Social Cognitive and Affective Neuroscience, 2017, 1–10

doi: 10.1093/scan/nsx145
Original article

Forgetting emotional material in working memory

Eda Mızrak,[1,2] Henrik Singmann,[3] and Ilke Öztekin[1]

[1]Department of Psychology, Koç University, Istanbul 3310, Turkey, [2]Department of Psychology, Center for Neuroscience, University of California, Davis, CA 95616, USA, and [3]Department of Psychology, University of Zurich, 8006 Zürich, Switzerland

Correspondence should be addressed to Eda Mızrak, Department of Psychology, Center for Neuroscience, University of California, 1544 Newton Court, Davis, CA 95616, USA. E-mail: edamizrak@gmail.com

Abstract

Proactive interference (PI) is the tendency for information learned earlier to interfere with more recently learned information. In the present study, we induced PI by presenting items from the same category over several trials. This results in a build-up of PI and reduces the discriminability of the items in each subsequent trial. We introduced emotional (e.g. disgust) and neutral (e.g. furniture) categories and examined how increasing levels of PI affected performance for both stimulus types. Participants were scanned using functional magnetic resonance imaging (fMRI) performing a 5-item probe recognition task. We modeled responses and corresponding response times with a hierarchical diffusion model. Results showed that PI effects on latent processes (i.e. reduced drift rate) were similar for both stimulus types, but the effect of PI on drift rate was less pronounced PI for emotional compared to neutral stimuli. The decline in the drift rate was accompanied by an increase in neural activation in parahippocampal regions and this relationship was more strongly observed for neutral stimuli compared to emotional stimuli.

Key words: proactive interference; emotion; memory retrieval; forgetting; hierarchical diffusion model

Introduction

One common scenario that makes forgetting cumbersome in our daily lives is when we change our password and use the old version the next time we sign into our account. Only after failing once or twice, we remember that we updated our password and retrieve the new version. We often suffer from retrieval failures as such due to interference of previously learned yet irrelevant information (i.e. the old password), known as proactive interference (PI). PI is a huge cost on our cognitive system, such that forgetting from working memory (WM) could be minimum if there was no interference (Jonides and Nee, 2006).

Considerable amount of research evaluated how we efficiently retrieve relevant information from WM in the presence of PI (Badre and Wagner, 2005; D'Esposito *et al.*, 1999; Jonides *et al.*, 1998; Öztekin and Badre, 2011; Öztekin *et al.*, 2012; Öztekin and McElree, 2007). While these studies provided rich findings suggesting that we adapt retrieval strategies and gear up with controlled retrieval processes to resist PI, these effects were mostly tested and demonstrated with neutral stimuli. However, in many everyday situations we deal with emotionally loaded stimuli. In a recent study, Mizrak and Öztekin (2016) showed that PI affected WM retrieval dynamics for two different stimulus types, emotionally arousing stimuli and neutral stimuli, differently. They used the Recent-Probes task in which participants are asked to hold a small set of target items in memory over a brief retention interval and then make a *yes* (i.e. member of the study list) or *no* (i.e. not a member of the study list) decision to a recognition probe. The recognition probe can be an item from the study list (target), an item that participants did not encounter before (new lure), or an item that was presented in the study list of the previous trial but not the current one (recent lure). Typically, participants are more prone to falsely recognize the recent lure as a target item compared to

Received: 15 March 2017; **Revised:** 22 November 2017; **Accepted:** 11 December 2017

non-recent lures. The difference in false alarms between recent lures and non-recent lures is therefore one behavioral measure of PI. If the temporal information of the recent lure (the previous study list) is recovered, PI can be overcome. Mizrak and Öztekin (2016) used a response deadline procedure to investigate the full time course of retrieval. The general pattern (Öztekin *et al.*, 2012; Öztekin and McElree, 2007, 2010) is that PI quickly builds up due to item and/or familiarity information (i.e. the difference in false alarm is largest early in the trial), but can be resolved later on after retrieving source information (i.e. that a recent lure was only shown in the previous trial). Their findings revealed an interaction of PI time course and stimulus type: The buildup of PI was slower for emotional compared to neutral trials, leading to lower amounts of interference early in the retrieval process, extending a growing body of evidence indicating that emotion slows down forgetting from long term memory to WM (Yonelinas and Ritchey, 2015). However, detrimental effects of PI were stronger for emotional material later on which suggested that temporal context retrieval in order to resolve PI was weaker for emotional stimuli also extending the findings from the literature showing emotion improves item memory while it might impair context memory (Chiu *et al.*, 2013). Overall, their findings showed that PI effects for emotional and neutral stimuli are qualitatively similar (i.e. early build-up of PI followed by a later phase of resolution) but differ only quantitatively.

The current study extends these results and investigates whether the neural responses to PI and the latent cognitive processes underlying PI are different for emotional and neutral stimuli. Firstly, we aimed to replicate our previous results (Mizrak and Öztekin, 2016) showing that the buildup effect of PI on emotional stimuli was less pronounced compared to neutral stimuli. This would suggest that the detrimental effects of PI on memory performance would be weaker for emotional stimuli in the presence of PI compared to neutral stimuli. Secondly, we were interested whether the differences between stimulus types are qualitative or merely quantitative. By qualitative differences we mean different underlying processes, neural or cognitive, that contribute to the memory decision for emotional versus neutral stimuli. In contrast, quantitative differences entail the same underlying processes and only differences in the degree to which certain process react to increasing levels of PI. Because previous studies consistently show greater amygdala activity during encoding and retrieval of emotional stimuli compared to neutral stimuli (reviewed in Kensinger and Schacter, 2008), one could also expect such qualitative differences for our task.

Another way of inducing PI is by presenting categorically similar items. In the release-from-PI (rfPI) paradigm Wickens (1970) manipulated PI by presenting items from the same semantic category within a block (or mini-block) of several trials and by switching categories between the blocks of trials (note that there was no break between blocks). Within one block, the items share the same features and are highly similar to each other which reduces the discriminability of the items (Nairne, 2002). More specifically, a particular cue that can be used in retrieval will now be related to multiple items. Thus, the to-be-retrieved items in the later trials of one block are subject to more competition, which leads to a decline in performance (i.e. slower responses and lower memory accuracy). It is assumed that PI increases across trials within one block and the decline of performance associated with increasing levels of PI constitutes the PI effect. Studies that used the rfPI paradigm consistently showed that performance recovers when the category is changed and the items presented are dissimilar to the items presented in previous trials (i.e. when the next block begins; Öztekin and Badre, 2011; Öztekin *et al.*, 2009; Öztekin and McElree, 2007; Watkins and Watkins, 1975; Wickens, 1970). To resolve PI participants need to selectively retrieve unique details about the items beyond category membership.

