrast, according to the previous models, the LC-NE system changes the steepness of the sigmoid curve in the output function of *all* units in the model. Thus, the previous models posit global and diffuse effects of NE across all neurons in the brain. These models reproduce the enhancement effects of

arousal on signals for a higher-priority input and the impairment effects of arousal on signals for a lower-priority input when there are two competing items with different priority levels (Eldar et al., 2013). However, without modifications, these models would not predict that arousal impairs memory for high-priority information when there are multiple to-be-processed items with equally strong priority levels. According to these models, arousal leads to globally increased gain across all processing units, increasing the signal level of all strongly activated items, thereby resulting in arousal-induced memory enhancement effects for at least one of the competing goal-relevant items. However, our study did not find arousal-induced enhancement in variance in accuracy for the multiple condition (Study 1:  $M_{\text{arousal}} = .37$ ;  $M_{\text{no-arousal}} = .36$ ,  $p = .63$ ; Study 2:  $M_{\text{arousal}} = .20$ ;  $M_{\text{no-arousal}} = .19$ ,  $p = .30$ ; Study 3:  $M_{\text{arousal}} = .27$ ;  $M_{\text{no-arousal}} = .28$ ,  $p = .26$ ).

Second, our model explains how NE affects not only the signal strength during the processing/learning phase but also long-term memory strength. Based on GANE, our model posits that higher NE levels lead to LTP, whereas lower NE levels result in LTD. To the best of our knowledge, this is the first time such NE-based learning algorithms have been incorporated into a computational model, allowing simulation of human memory performance with the resultant model. Thus, our model is unique in being based on neurophysiological mechanisms concerning how NE levels are modulated *locally* and how NE interacts with different noradrenergic receptors to produce different effects both in a short-term manner (i.e., signal strength during initial processing) and in a long-term manner (i.e., synaptic connection strength after Hebbian learning).

There are several important questions for future studies to be noted. First, in the current model, we determined NE baseline levels and thresholds to activate  $\alpha$ -1/ $\beta$  receptors based on previous observations in physiological studies. Yet as described above, it is not clear how the parameters of the current model map to real biophysical quantities. Future research should address this issue by testing the robustness of the model's behaviour on this as well as other psychological findings while varying these values.

Second, in our behavioral studies, while we observed patterns consistent with GANE in the final memory test, performance in the WM test did not show selective enhancement effects of arousal on solo high-priority items. This suggests that the effects of arousal proposed in GANE are stronger after a delay possibly due to the role of NE on LTP and LTD as modelled by Equations 9-10 (see Appendix 1). However, previous research shows that arousal has selective facilitation effects for high-priority information during a short-term memory task (Sutherland & Mather, 2012) and an attention task (Lee et al., 2018; Lee, Sakaki, et al., 2014), both of which did not have a long delay. Future research should systematically manipulate the length of delay and test the interaction between arousal and delay length.

Third, participants in our behavioral studies were not balanced on their gender and we had more females than males across all three studies. Recent research indicates that females show stronger interaction between arousal and priority in memory than do males (Clewett, Sakaki, Huang, Nielsen, & Mather, 2017). Previous studies also suggest that ovarian hormone levels interact with arousal in affecting memory in women (Clewett, Sakaki, Huang, et al., 2017; Nielsen, Barber, Chai, Clewett, & Mather,

2015). Future research should take into account sex and the action of ovarian hormones in the GANE model.

Another remaining question concerns whether CS+ induced similar emotional reactions across the multiple and solo conditions. In fact, previous research indicates that the ability to focus attention without being distracted by distractors decreases under high working memory load (Lavie, Hirst, de Fockert, & Viding, 2004). Given that working memory load should be higher in the multiple condition than in the solo condition, participants may be more distracted by the task-irrelevant CS+ in the multiple condition than in the solo condition, which may have resulted in greater emotional reactions to CS+ in the multiple condition than in the solo condition. While this account may be able to explain the selective impairment effects due to CS+ in the multiple condition in the first two studies, it cannot easily explain better memory for solo faces in the CS+ condition than in the CS- condition in Study 3. Therefore, our results are more consistent with the GANE's prediction that arousal impairs memory for multiple high-priority items via increased competitions among high-priority representations. Nevertheless, future research should measure emotional reactions during the learning phase to address this possibility.

In conclusion, in the current paper, we found that emotional arousal amplifies competitions across goal-relevant items. We also found that a computational model which was developed in a recent separate study (Lee et al., 2018) can reproduce these behavioral results. This model should help clarify how and why emotional arousal sometimes enhances memory and sometimes impairs memory. The current model also provides a critical step to bridge a gap between human psychological research on

emotion and memory and neurophysiological research on NE and glutamate. In addition, NE has been implicated not only in emotion and memory interactions, but also in general cognitive processing such as attention and learning (Aston-Jones & Cohen, 2005; Sara, 2009). Thus, the current results should also provide insights into how the LC-NE system supports human mental functioning in general.

