Unintentional and intentional recognition rely on dissociable neurocognitive mechanisms

Zara M. Bergström, David G. Williams, Mariam Bhula & Dinkar Sharma

School of Psychology, Keynes College, University of Kent, Canterbury, CT2 7NP, UK

Abstract

Distractibility can lead to accidents and academic failures, as well as memory problems. Recent evidence suggests that intentional recognition memory can be biased by unintentional recognition of distracting stimuli in the same environment. It is unknown whether unintentional and intentional recognition depend on the same underlying neurocognitive mechanisms. We assessed whether human participants' recognition of previously seen (old) or not seen (new) target stimuli was affected by whether a to-be-ignored distractor was old or new. ERPs were recorded to investigate the neural correlates of this bias. The results showed that the old/new status of salient distractors had a biasing effect on target recognition accuracy. Both intentional and unintentional recognition elicited early ERP effects that are thought to reflect relatively automatic memory processes. However, only intentional recognition elicited the later ERP marker of conscious recollection, consistent with previous suggestions that recollection is under voluntary control. In contrast, unintentional recognition was associated with an enhanced late posterior negativity, which may reflect monitoring or evaluation of memory signals. The findings suggest that unintentional and intentional recognition involve dissociable memory processes.

Introduction

The ability to ignore irrelevant distracting stimuli that interfere with our current goals is a critical skill for achieving many everyday tasks, such as driving or studying. Although progress has been made in understanding how distraction can impair perception and general decision-making (e.g. Gazzaley & Nobre, 2012), less is known regarding how distraction affects episodic recognition. In most lab-based memory experiments, researchers do their best to minimise distraction by presenting only the stimuli that participants are supposed to evaluate. However, in real life, we often need to recognise one stimulus in the context of multiple irrelevant stimuli in the same environment. Thus, distraction effects on recognition may be prevalent outside of the laboratory.

Previous research has shown that participants' recognition of previously seen (old) or not seen (new) target stimuli can be biased by whether a simultaneously presented to-be-ignored distractor is old or new (Anderson, Jacoby, Thomas, & Balota, 2011; Ste-Marie & Jacoby, 1993). That is, people are more likely to report that a target item is old if the distractor item is also old, and are more likely to report that a target item is new if the distractor item is also new, despite being explicitly instructed to always ignore the distractors. Similar results have been found in the literature on context effects on recognition memory, where previously encountered contexts can bias recognition responses to items superimposed on these contexts (e.g. Murnane, Phelps, & Malmberg, 1999; Hockley, Bancroft, & Bryant, 2012). Distractor-induced recognition biases are enhanced in young people when an additional task is conducted simultaneously that taxes their cognitive control abilities (Anderson et al., 2011; Ste-Marie & Jacoby, 1993), whereas older people show large recognition biases even without a secondary task (Anderson et al., 2011). In both young and old people, recognition biases are more likely when the distractors are pictures and targets words, rather than vice versa (Anderson et al., 2011). These findings suggest that a failure of control mechanisms to confine processing to targets can trigger unintentional recognition of particularly salient distractors, and that the memory signal from distractors is then misattributed to targets.

A large body of research has suggested that intentional recognition is supported by multiple distinct retrieval processes, including rapid and relatively automatic assessments of item familiarity as well as slower, more controlled recollection of contextual details from a specific previous encounter with an item (Mandler, 1980; Yonelinas & Jacoby, 2012). This account predicts that unintentional recognition of distractors is more likely driven by familiarity than recollection, because familiarity is more automatic than recollection (Anderson et al., 2011). Thus far however there is no direct evidence on this issue, perhaps because the memory processes associated with unintentional distractor recognition can only be indirectly observed with behavioural measures in terms of their biasing effect on target memory judgements.