In order to study the effects of PI on emotion, we used the rfPI task and manipulated the content of the categories. We introduced emotional (i.e. disgust and fear) and neutral (i.e. furniture and kitchen utensils) categories and examined how consecutively presenting items from either of these (sub-)categories in one block affected memory performance and the underlying neural mechanisms. For the two neutral categories we expected participants to use the specific semantic category (furniture or kitchen utensil) as a retrieval cue, whereas for emotional items we expected participants to use the specific emotion an item elicits (disgust or fear) as a retrieval cue. Since the items that are presented in the trials within a block come from the same category, participants would not benefit from relying on the category membership performance would decrease unless participants retrieve more diagnostic features unique to the item.

Previous research has shown that emotion enhances source memory for those features that are perceptually bound to the item while it impairs contextual/relational details that are not inherent to the item (see Chiu *et al.*, 2013, for the opposing effects of emotion on contextual memory; and Kensinger, 2009). While contextual details of the item that were not bounded to the item itself might be impaired for emotional items, visual details that were central to the item can be remembered better compared to neutral memoranda (see Kensinger *et al.*, 2006). Recently, Yonelinas and Ritchey (2015) proposed an account that provides a partial explanation to the findings in this literature. They suggested that emotion enhances recollection of item-emotion bindings mediated by amygdala, which might be more resistant to forgetting than hippocampus-dependent item-context bindings. Collectively, these findings suggest that emotion might increase the likelihood of remembering item details rather than context details. In the present study, emotional stimuli might therefore be retrieved with more item-related details than neutral stimuli which would lead to higher availability of more diagnostic item features. Thus, we could expect a less strong effect of PI (i.e. performance decline within one block) for emotional compared to neutral items.

Previous studies investigating PI effects in WM point to two important regions showing activation during successful item recognition in the presence of PI; parahippocampal cortex (PHc) within the medial temporal lobe (MTL) and left ventrolateral prefrontal cortex (VLPFC). MTL activation is generally associated with successful retrieval from long term memory (Diana *et al.*, 2007; see Eichenbaum *et al.*, 2007 for a review) and has also been shown to contribute the retrieval of WM representations (Öztekin *et al.*, 2009, 2010). VLPFC on the other hand has been suggested as the critical region for the cognitive control of PI in working memory. For instance, Öztekin *et al.* (2009) combined the rfPI and recent-probe paradigms to conjointly induce PI by semantic similarity (i.e. rfPI paradigm) and episodic familiarity. Their results indicated that resolving interference, regardless of its nature, required engaging both the PHc and left VLPFC regions. Finally, Öztekin and Badre (2011) used the rfPI paradigm to test the effects of PI on the encoding and retrieval of semantically related words. They showed that during retrieval, activation in the anterior part of left VLPFC (aVLPFC) varied as a function of PI. Crucially, aVLPFC activation mediated the relationship between increasing levels of PI within one block of semantically related items and the behavioral measurements. Memory performance, reflected in response times, was less affected with higher aVLPFC activation suggesting that this region was particularly related to the cognitive control of PI. They also observed a linear increase in PHc activity in response to increasing levels of PI and that this activation was associated with retrieval success. The performance of the individuals with higher PHc declined less as a function of PI compared to individuals with lower PHc activity. Accordingly, in our task we also expected aVLPFC and parahippocampal regions to respond to increasing levels of PI in a linear fashion. We were particularly interested whether neural response to PI differed for emotion and neutral stimuli. Our experimental design followed Öztekin and Badre (2011) with the addition of two emotional categories.

To summarize, our main goal was to investigate how PI in WM leads to forgetting of neutral and emotional information. Particularly, we were interested whether latent processes and neural markers underlying PI differed qualitatively or quantitatively between emotional and neutral stimuli. We expected at least quantitative differences following mounting evidence suggesting that emotional memories are more resistant to forgetting compared to their non-emotional counterparts (reviewed in Yonelinas and Ritchey, 2015). For example, in our previous work (Mızrak and Öztekin, 2016), we modeled the full time course of retrieval and were able to show that PI for emotional items differed quantitatively from the PI for neutral items. Here, we used a sequential sampling model, the diffusion model, to decompose the recognition decisions into latent cognitive processes (Ratcliff and Starns, 2013; White and Poldrack, 2014, see Method section for detailed description of the model). We expected increasing levels of PI to decrease memory performance. In addition, we examined neural activity during retrieval of emotion and neutral trials and expected that increasing level of PI lead to a linear increase in activation in two ROIs (aVLPFC and parahippocampal gyrus). We also expected the amygdala to be more strongly activated by emotional compared to neutral stimuli with no specific predictions regarding the effect of PI. Finally, we were interested how neural activity in response to increasing levels of PI was related to the diffusion model parameters.

Materials and methods

Participants

Twenty-one healthy adults (9 females) participated in this study. Data from two participants (one male, one female) were excluded due to excessive motion in the scanner. Participants had normal or corrected to normal vision and were screened for medical conditions that could contradict with MRI protocols. Participants gave written consent and were compensated for their participation.

Stimuli selection

The stimuli consisted of two neutral (Kitchen utensils and Furniture) and two emotional (Fear and Disgust) categories of images. Some of the emotional images were chosen from the International Affective Picture System (IAPS) database (Lang *et al.*, 2005). The rest was selected from Google images and rated by 18 different participants that did not participate in the actual experiment (See Supplementary Material for the detailed description of stimuli selection).

Experimental design

The experiment consisted of two sessions of six runs that took place on two different days. Each run contained 36 experimental trials. In each trial, participants first studied five-items in the encoding phase, solved three math problems, and then responded to a recognition probe. For each probe, participants were asked to indicate whether it appeared in the encoding phase (See Figure 1 for the procedure).