### Footnotes

<sup>1</sup> We used celebrity faces because faces are in general more salient than objects and thus have high priority (Lee, Sakaki, et al., 2014). In addition, each celebrity has many available images, which allowed us to select two images that differed in various features and to make it difficult for participants to predict a subsequent working memory test image. Thus, these stimuli helped us encourage participants to pay attention to all details of each memorandum image.

<sup>2</sup> Given that the same results were obtained when the number of hidden units for each item was increased, in the main model we present in the paper, we used the simplest model which has one hidden unit per item.

<sup>3</sup> This within-trial variability stems from the pretraining algorithm. During the pretraining stage, the network was pretrained to activate each output unit more than 0.9. Thus, there were variations in the resultant output values across items; some items had higher output values (e.g., 0.915) than other items (e.g., 0.901). This resulted in differences in the activity levels across items when they were presented at the same time.

### **Data References**

Summarized data from studies reported in this paper and codes associated with the simulation and the experimental tasks are available at: <https://osf.io/4mer6/>

## Appendix 1: Model details and pre-training

### Model details

The output of  $i$ -th unit ( $a_i$ ) is computed by a standard sigmoid function as follows:

$$a_i = \frac{1}{1 + e^{(-s_i * gain)}} \quad (1)$$

where  $gain$  is set to 1.0;  $s_i$  is the net input for unit  $i$  determined by weights and output values of all units that send their outputs to the unit (see Equation 2):

$$s_i = \sum_j (w_{j,i} * a_j) + b_i \quad (2)$$

where  $w_{j,i}$  is a weight of the connection from the sending unit  $j$  to the receiving unit  $i$ , and  $b_i$  is a bias input set to be -5. The network is operated in a continuous manner across  $N$  time steps, during which the time-integrated output value of each unit gradually changes by Equation (3):

$$time\ integrated\ a_{i,t} = a_{i,t-1} + 0.1 * (a_{i,t} - a_{i,t-1}) \quad (3)$$

where the second subscript  $t$  denotes the time step.

In addition to the above functions used in standard neural network models, we implemented core assumptions of GANE in our model. We used the same set of

assumptions and the same set of parameters as Lee et al. (2018) except for the additional parameters that are relevant to learning algorithm (see below). The specific values of key parameters were changed from Lee et al. (2018) in order to offer the best fit to the study-specific behavioural data in this article. More specifically, we explored the model's predictions across a range of parameter values using a coarse grid and compared it with data from the behavioral experiments (Plaut, 1997) (see Supplemental materials for the model's prediction when we used other values for these parameters).

**a) Glutamate signals priority.** According to GANE, glutamate is the brain's primary excitatory neurotransmitter and signals priority, such that neurons that process high-priority information release a larger amount of glutamate than those that process low-priority information. Based on this assumption, in our model, we posit that the local unit activation level corresponds to the amount of glutamate released from the unit.

**b) Each unit is associated with different levels of NE.** The second assumption of GANE is that NE concentration levels are controlled locally. To implement this, we introduced a unit-specific parameter  $NE_{i,t}$  to represent NE level for each unit  $i$  for each time step  $t$ . The baseline of  $NE_{i,t}$  is set to 0.000000001 (1e-9) mol/liter based on the baseline NE level observed in previous physiological studies (approx 1nM in the cortex; Slaney, Mabrouk, Porter-Stransky, Aragona, & Kennedy, 2013).

**c) Local NE levels are modulated depending on arousal and local glutamate levels.** According to GANE, the elevated glutamate associated with high-priority neural representations stimulates NMDA receptors on LC axonal varicosities to release more NE to create NE hotspots when LC neurons are activated (i.e., immediately after phasic arousal is induced) and when the local glutamate neurons release sufficient amount of

glutamate (Fink, Göthert, Molderings, & Schlicker, 1989; Lüscher & Malenka, 2012; Pittaluga & Raiteri, 1990). To implement this assumption, we defined  $NE_{i,t}$  under arousal as a decreasing function of the elapsed time after an arousing event and as an increasing function of the activity of the local unit as follows.

$$NE_{i,t} \text{ under arousal} = \tau * (a_{i,t})^2 * 0.9^s \quad (4)$$

where  $s$  denotes a time step elapsed after the onset of an arousing event;  $a_{i,t}$  is the output value of the unit  $i$  at  $t$ -th time step;  $\tau$  is a constant parameter which is set to 0.0001 to scale the product of the right side of Equation (4) to become a realistic NE concentration level (in mol/liter) under arousal based on previous physiological observations (Harley, Lalies, & Nutt, 1996; Salgado et al., 2012). The last term (0.9) was determined in a conservative way: Specifically, this should be less than 1 in order to realize the decreasing NE as a function of the elapsed time after an arousing event.