Therefore, we used EEG to directly measure recognition-related brain responses elicited by both targets and distractors, in order to reveal the underlying neurocognitive processes associated with intentional and unintentional recognition. Young healthy adults were given a recognition test, where in each trial, an old or new picture were simultaneously presented with an old or new word. In the first experiment, participants made recognition judgements on words whilst ignoring pictures, whereas they made judgements on pictures whilst ignoring words in the second experiment. We then measured how well-known ERP markers of familiarity and recollection (Rugg & Curran, 2007; Rugg et al., 1998; Wilding & Ranganath, 2012) were modulated by distractor and target old/new status.

We expected both intentional and unintentional recognition to elicit an early frontal ERP positivity– the FN400 – which is thought to reflect a relatively automatic familiarity process (Curran, 2000; Rugg, et al., 1998; although see Paller, Voss, & Boehm, 2007). In contrast, intentional target recognition was predicted to be uniquely associated with a later parietal ERP positivity that indexes conscious recollection and that can be voluntarily controlled (Bergström, Velmans, De Fockert, Richardson-Klavehn, 2007; Bergström, De Fockert, Richardson-Klavehn,2009a; 2009b; Bergström, Anderson, Buda, Simons, & Richardson-Klavehn, 2013a; Hanslmayr et al., 2009; Hu, Bergström, Bodenhausen, & Rosenfeld, 2015; Mecklinger, Parra, & Waldhauser, 2009). Unintentional recognition was instead expected to engage additional post-retrieval monitoring processes (Rugg & Wilding, 2000) that may be recruited to evaluate the automatic memory signals elicited by old distractors in order to counteract their biasing influence. Such post-retrieval monitoring was expected to be manifest as late ERP slow-drifts (e.g. Johansson & Mecklinger, 2003). Finally, we expected that behavioural and ERP correlates of unintentional recognition to be primarily expressed when distractors were visually salient pictures rather than less salient words (Anderson et al., 2011).

Methods

Participants

Twenty-four right-handed, native English speakers participated in each experiment (Experiment 1: *M* age 21, range 18-24, 8 males, 16 females; Experiment 2: *M* age 19, range 18-22, 5 males, 19 females). Participants were recruited through the offer of course credit or were awarded money for their participation. All participants gave written informed consent and the experiment was approved by the University of Kent Psychology Research Ethics Committee.

Materials

Stimuli were 336 words taken from the ANEW database (Bradley & Lang, 1999) with valence ratings ranging from 3.79 to 7.58 on a 9-point scale, and 336 colour photographs of a range of objects, events and scenes. Of the photographs, 277 were from the IAPS database (Lang, Bradley, & Cuthbert, 2008) with valence ratings between 1.51 to 6.62 on a 9-point scale, and 43 were from the GAPED database (Dan-Glauser & Scherer, 2011) with valence ratings between1.35 to 45.7 on a 100-point scale. We initially aimed to investigate the effect of emotional valence of distracting pictures on word target recognition and ERPs, so half of the pictures were emotionally negative whereas the other half were neutral. However as there were no effects of valence on behaviour or ERPs in either experiment, all results are presented collapsed across this factor. The words ranged from four to eight letters in length, each of no more than two syllables. Of each type, 16 were assigned to a practice round and the remaining 320 were used in the experiment. Assignment of words and pictures to experimental conditions was fully counterbalanced across participants.

Design and Procedure

Our experimental design was closely based on the "Memory Stroop" paradigm devised by Anderson et al. (2011), which was designed to investigate the effect of unintentional distractor recognition on intentional target recognition performance. The name of this paradigm stems from its similarities with the traditional colour-word Stroop (1935) task,

which investigates the effect of unintentional word reading on people's ability to name the ink in which the word is printed (see Anderson et al., 2011, for more detail). Participants were first given task instructions and completed a short practice phase. The actual experiment consisted of ten study-test cycles (split into multiple cycles to ensure adequate recognition test performance, in line with Anderson, et al., 2011). In each study phase, 16 pictures and 16 words were presented individually and randomly interspersed at the centre of the screen for 3000ms, preceded by a 500ms fixation cross. Participants rated the pleasantness of the words and pictures on a scale between one and four (with one being very unpleasant and four being very pleasant) by pressing buttons on a keyboard, and were told that their memory for all items would later be tested.