The release from PI paradigm was employed in which images from the same category were presented for three consecutive trials. The three consecutive trials from the same category formed a block and the first trial within this triplet will be referred to as (PI) Level 1, the second as Level 2, and the third as Level 3, as we expected PI to increase monotonically across the three trials. In total, there were 432 (216 × 2) trials (108 trials from each category), 144 blocks (36 blocks from each category).There were equal numbers of targets (i.e. probes that were shown in the encoding phase) and lures (i.e. probes that were not shown in the encoding phase) for each category. The probes of the three trials within one block were randomly chosen to be a target or a lure (e.g. it was possible that all three trials were lures or targets). After the presentation of one block (i.e. three trials) consisting of same category images, the next trial employed stimuli from another category. Note that there was no break between blocks (only between runs).

Lures were drawn from members of the same category as the studied items. Both targets and lures were recycled throughout the whole experimental session with the restriction that neither had been presented within the last three runs before it was repeated. That is, all images used in the first three runs were novel, and in the following runs, images that were presented in the previous three runs were shown.

The order of lists from different categories was randomized in such a way that the number of switches between different categories was identical. This was done to make sure that there were approximately equal numbers of switches between different categories so that neutral to neutral or emotional to emotional switches happened equally often.

fMRI protocol

Scanning was performed on a Siemens 3 T Magnetom TRIO MRI system with a 32-channel head coil in the National Magnetic Resonance Imaging Center at Bilkent University. Functional images were acquired over six runs in each session using a gradient echo planar imaging (EPI) sequence (TR = 2000 ms, TE = 30 ms, flip angle = 90°, FOV = 192 mm, 34 interleaved axial slices, voxel size = 3 mm × 3 mm × 3 mm with 0.3 mm interslice gap). After each session, high-resolution T1-weighed (MP-RAGE) anatomical images were acquired.

Pre-processing

SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) was used to process images which included slice timing correction, realignment, normalization, and smoothing. Functional images were first corrected for differences in slice acquisition timing by resampling all slices in time to match the first slice, then realigned for motion correction and resliced. Resliced images were then normalized to MNI stereotaxic space using a 12-parameter affine transformation along with a non-linear transformation using cosine basis functions. Functional images were then smoothed with an 8-mm FWHM isotropic Gaussian kernel. Image data quality was assured via visual inspection and runs

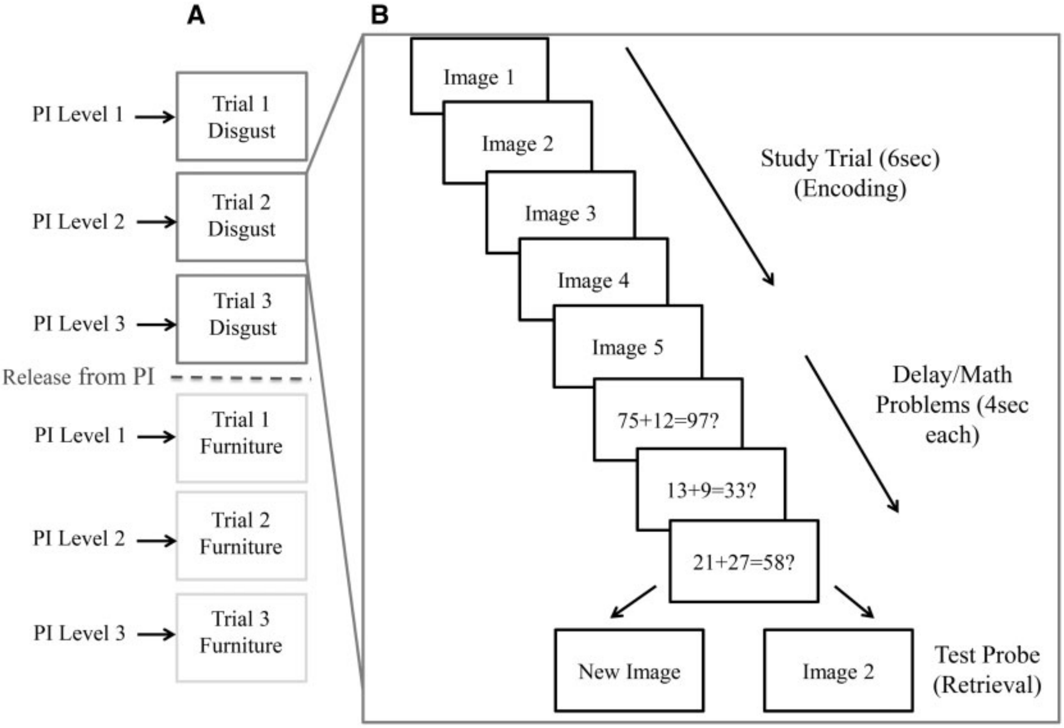

Fig. 1. (A) Illustration of the proactive interference (PI) manipulation. Each block (in blue or green) consists of three trials in which the presented images come from the same category. The category is switched for the next block of trials. PI is expected to increase from Level 1 at Trial 1 to Level 3 at Trial 3 within a block. PI will be released when the category is switched and the PI Level will return to 1. (B) Illustration of the sequence of events within a single trial. Each trial began with the encoding phase in which 5-images were presented sequentially for 1200 ms each. Following the fifth image, participants solved three math problems consisting of addition or subtraction of two randomly selected two-digit numbers which were presented for 4000 ms each. Participants indicated whether the solution presented next to the math problem was accurate by pressing either the middle or index finger on the button box. Following the third math problem, participants were presented with a test image for 2000 ms and asked to indicate whether the image was shown during the current encoding phase. The test image was either a study list item (e.g. Image 2) or an image which was not presented within the experimental session (e.g. New Image). The inter-trial interval consisted of the presentation of a fixation cross in the center of the screen for a fixed duration of 12 000 ms.

Note. Images that are used in the study are not presented due to copyright reasons. For detailed description of the stimuli and the stimuli selection process please see Supplementary Material. The IAPS image numbers of the stimuli used here are also given there.

in which there were excessive motion (>3 mm) were excluded from the analysis.

fMRI data analysis

A general linear model was built in SPM8 by generating separate regressors for each condition [separate encoding for each of the three trials within a block for emotion and neutral stimuli, distractor period (collapsed across emotion and neutral trials), recognition probes for each of the three trials within a block for emotion and neutral stimuli] and were modeled using a canonical hemodynamic response function and its temporal derivative. For each of the two scanning sessions, the data across the six runs were concatenated and modeled as one session, with mean signal and scanner drift entered as covariates to account for the variability across runs.