**d) Phasic NE release is returned to baseline.** Due to the NE reuptake process (Amara & Kuhar, 1993), the increased NE level will gradually return to the baseline value of 0.000000001 (1e-9) (see Equation 5).

$$NE_{i,t} = 0.9 * NE_{i,t-1} + 0.000000001 \quad (5)$$

By combining Equations (4) and (5), the local NE level for each unit is defined as follows:

*NE levels after phasic arousal induction:*

$$NE_{i,t} = \left[ \tau * (a_{i,t-1})^2 * 0.9^{s-1} \right] * 0.9 + 0.0000000001 \quad (6)$$

*NE levels without phasic arousal:*

$$NE_{i,t} = NE_{i,t-1} * 0.9 + 0.0000000001 \quad (7)$$

**e) NE modulates additional glutamate release and GABAergic activity.**

Another critical assumption of GANE is that higher NE concentrations in NE hot spots lead to stimulation of low-affinity  $\beta$ -adrenergic receptors, resulting in more glutamate release (e.g., Ferrero et al., 2013), and that strong glutamate signals (e.g., Xue, Atallah, & Scanziani, 2014) and high NE concentrations (e.g., Nai, Dong, Hayar, Linster, & Ennis, 2009) in NE hot spots stimulate GABAergic activity to inhibit other neurons that process competing representations.

To implement the interactions across NE, glutamate and GABA, we included two additional parameters for each unit  $i$  and each time  $t$ : a)  $Glutamate_{i,t}$  which refers to the excitatory effects of NE on glutamate release from the unit  $i$  via  $\beta$ -adrenergic receptors and b)  $GABA_{k,t}$  which refers to the inhibitory effects from unit  $k$  (other units within a layer) on unit  $i$  due to GABAergic activity. These two parameters are added to Equation (2) as follows:

$$s_{i,t} = \sum_j (w_{j,i} * a_{j,t}) + Glutamate_{i,t} - \sum_k GABA_{k,t} \quad (8)$$

where  $k \neq i$ . The value of  $Glutamate_{i,t}$  and  $GABA_{k,t}$  are determined based on the local  $NE_{i,t}$  value as follows:

*if*  $NE_{i,t} \leq 0.000007$ ;

$$Glutamate_{i,t} = 0$$

$$GABA_{i,t} = 0$$

*if*  $NE_{i,t} > 0.000007$ ;

$$Glutamate_{i,t} = 0.15$$

$$GABA_{i,t} = 0.15$$

Due to the action of  $Glutamate_{i,t}$  the outcome of Equation (8) gets larger when NE levels near unit  $i$  are higher than 0.000007 (7e-6) – a threshold value to activate  $\beta$  receptors. In contrast, due to the action of  $GABA_{k,t}$  the output of unit  $i$  gets smaller when NE levels near *other* units within the same layer (unit  $k$ ) are higher than 0.000007 (7e-6). The threshold value to induce the  $\beta$ -driven facilitative effects was estimated based on physiological findings (Harley et al., 1996; Salgado et al., 2012).

Note that even after NE levels reach the threshold to activate  $\beta$ -adrenergic receptors, they will gradually decline and fall below that threshold via the reuptake

process (see Equation 6). We assume that on these occasions, the effects of NE on glutamate and GABA gradually decline as follows:

$$Glutamate_{i,t} = 0.15 * 0.9^u$$

$$GABA_{i,t} = 0.15 * 0.9^u$$

where  $u$  steps have elapsed since  $NE_{i,t}$  gets below 0.000007 (7e-6).

**f) NE gates LTP or LTD depending on its concentration levels.** According to GANE, NE should also modulate memory processing in addition to glutamate release. Specifically, high NE levels should allow low-affinity  $\beta$ -receptors to initiate long-term potentiation (LTP), whereas moderate or lower NE levels should stimulate only high-affinity  $\alpha$ 1-adrenoreceptors, resulting in long-term depression (LTD; Salgado et al., 2012). These effects are implemented by the following learning algorithms for each unit  $i$  after the last time step ( $t = N$ ) in each learning trial:

$$w_{i,,after\_updating} = w_{i,,before\_updating} * (1 + LTP\_learning\_rate)^v * (1 + LTD\_learning\_rate)^w$$

(9)

where  $w_{i,,}$  is all of the weights that project from unit  $i$ . The threshold values to induce either LTP driven by  $\beta$ -receptors or LTD driven by  $\alpha$ 1-receptors were determined based on previous findings that  $\beta$ -receptors require 20~30-fold higher concentrations of NE to be active than  $\alpha$ 1-receptors (Harley et al., 1996; Salgado et al., 2012). Specifically,  $v$  in

Equation (9) denotes the number of time steps in which  $NE_{i,t} > 0.000007$ , and the value of  $LTP\_learning\_rate$  was set to 0.015. In contrast,  $w$  in Equation (9) denotes the number of time steps in which  $0.000007 \geq NE_{i,t} > 0.0000003$ , and the value of  $LTD\_learning\_rate$  was set to -0.015.