In each target recognition test phase, 32 pairs of pictures and words were presented with the word superimposed over the picture. Each phase contained four combinations: old word and old picture (8 trials), old word and new picture (8 trials), new word and old picture (8 trials), and new word and new picture (8 trials), displayed in random order. Participants were asked to press one keyboard button if they recognised the target stimulus (words in Experiment 1 and pictures in Experiment 2) as "old" (i.e. previously presented in the experiment), and another button to classify the target as "new" (i.e. not previously presented in the experiment), with response hand counterbalanced across participants. Participants were instructed to always ignore distractor stimuli (pictures in Experiment 1 and words in Experiment 2). Each trial began with a 500ms fixation cross, after which the word and picture pair was presented for 3000ms, and participants were asked to respond whilst the pair was still on the screen.

During each test phase, participants also did a simultaneous working memory task, since previous research had shown that unintentional distractor recognition is more likely during dual task conditions. Anderson et al. (2011) and Ste-Marie and Jacoby (1993) found that young participants showed larger distractor-induced recognition biases when they were given a secondary task that involved listening to a recorded list of digits and verbally responding whenever they detected particular sequences of digits. However, since a continuous listening task with verbal responses would interfere with the EEG recording, we used an alternative secondary task that involved covert rehearsal of digit sequences, which have been shown to increase interference from distractor processing in other tasks (e.g. De Fockert, Rees, Frith, & Lavie, 2001). To this end, a number string of five digits (0-4, always beginning with 0 but with 1-4 in random order) was shown for 3000ms every 4-6 trials (randomly determined), and participants were instructed to maintain the sequence of numbers in WM whilst completing the episodic recognition task. After 4-6 trials, a single digit probe was displayed for 3000ms and participants pressed the number corresponding to the next digit in the number sequence that they were currently rehearsing. To encourage participants to pay attention to the WM task, they were given visual feedback regarding the accuracy of each response (either "incorrect", "correct", or "no response"). Next, participants were shown a new number sequence to maintain in WM during the following 4-6 recognition trials until the next probe. No number sequence was repeated within the same test cycle.

Following the target recognition test in each cycle, participants were given a very short distractor recognition test consisting of two previously seen distractors intermixed with two novel item from the same stimuli class (pictures in Experiment 1 and words in Experiment 2) and were asked to press one button to classify distractors as "old" (previously seen) and another to classify them as "new" (not seen at any point in the experiment). The purpose of this test was to ensure that participants attempted to memorise distractors as well as targets during the study phases since both item types would be tested. Stimulus

presentation durations and response buttons during the distractor test were the same as in the target recognition test.

EEG recording and analysis

EEG was recorded at 500 Hz with a 0.05-70Hz bandwidth using FCz as the reference electrode for 64 scalp electrodes placed in an actiCAP (Brain Products GmbH), with locations according to the extended 10-20 system. EOG was recorded from below the left eye (VEOG) and from the right outer canthi (HEOG). Recorded data were analysed using EEGLAB (Delorme & Makeig, 2004). The EEG was re-referenced to the average of the mastoids and segmented into 1700ms epochs (including a 200ms pre-stimulus baseline) that were time-locked to the onset of the word-picture pair in the target recognition test. Epochs were concatenated and submitted to extended infomax Independent Component Analysis using runica from the EEGLAB toolbox, with default extended-mode training parameters (Delorme & Makeig, 2004). Independent components reflecting eye movements and other sources of noise were identified by visual inspection of component scalp topographies, time courses and activation spectra, and were discarded from the data by backprojecting all but these components to the data space. Corrected data were subsequently lowpass filtered digitally at 30Hz (two-way least-squares finite impulse response filter). Any trials that still contained visible artefacts following filtering were removed, as were trials were participants failed to respond within the allocated time. Only a very small percent of trials (5% in Experiment 1 and 3% in Experiment 2) were deleted in total. Finally, ERPs were formed for the four conditions: Old Word Old Picture (Mean trial numbers in Experiment 1: 76.4; Experiment 2: 77.7), Old Word New Picture (Mean trial numbers in Experiment 1: 75.8; Experiment 2: 77.4), New Word Old Picture (Mean trial numbers in Experiment 1: 76.3; Experiment 2: 77.8), and New Word New Picture (Mean trial numbers in Experiment 1: 76.4; Experiment 2: 77.4).