Percent BOLD signal change was assessed by averaging the time-series signal across the region of interests (ROIs) by using the MarsBaR region of interest toolbox for SPM8 (http://marsbar.sourceforge.net/). We calculated integrated percent signal change (iPSC) by averaging the peak time point and the two adjacent time points to the peak (±1 TR) for each ROI. Bilateral parahippocampal gyrus (PHg) and amygdala ROIs were defined using anatomical masks from the Automated Anatomical Labeling (AAL) database (Tzourio-Mazoyer *et al.*, 2002). For left anterior VLPFC ROIs we assessed pre-defined anatomical masks which were used in Öztekin and Badre's (2011) study. Specifically, the anterior VLPFC ROI was restricted to the pars orbitalis portion of the left inferior frontal gyrus, located ventrally to the horizontal Sylvian ramus.

Modeling recognition decision

The diffusion model was used to model responses and corresponding response times (Ratcliff, 1978; Ratcliff and Rouder, 1998; Ratcliff *et al.*, 2016).We employed the four-parameter Wiener model (e.g. Wabersich and Vandekerckhove, 2014) augmented by a drift criterion parameter (Ratcliff and McKoon, 2008). In our task, the lower decision bound corresponded to 'new' responses (i.e. when the participant decided that the probe was novel); the upper bound corresponded to 'old' responses (i.e. when the participant decided that the probe was shown during encoding). The diffusion model assumes that across time within each trial evidence accumulates in a noisy (diffusion) process. If the evidence hits one of the two decision boundaries the corresponding response is given. The first parameter, the drift rate, v, with which the accumulation process approaches one of the boundaries given a specific stimulus class (i.e. old or new probe), is a measure of the strength of evidence resulting from the retrieval processes. In our model, the same drift rate was used for old and new probes with the only difference being their sign (positive for old probes and negative for new probes). The larger the absolute value of the drift rate the stronger the mnemonic evidence in that condition. The position of the response boundaries is captured by the boundary separation parameter a. The larger the a, the more evidence is necessary until a decision is made (i.e. the decision maker is more cautious).

Evidence starts to accumulate from a starting point which might be closer or at equal distance to one of the boundaries. This response bias parameter, z, measures the bias participants have towards one of the boundaries, with $z = 0.5$ indicating no bias. Participants might be biased towards an old response (if $z > 0.5$) in which case giving an old response would require less evidence. Note that response bias shifts do not have to result in a change in the drift rate.

An additional possibility to account for response bias is via the drift criterion parameter, d_c. As mentioned above, we initially assume that the absolute value of the drift rate is identical for old and new probes. The drift criterion removes this assumption. More specifically, the value of the drift criterion is added to both drift rates, the positive drift for new items and the negative drift rate for old items and thereby represents a symmetric shift in the available evidence (while not affecting its absolute magnitude). Values above zero indicate that in a given condition the evidence is shifted towards old probes whereas values below zero indicate a shift towards new probes (Ratcliff and McKoon, 2008). Finally, our model also included a non-decision parameter, t_0, to capture all processes not related to the mnemonic decision such as encoding and motor processes. Our model can be seen as an extension of a signal-detection model (Macmillan and Creelman, 2005) that incorporates response times in addition to binary responses. In this view, the drift rate corresponds to the signal-detection sensitivity (e.g. d') and the drift criterion to the signal-detection response bias (e.g. c).

Studies examining the impact of emotion on recognition memory suggest that emotion has an impact on the response criterion, traditionally measured within signal detection framework. By separating response bias and drift criterion, we can a) evaluate the retrieval strategies participants adapt to changing levels of PI and b) investigate whether emotion has an impact on the response bias or drift criterion.

We employed a hierarchical Bayesian approach to model the data (Gelman *et al.*, 2013). For each of the five diffusion model parameters (i.e. v, a, z, d_c, t_0) we estimated six hyperparameters to describe the full design [(i.e. stimulus type (emotion & neutral) and PI level (1, 2, & 3)] using treatment contrasts separately for emotional and neutral stimuli (i.e. in each emotion condition Level 1 was the intercept and we estimated two parameters quantifying the differences of Level 2 and Level 3 from the intercept). In total, the model had 30 hyperparameters and we estimated the full variance-covariance matrix for all individual level parameters (Klauer, 2010). The model provided a good account to the data. See Supplementary Material for more details.

Results

Hierarchical-Bayesian diffusion model analysis

Diffusion models have often been used to model recognition memory decisions (e.g. Ratcliff and Starns, 2013; Starns *et al.*, 2012; White and Poldrack, 2014). Consistent with prior work showing that PI slowed down the retrieval speed (Öztekin and McElree, 2007), we hypothesized that PI would reduce the efficiency of evidence accumulation (i.e. decrease the drift rate) due to reduced memory signal quality with increasing similarity between memoranda.

Results confirmed this prediction. As can be seen in Figure 2 (panel 'Drift Rate') the drift rate decreased linearly for neutral trials with increasing PI levels (Level 1 vs. Level 2: $p_B = 0.01$; Level 2 vs. Level 3: $p_B = 0.01$).[1] However, while descriptively such a pattern was also obtained for emotion trials, the evidence for it was comparatively weak. There was no credible decrease from Level 1 to Level 2 ($p_B = 0.27$) and neither from Level 2 to Level 3 ($p_B = 0.11$). Only when comparing Level 1 and Level 3 we found a credible decrease in drift rate ($p_B = 0.03$). When comparing the decreases in drift rate due to PI across emotion and neutral stimuli there was some evidence for a differential pattern (smallest $p_B = 0.06$). This suggests that PI affected both stimulus types, although there was some evidence that this decrease was less pronounced for emotional stimuli.

For the remaining diffusion model parameters, we did not observe systematic differences between stimulus types and the levels of PI (See the compact letter display in Figure 2). We found that overall, participants showed a response bias towards old responses (i.e. response bias above 0.5) whereas they were more inclined towards evaluating the evidence as more new than old (i.e. drift criterion below 0).