In addition, the weight matrix was updated by the standard Hebbian learning algorithm in each learning trial:

$$w_{i,j,t+1} = w_{i,j,t} * (1 + hebbian\_learning\_rate * a_{i,t} * a_{j,t}) \quad (10)$$

where  $hebbian\_learning\_rate$  was 0.15. Again, this learning algorithm was conducted at the end of trial for the sake of simplicity. Specifically, we saved the time-specific values of Equation (10) in each time step during the trial, and then this learning algorithm was repeated 50 times (i.e., the number of time steps) at the end of each trial by retrieving the stored values.

### **Network pre-training**

The weight matrix of the model was determined so that activation of each input unit would activate one unit in the hidden layer (hereafter denoted as *target hidden unit*), and this unit in turn activated the corresponding output unit (hereafter denoted as *target output unit*). This means that units that are responsible for the same item were positively connected. The remaining links were inhibitory except for the self-recurrent connectivities. The connection strength was determined by pre-training the model with a back-propagation learning algorithm.

The pre-training stage consisted of a) training of a working memory task (where the model was required to reproduce a presented stimulus in the target output unit even after the input stimulus disappeared; each item was repeated for 100 times), b) a test of the memory task, c) training of a perception task (where the model was required to recognize and reproduce a presented stimulus in the target output unit during the stimulus presentation; each item was repeated for 100 times) and d) a test of the perception task. Items for which the model produced the correct answer during the tests were excluded from the subsequent training stage of the same task, and this cycle was repeated until the model produced the correct response to all items in both the memory test and the perception test.

**Working memory task.** Each trial during the memory task included 50 time steps. In the first 20 steps, the model received an external input for one item, an input unit which corresponded to the presented item was activated (i.e.,  $a_i = 1.0$ ), and this activation spread to other units in the same and other layers. In the next 30 time steps (21-50 time steps), the model no longer received an external input and therefore the output value for the input unit was changed to 0 (i.e.,  $a_i = 0.0$ ) but the model updated its internal status (i.e., activations in the hidden and output units). If the activity of the target output unit is greater than 0.9 either at the penultimate time step (49<sup>th</sup> time step) or the final time step (50<sup>th</sup> time step), we considered that the model correctly maintained the presented stimulus, and therefore the weight matrix was maintained. In contrast, if the activity of the target output unit is smaller than 0.9 both at the 49<sup>th</sup> and the 50<sup>th</sup> time steps, we regarded that the model missed the presented stimulus and back-propagation learning algorithm was applied to the weight matrix to improve the model's behavior in a

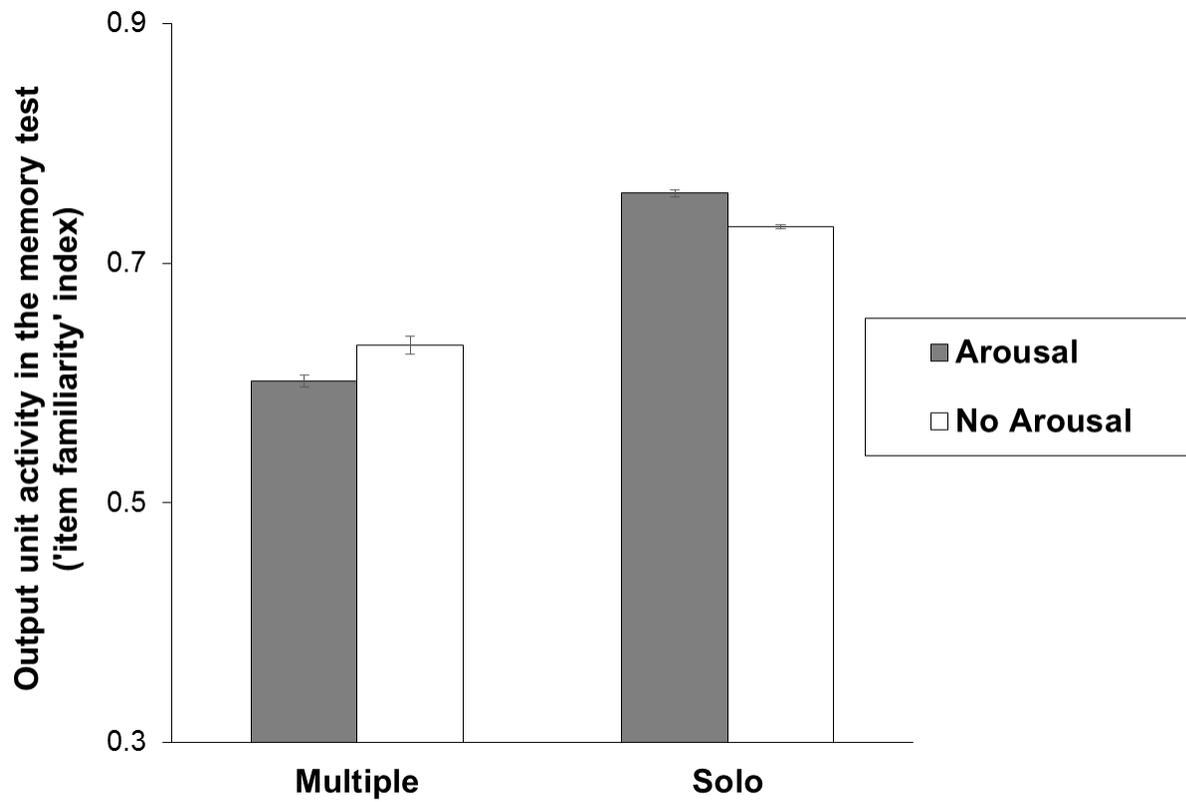
next trial. In addition, if the activities of the other output units (i.e., non-target output units) were bigger than 0.1 both at the 49<sup>th</sup> and the 50<sup>th</sup> time steps, we regarded that the model produced a false alarm and applied the back-propagation algorithm.