We first tested our specific predictions by statistically analysing ERP mean amplitudes from two time windows and electrode sites where the FN400 and left parietal old/new effects are typically maximal, 300-500ms at the mid frontal site (Fz) and 500-800ms at the left parietal site (P3) respectively. These *a-priori* selected time windows and locations were based on a large body of previous research (reviewed in Rugg & Curran, 2007). The targeted analysis did not include the late ERP slow-drifts that are thought to index retrieval monitoring processes. This is because retrieval monitoring-related slow drifts can have very different scalp distributions across studies (c.f. Johansson & Mecklinger, 2003; Hayama, Johnson & Rugg, 2008) so we were unable to make clear predictions about their spatial locations.

Because selecting only a few time-windows and electrode sites for analysis may overlook effects at other sites and time-points, we also conducted a whole-head, fully data-driven multivariate "non-rotated" Task Partial Least Square analysis (Task-PLS, McIntosh & Lobaugh, 2004). PLS allows the examination of distributed patterns of spatial and temporal dependencies in the ERP data with minimal assumptions regarding the timing and distribution of potential effects. Task-PLS analyzes the "cross-block" covariance between the spatiotemporal ERP distribution and orthogonal contrast vectors representing differences between experimental conditions. In nonrotated PLS (Bergström et al., 2009a, 2009b; 2013a; McIntosh & Lobaugh, 2004) the sums of squares of the cross-block covariance between the contrast vector and the spatiotemporal data matrix are directly tested for significance using random permutation test, thus allowing a direct assessment of the hypothesised experimental effects. Correction for multiple comparisons is not required, because the PLS only tests the

same number of contrasts as degrees of freedom in the design. The PLS analysis outputs electrode saliences that identify the electrodes that most strongly covary at a particular point in time with the experimental effect expressed in the contrast vector. The standard errors of the electrode saliences are estimated through bootstrap resampling. The ratio of the electrode salience to the bootstrap standard error gives a standardized measure of reliability that is approximately equivalent to a z score, whereby values above 1.96 and below −1.96 are reliably different from zero with a 95% confidence interval (McIntosh & Lobaugh, 2004).

In the current analysis, non-rotated Task PLS was used to test the full factorial design with contrasts coding for the main effects of Word and Picture memory status (old vs. new) as well as their interaction term. Data from all scalp channels across the time-window from 0-1000ms post-stimulus were included. The covariance of the experimental contrasts with the spatiotemporal data was tested for significance using 1000 permutations, and the reliability of the electrode saliences was tested using 200 bootstraps. See McIntosh & Lobaugh (2004) for full description of PLS. Matlab code to perform PLS is available at www.rotman-baycrest.on.ca/pls/.

Results

Behaviour

WM task accuracy was similar and high in both experiments (Experiment 1: $M =$ 0.84, $SD = 0.10$; Experiment 2: $M = 0.83$, $SD = 0.13$; $t<1$, $p>0.79$), suggesting that participants complied with instructions and successfully managed to combine performing both tasks. One participant scored lower than 2 standard deviations below the mean in each experiment (Accuracy of 0.57 in Experiment 1 and 0.41 in Experiment 2), but excluding those participants did not change the pattern of results on the target recognition task, therefore all participants were included in the final analysis.

For the recognition task, we first analysed raw hit rates and correct rejection rates in order to make our results directly comparable with previous research (Anderson et al., 2011; Ste-Marie & Jacoby, 1993). Mean accuracy and RT are presented in Table 1.

Table 1. Mean Proportion Accurate Responses and Reaction Times (in msec) of Target Recognition Decisions in Both Experiments.