Neuroimaging data

Effects of PI on the recognition of emotion and neutral stimuli. We expected a linear increase in the neural activation as a response to increasing levels of PI for two ROIs, the aVLPFC, and the parahippocampal gyrus of MTL. To test this, we performed two separate linear mixed model (LMM; e.g. Singmann and Kellen, in press) analyses on the integrated percent signal change (iPSC) in the aVLPFC, and parahippocampal gyrus to hit responses only (see method for the definition of regions). Each LMM had fixed effects for stimulus type (Emotion vs. Neutral), a linear and quadratic polynomial for level of PI, and the interaction between stimulus type and the polynomial terms as well as by-participant random intercepts and random slopes for the two main effects (the random slope of the interaction term was non-identified and this model thereby represents the 'maximal model'; Barr *et al.*, 2013). To evaluate the significance of the fixed effects we employed the Satterthwaite approximation via the methods implemented in afex (Singmann *et al.*, 2017). To control for multiple testing adjusted P-values are Bonferroni-Holm corrected across ROIs for each family of tests (i.e. each effect).

Results of the LMMs are displayed in Table 1 and show that none of the adjusted p-values reaches significance. However, the interaction of the linear effect for PI with stimulus type for PHg was marginally significant (as well as unadjusted $P < 0.05$). To assess whether we could directly replicate the results from Öztekin and Badre (2011) showing a linear increase in PHg activity in response to increase in PI levels, we tested the linear effect of PI separately for both stimulus types. For neutral stimuli we found a linear effect, $\beta = 0.05$, 95%-CI [0.02, 0.10], but not for emotional stimuli, $\beta = 0.00$, 95%-CI [−0.04, 0.04]. Thus, we could replicate

[1] p_B-values reported in this manuscript are calculated from the posterior distributions of the hyperparameters and based on the probability that specific parameter or difference of parameters is smaller or larger than 0 (i.e. values > 0.5 are subtracted from 1). In line with a Bayesian framework we speak of *credible* differences if $p_B < 0.05$.

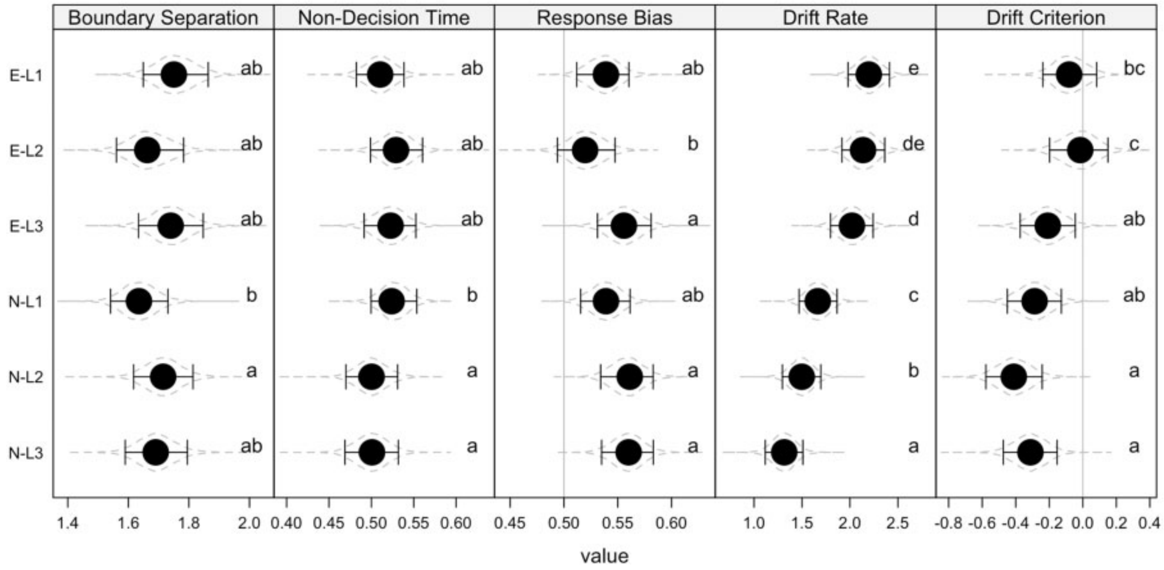

Fig. 2. Parameter estimates from the diffusion model. E = Emotion, *n* = Neutral, L1 = Level 1, L2 = Level 2, L3 = Level 3. The points show the posterior modes, the error bars the 90% highest-posterior density regions, and the gray dashed lines the (mirrored) density estimates of the full posterior. The letters in each plot represent a compact letter display (CLD; Piepho, 2004) presentation of the difference between conditions. Conditions that do *not* share a letter within one plot differ significantly from each other with $p_B < 0.05$. For 'Response Bias' and 'Drift Criterion' the vertical gray line indicates no bias.

Table 1. LMM results for ROIs which were expected to be affected by PI

ROI	Effect	df	F	P	P_{adj}
Anterior VLPFC	Stim Type	1, 18.12	$3.99^{\dagger}$	0.06	0.07
	PI (linear)	1, 18.49	0.65	0.43	0.43
	PI (quadratic)	1, 29.78	0.22	0.64	0.85
	Stim Type × PI (linear)	1, 54.00	0.26	0.61	0.61
	Stim Type × PI (quadratic)	1, 54.00	1.31	0.26	0.52
PHg	Stim Type	1, 18.13	$5.21^{\dagger}$	0.03	0.07
	PI (linear)	1, 27.49	3.90	0.06	0.12
	PI (quadratic)	1, 18.98	0.66	0.43	0.85
	Stim Type × PI (linear)	1, 54.00	$4.15^{\dagger}$	0.05	0.09
	Stim Type × PI (quadratic)	1, 54.00	1.13	0.29	0.52

Notes. The dependent variable was iPSC. P_{adj}: Adjusted *P*-values are Bonferroni-Holm corrected across ROIs for each effect. †: adjusted-$P < 0.1$.

We were also interested in the effect of stimulus class and PI on amygdala activation. However, we had no reasons to expect a linear effect of PI on amygdala. Therefore, we estimated an LMM with amygdala iPSC as dependent variable, fixed effects for stimulus type, PI (with three levels and no linear or quadratic effect), and their interaction, and by-participants random intercepts and random slopes for the two main effects. For this analysis we restricted the overall probability of a Type I error to 0.05 using the Bonferroni-Holm correction. This analysis revealed a significant effect of stimulus type [$F(1, 18.03) = 12.37$, $P_{adj} = 0.007$] in the expected direction. Emotional stimuli lead to larger iPSC values than neutral stimuli. None of the other effects reached significance, smallest $P_{adj} = 0.16$.