After we completed the working memory training for 100 times for each item, we tested the model's behavior. The procedures of this test were identical to those in the working memory task, except that we did not update the weight matrix based on the output. As done in the working memory task, we considered that the model correctly maintained the presented stimulus if the activity of the target output unit is greater than 0.9 either at the penultimate time step (49<sup>th</sup> time step) or the final time step (50<sup>th</sup> time step) and if the activities of the non-target output units were smaller than 0.1 both at the 49<sup>th</sup> and the 50<sup>th</sup> time steps.

***Perception task.*** Procedures of the perception task and a subsequent test were similar to those in the working memory task, except that we maintained the external input (i.e.,  $a_i = 1.0$ ) for 21-50 time steps.

**Appendix 2**

The results from simulation where we posited that participants engaged in parallel processing in the multiple condition.



### References

- Amara, S. G., & Kuhar, M. J. (1993). Neurotransmitter Transporters: Recent Progress. *Annual Review of Neuroscience*, 16(1), 73-93. doi: 10.1146/annurev.ne.16.030193.000445
- Anderson, A. K., Wais, P. E., & Gabrieli, J. D. E. (2006). Emotion enhances remembrance of neutral events past. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), 1599-1604. doi: 10.1073/pnas.0506308103
- Anderson, C. J., Colombo, J., & Unruh, K. E. (2013). Pupil and salivary indicators of autonomic dysfunction in autism spectrum disorder. *Developmental Psychobiology*, 55(5), 465-482. doi: 10.1002/dev.21051
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28(1), 403-450. doi: 10.1146/annurev.neuro.28.061604.135709
- Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, 42(1), 33-84. doi: 10.1016/s0165-0173(03)00143-7
- Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research. Brain Research Reviews*, 28.
- Bouret, S., & Sara, S. J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences*, 28(11), 574-582. doi: <http://dx.doi.org/10.1016/j.tins.2005.09.002>

- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325-14329. doi: 10.1073/pnas.0803390105
- Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215-222.
- Chiu, Y.-C., Dolcos, F., Gonsalves, B., & Cohen, N. (2013). On opposing effects of emotion on contextual or relational memory. *Frontiers in Psychology*, *4*(103). doi: 10.3389/fpsyg.2013.00103
- Clark, K. B., Naritoku, D. K., Smith, D. C., Browning, R. A., & Jensen, R. A. (1999). Enhanced recognition memory following vagus nerve stimulation in human subjects. *Nature Neuroscience*, *2*(1), 94-98.
- Clewett, D., Sakaki, M., Huang, R., Nielsen, S. E., & Mather, M. (2017). Arousal amplifies biased competition between high and low priority memories more in women than in men: The role of elevated noradrenergic activity. *Psychoneuroendocrinology*, *80*, 80-91. doi: <https://doi.org/10.1016/j.psyneuen.2017.02.022>
- Clewett, D., Sakaki, M., Nielsen, S., Petzinger, G., & Mather, M. (2017). Noradrenergic mechanisms of arousal's bidirectional effects on episodic memory. *Neurobiology of Learning and Memory*, *137*(Supplement C), 1-14. doi: <https://doi.org/10.1016/j.nlm.2016.10.017>
- Dolcos, F., Diaz-Granados, P., Wang, L. H., & McCarthy, G. (2008). Opposing influences of emotional and non-emotional distracters upon sustained prefrontal cortex activity during a delayed-response working memory task.