Öztekin and Badre (2011) for the type of stimuli they had also used (i.e. neutral items), but not for the novel stimuli introduced here.

Linking latent processes with neural measures

Our ROI analysis replicated Öztekin and Badre's (2011) results and showed a linear increase in PHg activation for hits to neutral items as a function of PI. As we also observed a linear decrease in drift rate as a function of PI, we investigated the relationship between PHg activation for hit responses and drift rate. More specifically, we estimated a LMM with the individual drift rate estimates (3 per participant and stimulus type, we used the individual posterior means) as dependent variable with fixed effects for PHg iPSC, stimulus type, and their interaction, as well as a by-participant random intercept and by-participant random slopes for all fixed effects.

In addition to a significant effect of stimulus type [$F(1, 37.89) = 37.54$, $P < 0.001$] (see Figure 3), the model showed a significant effect of PHg [$F(1, 50.45) = 7.92$, $P = 0.007$]. This latter effect indicated that within an individual participant a decrease in drift rate was associated with an increase in PHg activation to hits – successfully recognized items. Previous studies showed retrieval of more detailed episodic information recruits regions within parahippocampal gyrus (Diana *et al.*, 2007). In our task, we observe that there is higher PHg activation to correctly recognize items even though the performance lowers due to PI. The negative relationship we observe here between the PHg activation and the drift rate can be interpreted as a response to retrieval demands. The harder the retrieval gets, drift rate decreases and more PHg activation is recruited for successfully retrieving more detailed information which leads to a correct response. The model also revealed a marginally significant interaction between PHg and

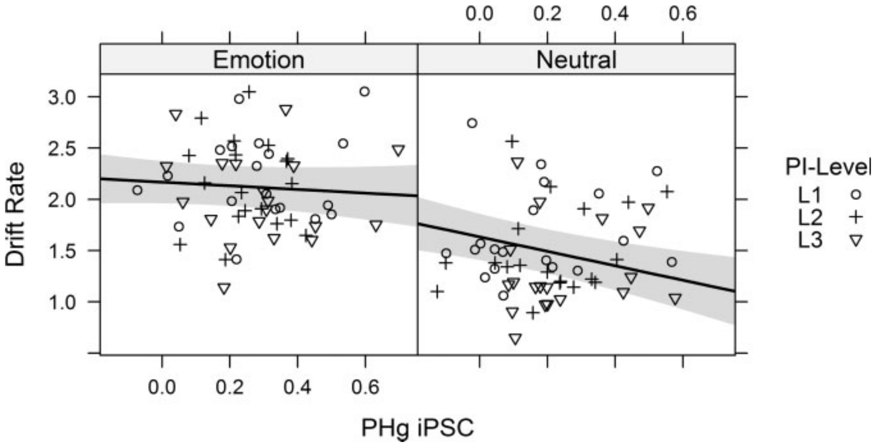

Fig. 3. The interaction between parahippocampal gyrus activation (PHg) and the drift rate for neutral and emotional stimuli. Shaded areas show 95%-confidence bands from the LMM. Note that PI level (L1, L2, or L3) was not part of the LMM and is only added for illustrative purposes.

stimulus type [$F(1, 45.95) = 3.77$, $P = 0.06$] which is displayed in Figure 3. From this interaction, we can see that PHg activation is a predictor for the drift rate for neutral stimuli ($\beta = -0.70$, 95%-CI [-1.14, -0.27]), but this relationship is essentially absent for emotional stimuli ($\beta = -0.18$, 95%-CI [-0.57, 0.22]). This finding mimics our ROI finding that the effect of PI has essentially no PHg involvement for emotional stimuli.

Discussion

The present study aimed to provide new insights into the relationship between emotion and forgetting due to PI from WM. We were specifically interested in differences in how PI affected emotional and neutral stimuli for (i) the buildup of PI reflected in latent processes, and (ii) neural responses to PI. To this end, we used a task that progressively induced PI by presenting categorically similar items over three trials. After one block of three trials, the category was changed which led to a release from PI. We employed a formal measurement model, the diffusion model, that allowed us to estimate and compare how the built up of PI affected latent cognitive processes of recognition memory decisions for emotional and neutral material. We found that PI decreased the rate of evidence accumulation (i.e. the drift rate). Whereas we observed this for both neutral and emotional stimuli, the decrease appeared to be somewhat smaller for emotional stimuli. We also tracked the neural responses to PI during retrieval of emotional and neutral memoranda in regions that have been previously related to overcoming PI. We found that the buildup of PI was associated with a linear increase in PHg activation for neutral stimuli but not for emotional stimuli (although the interaction was only marginally significant after adjusting for multiple testing). Finally, we found that for neutral stimuli an increase in PHg activity was associated with a decrease in drift rate, whereas this association was absent for emotional stimuli (again the interaction was only marginally significant).

Effects of PI on latent processes

In memory studies the drift rate generally represents the quality of the match between probe and memoranda (Ratcliff and McKoon, 2008), it is a measure of the strength of evidence. Items with higher memory strength are expected to have a higher drift rate. Our findings showed that PI decreased the drift rate (i.e. lowered the mnemonic quality/degree of match) in a linear fashion. It has been suggested that one of the factors that affects the strength of evidence in memory retrieval might be the degree of difficulty of recovering the evidence (Badre *et al.*, 2014). In our task, the shared feature (category) elicits a level of match between the probe and memoranda, contaminating the quality of the match and adding noise to the evidence making the recovery of evidence harder.

Our results indicate that the PI-related decrease in the quality of the probe-memoranda match is similar for emotion and neutral stimuli. This is an important finding showing that PI in WM affects memory performance for emotion and neutral stimuli in a similar way. It is noteworthy to state that the decline in drift rate tended to be less pronounced for emotion stimuli such that medium levels of PI did not lead to an effective decline in drift rate. This is consistent with our previous findings in Mizrak and Öztekin (2016) study showing that PI build up was slower and smaller in magnitude for emotion trials compared to neutral trials. We discuss potential explanations for this finding below.