- Neuropsychologia*, 46(1), 326-335. doi: 10.1016/j.neuropsychologia.2007.07.010
- Dolcos, F., Iordan, A., Kragel, J., Stokes, J., Campbell, R., McCarthy, G., & Cabeza, R. (2013). Neural Correlates of Opposing Effects of Emotional Distraction on Working Memory and Episodic Memory: An Event-Related fMRI Investigation. *Frontiers in Psychology*, 4(293). doi: 10.3389/fpsyg.2013.00293
- Dolcos, F., Iordan, A. D., & Dolcos, S. (2011). Neural correlates of emotion-cognition interactions: A review of evidence from brain imaging investigations. *Journal of Cognitive Psychology*, 23, 669-694.
- Dolcos, F., Katsumi, Y., Denkova, E., & Dolcos, S. (2017). Factors Influencing Opposing Effects of Emotion on Cognition: A Review of Evidence from Research on Perception and Memory. In I. Opris & M. F. Casanova (Eds.), *The Physics of the Mind and Brain Disorders: Integrated Neural Circuits Supporting the Emergence of Mind* (pp. 297-341). Cham: Springer International Publishing.
- Dolcos, F., Katsumi, Y., Weymar, M., Moore, M., Tsukiura, T., & Dolcos, S. (2017). Emerging Directions in Emotional Episodic Memory. *Frontiers in Psychology*, 8(1867). doi: 10.3389/fpsyg.2017.01867
- Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron*, 42(5), 855-863. doi: 10.1016/S0896-6273(04)00289-2
- Dolcos, F., & McCarthy, G. (2006). Brain systems mediating cognitive interference by emotional distraction. *Journal of Neuroscience*, 15(7), 2072-2079. doi: 10.1523/JNEUROSCI.5042-05.2006
- Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006).

- Domain Specificity in Visual Cortex. *Cerebral Cortex*, 16(10), 1453-1461. doi: 10.1093/cercor/bhj086
- Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*, 16(8), 1146-1153. doi: 10.1038/nn.3428
- Eldar, E., Niv, Y., & Cohen, J. D. (2016). Do You See the Forest or the Tree? Neural Gain and Breadth Versus Focus in Perceptual Processing. *Psychological Science*, 27(12), 1632-1643. doi: 10.1177/0956797616665578
- Ferrero, J. J., Alvarez, A. M., Ramírez-Franco, J., Godino, M. C., Bartolomé-Martín, D., Aguado, C., . . . Sánchez-Prieto, J. (2013).  $\beta$ -Adrenergic Receptors Activate Exchange Protein Directly Activated by cAMP (Epac), Translocate Munc13-1, and Enhance the Rab3A-RIM1 $\alpha$  Interaction to Potentiate Glutamate Release at Cerebrocortical Nerve Terminals. *Journal of Biological Chemistry*, 288(43), 31370-31385. doi: 10.1074/jbc.M113.463877
- Fink, K., Göthert, M., Molderings, G., & Schlicker, E. (1989). N-Methyl-d-aspartate (NMDA) receptor-mediated stimulation of noradrenaline release, but not release of other neurotransmitters, in the rat brain cortex: receptor location, characterization and desensitization. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 339(5), 514-521. doi: 10.1007/BF00167254
- Finn, B., & Roediger, H. L. (2011). Enhancing retention through reconsolidation: Negative emotional arousal following retrieval enhances later recall. *Psychological Science*, 22, 781-786. doi: 10.1177/0956797611407932
- Foote, S. L., Freedman, R., & Oliver, A. P. (1975). Effects of putative neurotransmitters on neuronal activity in monkey auditory cortex. *Brain Research*, 86(2), 229-242.

doi: 10.1016/0006-8993(75)90699-X

Harley, C. W., Lalies, M. D., & Nutt, D. J. (1996). Estimating the synaptic concentration of norepinephrine in dentate gyrus which produces  $\beta$ -receptor mediated long-lasting potentiation in vivo using microdialysis and intracerebroventricular norepinephrine. *Brain Research*, 710(1–2), 293-298. doi: [http://dx.doi.org/10.1016/0006-8993\(95\)01443-8](http://dx.doi.org/10.1016/0006-8993(95)01443-8)

Hassert, D. L., Miyashita, T., & Williams, C. L. (2004). The effects of peripheral vagal nerve stimulation at a memory-modulating intensity on norepinephrine output in the basolateral amygdala. *Behavioral Neuroscience*, 118(1), 79-88. doi: <http://dx.doi.org/10.1037/0735-7044.118.1.79>

Hurlemann, R., Hawellek, B., Matusch, A., Kolsch, H., Wollersen, H., Madea, B., . . . Dolan, R. J. (2005). Noradrenergic modulation of emotion-induced forgetting and remembering. *Journal of Neuroscience*, 25, 6343-6349. doi: 10.1523/JNEUROSCI.0228-05.2005

Kensinger, E. A., & Corkin, S. (2004). Two routs to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Science*, 101(9), 3310-3315. doi: 10.1073/pnas.0306408101

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language*, 54(1), 99-112. doi: 10.1016/j.jml.2005.05.005

Knight, M., & Mather, M. (2009). Reconciling findings of emotion-induced memory enhancement and impairment of preceding items. *Emotion*, 9(6), 763-781. doi: 10.1037/a0017281

- LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, 7(1), 54-64. doi: 10.1038/nrn1825
- Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, 133(3), 339-354.
- Leal, S. L., Tighe, S. K., Jones, C. K., & Yassa, M. A. (2014). Pattern separation of emotional information in hippocampal dentate and CA3. *Hippocampus*, n/a-n/a. doi: 10.1002/hipo.22298
- Lee, T.-H., Baek, J., Lu, Z.-L., & Mather, M. (2014). How Arousal Modulates the Visual Contrast Sensitivity Function. *Emotion*, 14(5), 978-984.
- Lee, T.-H., Greening, S., Ueno, T., Clewett, D., Ponzio, A., Sakaki, M., & Mather, M. (2018). Arousal increases neural gain via the locus coeruleus-norepinephrine system in younger adults but not in older adults. *Nature Human Behaviour*, 2, 356–366. doi: 10.1038/s41562-018-0344-1
- Lee, T.-H., Greening, S. G., & Mather, M. (2015). Encoding of goal-relevant stimuli is strengthened by emotional arousal in memory. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.01173
- Lee, T.-H., Sakaki, M., Cheng, R., Velasco, R., & Mather, M. (2014). Emotional arousal amplifies the effects of biased competition in the brain. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nsu015
- Levine, L. J., & Edelman, R. S. (2009). Emotion and memory narrowing: A review and goal-relevance approach. *Cognition & Emotion*, 23(5), 833-875. doi: 10.1080/02699930902738863

- Lisman, J., Grace, A. A., & Duzel, E. (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends in Neurosciences*, 34(10), 536-547. doi: <http://dx.doi.org/10.1016/j.tins.2011.07.006>
- Loftus, E., & Burns, T. (1982). Mental shock can produce retrograde amnesia. *Memory and Cognition*, 10(4), 318-323. doi: 10.3758/BF03202423
- Lüscher, C., & Malenka, R. C. (2012). NMDA receptor-dependent long-term potentiation and long-term depression (LTP/LTD). *Cold Spring Harbor Perspectives in Biology*, 4(6). doi: 10.1101/cshperspect.a005710
- Markovic, J., Anderson, A. K., & Todd, R. M. (2014). Tuning to the significant: Neural and genetic processes underlying affective enhancement of visual perception and memory. *Behavioural Brain Research*, 259(0), 229-241. doi: <http://dx.doi.org/10.1016/j.bbr.2013.11.018>
- Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nat Rev Neurosci*, 5(10), 793-807. doi: [http://www.nature.com/nrn/journal/v5/n10/supinfo/nrn1519\\_S1.html](http://www.nature.com/nrn/journal/v5/n10/supinfo/nrn1519_S1.html)
- Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences*, 39. doi: 10.1017/S0140525X15000667
- Mather, M., & Sutherland, M. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science*, 6, 114-133. doi: 10.1177/1745691611400234

- McIntyre, C. K., McGaugh, J. L., & Williams, C. L. (2012). Interacting brain systems modulate memory consolidation. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1750-1762. doi: 10.1016/j.neubiorev.2011.11.001
- Miu, A. C., Heilman, R. M., Opre, A., & Miclea, M. (2005). Emotion-induced retrograde amnesia and trait anxiety. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1250-1257. doi: 10.1037/0278-7393.31.6.1250
- Nai, Q., Dong, H.-W., Hayar, A., Linster, C., & Ennis, M. (2009). Noradrenergic Regulation of GABAergic Inhibition of Main Olfactory Bulb Mitral Cells Varies as a Function of Concentration and Receptor Subtype. *Journal of Neurophysiology*, *101*(5), 2472-2484. doi: 10.1152/jn.91187.2008
- Nielsen, S. E., Barber, S. J., Chai, A., Clewett, D. V., & Mather, M. (2015). Sympathetic arousal increases a negative memory bias in young women with low sex hormone levels. *Psychoneuroendocrinology*, *62*, 96-106. doi: <http://dx.doi.org/10.1016/j.psyneuen.2015.08.001>
- Nielsen, S. E., & Mather, M. (2015). Comparison of two isometric handgrip protocols on sympathetic arousal in women. *Physiology & Behavior*, *142*, 5-13. doi: <http://dx.doi.org/10.1016/j.physbeh.2015.01.031>
- Nielson, K. A., & Jensen, R. A. (1994). Beta-adrenergic receptor antagonist antihypertensive medications impair arousal-induced modulation of working memory in Elderly Humans. *Behavioral and Neural Biology*, *62*(3), 190-200. doi: [http://dx.doi.org/10.1016/S0163-1047\(05\)80017-2](http://dx.doi.org/10.1016/S0163-1047(05)80017-2)
- Nielson, K. A., & Powless, M. (2007). Positive and negative sources of emotional arousal enhance long-term word-list retention when induced as long as 30 min