Neural mechanisms that overcome PI for emotion and neutral stimuli

Neutral stimuli. For neutral stimuli, increasing levels of PI led to a gradual decline in the rate of evidence accumulation. This decrease was however accompanied by an increase in activation in parahippocampal regions (PHg) for hits (i.e. 'old' responses to targets). This suggests that participants tried to counteract the effect of PI by increasing memory retrieval. This is in line with previous work showing that MTL regions are involved in successful episodic retrieval (e.g. Dobbins *et al.*, 2003). Additionally, it has been shown that MTL activation during encoding predicted memory success (e.g. Davachi *et al.*, 2003; Paller and Wagner, 2002; Staresina and Davachi, 2006). More recently, MTL activation has been observed during retrieval from WM (Oztekin *et al.*, 2010; Öztekin *et al.*, 2008), also suggesting a relationship between memory accuracy in a WM task and MTL activation. With increasing PI levels, participants need to retrieve more information in order to correctly identify items. This incease in retrieval demands was reflected in PHg activation as well.

We did not detect systematic changes in left aVLPFC activation in response to PI. This finding is surprising as this region had often been suggested to mediate the detrimental effects of PI on memory performance (Badre and Wagner, 2005; Jonides and Nee, 2006; Oztekin *et al.*, 2009). One reason for this might be the use of visual material in our study. Previous studies which showed left anterior VLPFC activation in response to PI used verbal material and there is considerable evidence showing that left VLPFC can be domain sensitive (see Badre and Wagner, 2007 for a review). For instance, Dobbins and Wagner (2005) found that when participants were asked to recollect conceptual details about the studied objects, higher left aVLPFC activation was observed compared to recollection of perceptual details. On the other hand, during the recollection of perceptual details higher right VLPFC activation was observed compared to recollection of conceptual details. Although we expected left VLPFC activation due to previous findings showing that left VLPFC is an important region that mediates the control of PI in working memory, it is possible that left VLPFC effects are specific to verbal material.

Emotional stimuli. For emotional stimuli we only observed a comparatively small linear decline in drift rate as a function of PI. In addition, this decrease was not as strongly accompanied by a corresponding PHg activation increase to hits, if at all. In addition, we did not observe aVLPFC activation in response to PI. Overall, our data suggest no specific neural response as a function of PI for emotional stimuli in our ROIs. However, the evidence for this qualitative difference as a function of PI between emotional and neutral stimuli was not as strong as we expected. Given the clear replication of the PI effect for neutral items on PHg and the complete absence of this effect for emotional items we are nevertheless confident that this differential pattern is noteworthy. One possible explanation for this differential pattern is given in the following section.

Effect of emotion on PI. It has been suggested that the release from PI task specifically necessitates controlled retrieval at the item level. If item-level details are not immediately or automatically available at retrieval, participants need to access more diagnostic information about the item on top of the category membership. One common finding from the emotion-recognition memory literature (Kensinger and Corkin, 2003; Ochsner, 2000; Pierce and Kensinger, 2011; Ritchey *et al.*, 2008; Sharot and Yonelinas, 2008) is that emotion impacts retrieval of detailed information about the event from memory, often called *recollection*, rather than assessments of the overall memory strength of the items in memory, often called *familiarity*. However, emotion specifically enhances recollection of within-item features such as identifying details about the item rather than context (Chiu *et al.*, 2013; Yonelinas and Ritchey, 2015). For instance, font color or spatial location information of emotional items was remembered better than for neutral items (D'Argembeau and Van der Linden, 2005; Doerksen and Shimamura, 2001; Mather and Nesmith, 2008). In contrast, when participants were asked about the scene in which an item was presented, memory performance was worse for emotional items compared to neutral items (see Kensinger *et al.*, 2006). In our task, participants might have benefited from a focus on the central details (item features) in exchange for remembering peripheral details (context) for emotional stimuli. If emotional stimuli are activated in WM with more diagnostic item features, these features can be used as additional cues while making the recognition judgment and the category membership as a cue will have less weight on the recognition judgment. This can explain the less pronounced PI-related decline in drift rate for emotion stimuli compared to neutral stimuli. In this case, the category cue will contaminate the quality of the match between probe and memoranda, but for emotional stimuli not as much as it does for neutral stimuli. Additionally, if item-specific details are enhanced for emotion items, these details will be readily available at retrieval and the need for accessing more detailed information would be less for emotional information. Accordingly, this can explain the absence of the relationship between the drift rate decrease and increase in parahippocampal gyrus activation for emotional stimuli since this region is specifically associated with retrieval demands.

Conclusion

We provided an in-depth investigation of the relationship between emotion and forgetting in WM by modeling the latent processes that are involved in the recognition memory performance of emotioal and neutral items in the presence of PI. We complemented our investigation by analyzing neural responses to changing levels of PI and related our findings from brain and behavioral measurements. Our results replicated quantitative differences in the cognitive processes in response to PI which were extended by our neural measures suggesting a possible qualitative difference between emotional and neutral stimuli. For neutral stimuli individuals increased PHg activation as a response to PI and such an increase was associated with a decrease in drift rate. However, the data did not indicate evidence for such a mechanism for emotional stimuli.

Acknowledgements

We would like to thank Ege Ekin Ozer for her help in the data collection. We also thank the Computational Modeling of Cognition Summer School organized by Stephan Lewandowsky and Klaus Oberauer for their assistance in the modeling part of this project.

Supplementary data

Supplementary data are available at *SCAN* online.

Funding

This research was supported by a Science Academy Young Investigator Award (BAGEP), and grants from the Scientific and Technological Research Council of Turkey (111K220, 112E315), Tubitak 2214 fellowship, and FP7 Marie Curie IRG (277016).

Conflict of interest. None declared.

References

Badre, D., Lebrecht, S., Pagliaccio, D., Long, N.M., Scimeca, J.M. (2014). Ventral striatum and the evaluation of memory retrieval strategies. *Journal of Cognitive Neuroscience*, **26**(9), 1928–48.

Badre, D., Wagner, A.D. (2005). Frontal lobe mechanisms that resolve proactive interference. *Cerebral Cortex*, **15**(12), 2003–12.

Badre, D., Wagner, A.D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, **45**(13), 2883–901.

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**(3), 255.

Chiu, Y.C., Dolcos, F., Gonsalves, B.D., Cohen, N.J. (2013). On opposing effects of emotion on contextual or relational memory. *Frontiers in Psychology*, **4**, 103.

Davachi, L., Mitchell, J.P., Wagner, A.D. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(4), 2157–62.