- after learning. *Neurobiology of Learning and Memory*, 88(1), 40-47. doi: 10.1016/j.nlm.2007.03.005
- Nielson, K. A., Radtke, R. C., & Jensen, R. A. (1996). Arousal-induced modulation of memory storage processes in humans. *Neurobiology of Learning and Memory*, 66(2), 133-142. doi: <http://dx.doi.org/10.1006/nlme.1996.0054>
- Nielson, K. A., Yee, D., & Erickson, K. I. (2005). Memory enhancement by a semantically unrelated emotional arousal source induced after learning. *Neurobiology of Learning and Memory*, 84(1), 49-56. doi: 10.1016/j.nlm.2005.04.001
- Nieuwenhuis, S., Gilzenrat, M. S., Holmes, B. D., & Cohen, J. D. (2005). The role of the locus coeruleus in mediating the attentional blink: A neurocomputational theory. *Journal of Experimental Psychology: General*, 134(3), 291-307. doi: <http://dx.doi.org/10.1037/0096-3445.134.3.291>
- Norman, K. A., Newman, E., & Detre, G. (2007). A Neural Network Model of Retrieval-Induced Forgetting. *Psychological Review*, 114(4), 887-953.
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling Hippocampal and Neocortical Contributions to Recognition Memory: A Complementary-Learning-Systems Approach. *Psychological Review*, 110(4), 611-646.
- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, 16(2), 331-348.
- Pittaluga, A., & Raiteri, M. (1990). Release-enhancing glycine-dependent presynaptic NMDA receptors exist on noradrenergic terminals of hippocampus. *European*

*Journal of Pharmacology*, 191(2), 231-234. doi: [http://dx.doi.org/10.1016/0014-2999\(90\)94153-O](http://dx.doi.org/10.1016/0014-2999(90)94153-O)

- Plaut, D. C. (1997). Structure and Function in the Lexical System: Insights from Distributed Models of Word Reading and Lexical Decision. *Language and Cognitive Processes*, 12(5-6), 765-806. doi: 10.1080/016909697386682
- Ponzio, A., & Mather, M. (2014). Hearing something emotional influences memory for what was just seen: How arousal amplifies effects of competition in memory consolidation. *Emotion*, 14(6), 1137-1142. doi: 10.1037/a0037983
- Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1(3), 385-401. doi: 10.1177/014662167700100306
- Roosendaal, B., Castello, N. A., Vedana, G., Barsegyan, A., & McGaugh, J. L. (2008). Noradrenergic activation of the basolateral amygdala modulates consolidation of object recognition memory. *Neurobiology of Learning and Memory*, 90(3), 576-579. doi: <http://dx.doi.org/10.1016/j.nlm.2008.06.010>
- Sakaki, M., Fryer, K., & Mather, M. (2014). Emotion strengthens high-priority memory traces but weakens low-priority memory traces. *Psychological Science*, 25(2), 387-395. doi: 10.1177/0956797613504784
- Salgado, H., Köhr, G., & Treviño, M. (2012). Noradrenergic 'tone' determines dichotomous control of cortical spike-Timing-dependent plasticity. *Scientific Report*, 2.
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Review Neuroscience*, 10(3), 211-223.

Segal, S. K., Stark, S. M., Kattan, D., Stark, C. E., & Yassa, M. A. (2012).

Norepinephrine-mediated emotional arousal facilitates subsequent pattern separation. *Neurobiology of Learning and Memory*, *97*(4), 465-469. doi: 10.1016/j.nlm.2012.03.010

Shafer, A., & Dolcos, F. (2012). Neural Correlates of Opposing Effects of Emotional Distraction on Perception and Episodic Memory: An Event-Related fMRI Investigation. *Frontiers in Integrative Neuroscience*, *6*(70). doi:

10.3389/fnint.2012.00070

Slaney, T. R., Mabrouk, O. S., Porter-Stransky, K. A., Aragona, B. J., & Kennedy, R. T. (2013). Chemical Gradients within Brain Extracellular Space Measured using Low Flow Push–Pull Perfusion Sampling in Vivo. *ACS Chemical Neuroscience*, *4*(2), 321-329. doi: 10.1021/cn300158p

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventor STAI (Self-evaluation questionnaire)*. Palo Alto, CA: Consulting Psychologists Press.

Strange, B. A., Hurlmann, R., & Dolan, R. J. (2003). An emotion-induced retrograde amnesia in humans is amygdala- and beta-adrenergic-dependent. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 13626-13631. doi: 10.1073/pnas.1635116100

Sutherland, M. R., & Mather, M. (2012). Negative arousal amplifies the effects of saliency in short-term memory. *Emotion*, *12*(6), 1367-1372. doi:

10.1037/a0027860

Talmi, D. (2013). Enhanced emotional memory: Cognitive and neural mechanisms.

*Current Directions in Psychological Science*, 22(6), 430-436. doi:

10.1177/0963721413498893

Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999).

The role of locus coeruleus in the regulation of cognitive performance. *Science*, 283(5401), 549-554.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief

measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070. doi: 10.1037/0022-

3514.55.1.128

Wechsler, D. (2001). *Wechsler Test of Adult Reading: WTAR*: Psychological

Corporation.

Xue, M., Atallah, B. V., & Scanziani, M. (2014). Equalizing excitation-inhibition ratios

across visual cortical neurons. *Nature*, 511(7511), 596-600. doi:

10.1038/nature13321

Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. L.

(2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus*, 21(9),

968-979. doi: 10.1002/hipo.20808