D'Esposito, M., Postle, B.R., Jonides, J., Smith, E.E. (1999). The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related fMRI. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 7514–9.

Diana, R.A., Yonelinas, A.P., Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Sciences*, **11**(9), 379–86.

Dobbins, I.G., Rice, H.J., Wagner, A.D., Schacter, D.L. (2003). Memory orientation and success: separable neurocognitive components underlying episodic recognition. *Neuropsychologia*, **41**(3), 318–33.

Dobbins, I.G., Wagner, A.D. (2005). Domain-general and domain-sensitive prefrontal mechanisms for recollecting events and detecting novelty. *Cerebral Cortex* (New York, N.Y.: 1991), **15**(11), 1768–78.

Doerksen, S., Shimamura, A.P. (2001). Source Memory Enhancement for Emotional Words. *Emotion*, **1**(1), 5–11.

Eichenbaum, H., Yonelinas, A.P., Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, **30**, 123–52.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B. (2013). Bayesian Data Analysis, Third Edition. Hoboken: CRC Press.

Jonides, J., Nee, D.E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience* **139**(1), 181–93.

Jonides, J., Smith, E.E., Marshuetz, C., Koeppe, R.A., Reuter-Lorenz, P.A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Sciences of the United States of America*,**95**(14), 8410–3.

Kensinger, E.A. (2009). Remembering the details: effects of emotion. *Emotion Review*, **1**(2), 99–113.

Kensinger, E.A., Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words?. *Memory & Cognition*, **31**(8), 1169–80.

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.

Kensinger, E.A., Schacter, D.L. (2008). Memory and emotion, In M. Lewis, J.M. Haviland-Jones, L.F. Barrett (Eds). In Handbook of emotions (3rd ed., pp. 601–17). New York: Guilford.

Klauer, K.C. (2010). Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach. *Psychometrika*, **75**(1), 70–98.

Lang, P.J., Bradley, M.M., Cuthbert, B.I. (2005). *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual* (Technical Report A-6.). University of Florida, Gainesville, FL.: University of Florida.

Macmillan, N.A., Creelman, C.D. (2005). *Detection Theory: A User's Guide* (2nd ed). Mahwah, N.J: Lawrence Erlbaum Associates.

Mather, M., Nesmith, K. (2008). Arousal-Enhanced Location Memory for Pictures. *J Mem Lang*, **58**(2), 449–64.

Mızrak, E., Öztekin, I. (2016). Relationship between emotion and forgetting. *Emotion (Washington, D.C.)*, **16**(1), 33–42.

Nairne, J.S. (2002). Remembering over the short term: The Case Against the Standard Model. *Annual Review of Psychology*, **53**, 53–81.

Ochsner, K.N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology. General*, **129**(2), 242–61.

Öztekin, I., Badre, D. (2011). Distributed patterns of brain activity that lead to forgetting. *Frontiers in Human Neuroscience*, **5**, 1–8.

Oztekin, I., Curtis, C.E., McElree, B. (2009). The medial temporal lobe and the left inferior prefrontal cortex jointly support interference resolution in verbal working memory. *Journal of Cognitive Neuroscience*, **21**(10), 1967–79.

Oztekin, I., Davachi, L., McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store. *Psychological Science*, **21**(8), 1123–33.

Öztekin, I., Gungor, N.Z., Badre, D. (2012). Impact of aging on the dynamics of memory retrieval: A time-course analysis. *Journal of Memory and Language*, **67**(2), 285–94.

Öztekin, I., McElree, B. (2007). Proactive interference slows recognition by eliminating fast assessments of familiarity. *Journal of Memory and Language*, **57**(1), 126–49.

Öztekin, I., McElree, B., Staresina, B.P., Davachi, L. (2008). Working Memory Retrieval: Contributions of the Left Prefrontal Cortex, the Left Posterior Parietal Cortex, and the Hippocampus. *Journal of Cognitive Neuroscience*, **21**(3), 581–93.

Paller, K.A., Wagner, A.D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, **6**(2), 93–102.

Piepho, H.-P. (2004). An Algorithm for a Letter-Based Representation of All-Pairwise Comparisons. *Journal of Computational and Graphical Statistics*, **13**(2), 456–66.

Pierce, B.H., Kensinger, E.A. (2011). Effects of emotion on associative recognition: valence and retention interval matter. *Emotion*, **11**(1), 139–44.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**(2), 59–108.

Ratcliff, R., McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, **20**(4), 873–922.

Ratcliff, R., Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**(5), 347–56.

Ratcliff, R., Smith, P.L., Brown, S.D., McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends in Cognitive Sciences*, **20**(4), 260.

Ratcliff, R., Starns, J.J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychological Review*, **120**(3), 697–719.

Ritchey, M., Dolcos, F., Cabeza, R. (2008). Role of amygdala connectivity in the persistence of emotional memories over time: an event-related FMRI investigation. *Cerebral Cortex*, **18**(11), 2494–504.

Sharot, T., Yonelinas, A.P. (2008). Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition*, **106**(1), 538–47.

Singmann, H., Bolker, B., Westfall, J. (2017). *afex: Analysis of Factorial Experiments*.

Singmann, H., Kellen, D. (in press). An introduction to linear mixed modeling in experimental psychology. In: D.H. Spieler, E. Schumacher, New Methods in Neuroscience and Cognitive Psychology. Psychology Press.

Staresina, B.P., Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *The Journal of Neuroscience*, **26**(36), 9162–72.

Starns, J.J., Ratcliff, R., McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, **64**(1–2), 1–34.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., *et al.* (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, **15**(1), 273–89.

Wabersich, D., Vandekerckhove, J. (2014). Extending JAGS: a tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavior Research Methods*, **46**(1), 15–28.

Watkins, O.C., Watkins, M.J. (1975). Build-up of proactive inhibition as a cue overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, **1**(4), 442–52.

White, C.N., Poldrack, R.A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **40**(2), 385–98.

Wickens, D.D. (1970). Encoding categories of words: an empirical approach to meaning. *Psychological Review*, **77**(1), 1–15.

Yonelinas, A.P., Ritchey, M. (2015). The slow forgetting of emotional episodic memories: an emotional binding account. *Trends in Cognitive Sciences*, **19**(5), 259–67.