	Experiment 1, Word targets and Pictures distractors		Experiment 2, Picture targets and Words Distractors	
	Mean Accuracy (SD)	Mean RT (SD)	Mean Accuracy (SD)	Mean RT (SD)
Old Word Old Picture	.93(.06)	1206(200)	.95(.03)	1145(161)
Old Word New Picture	.90(.07)	1164(187)	.97(.03)	1234(187)
New Word Old Picture	.90(.11)	1280(201)	.95(.04)	1134(197)
New Word New Picture	.92(.06)	1248(219)	.97(.03)	1186(211)

In Experiment 1, a 2x2 ANOVA on the accuracy data revealed a significant interaction between Word memory status and Picture memory status ($F(1,23)=7.29$, $p=0.01$, $\eta p^2 = .24$), but no main effects ($Fs<1$, $p>0.75$). When the distracting Picture was New, it decreased the likelihood that Old target Words would be correctly recognised compared to when the distracting Picture was Old ($t(23)=2.84$, $p=0.009$, Cohen's $d = 0.46$, calculated here and subsequently as the difference between means divided by the pooled standard deviation to ensure unbiased effect size estimates; Dunlap, Cortina, Vaslow, & Burke, 1996) whereas New distracting Pictures facilitated correct rejections of New Words compared to Old distracting Pictures, although this difference was only at trend-level significance ($t(23)=1.83$, $p=0.08$, $d = 0.23$).

For reaction times in Experiment 1, the 2x2 ANOVA revealed only significant main effects and no interaction ($F<1$, $p=.70$). RTs were slower for New ($M=1264$, $SEM=42$) than Old ($M=1185$, $SEM=39$) target Words ($F(1,23)=25.89$, $p<0.001$, $\eta p^2 = .53$) and slower for Old ($M=1243$, $SEM=40$) than New ($M=1206$, $SEM=41$) distracting Pictures ($F(1,23)=12.58$, $p=0.002$, $\eta p^2 = .35$).

In Experiment 2, the accuracy pattern was different. A 2x2 ANOVA on the accuracy data revealed only a significant main effect of Picture memory status ($F(1,23)=11.60$, $p=0.002$, $\eta p^2 = .34$) with significantly higher accuracy for New Pictures ($M=.97$, $SEM=.01$) than Old Pictures ($M=.95$, $SEM=.01$). The main effect of Word memory status and the interaction were far from significant ($Fs<1$, $p>0.51$).

Reaction times differences in Experiment 2 were reversed compared to Experiment 1. The 2x2 ANOVA revealed only significant main effects and no significant interaction ($F(1,23)=2.24$, $p=.15$). Now, RTs were slower for New ($M=1210$, $SEM=39$) than Old ($M=1139$, $SEM=36$) target Pictures ($F(1,23)=12.52$, $p=0.002$, $\eta p^2 = .35$) and slower for Old ($M=1190$, $SEM=34$) than New ($M=1160$, $SEM=39$) distracting Words ($F(1,23)=4.87$, $p=0.04$, $\eta p^2 = .18$).

In order to formally assess whether accuracy and RT patterns were qualitatively different across the two experiments, we analysed both measures with 3-way mixed ANOVAs with the factors Experiment, Word memory status, and Picture memory status. For accuracy, there was indeed a significant three-way interaction ($F(1,46)=4.84$, $p=0.033$, $\eta p^2 = .10$), confirming that the Word x Picture memory status interaction was unique to Experiment 1. For RTs, the three-way interaction was not significant ($F(1,46)=1.85$, $p=0.18$, $\eta p^2 = .04$), consistent with the lack of Word x Picture memory status interactions in both experiments. However, Experiment interacted with Word memory status ($F(1,46)=27.89$, $p<0.001$, $\eta p^2 = .38$) and Picture memory status ($F(1,46)=22.87$, $p<0.001$, $\eta p^2 = .33$) factors individually. These two-way interactions arose because RTs were slower for New than Old targets (i.e. New Words > Old Words in Experiment 1, and New Pictures > Old Pictures in Experiment 2) and slower for Old than New distractors (i.e. New Pictures < Old Pictures in Experiment 1, and New Words < Old Words in Experiment 2) in both Experiments, leading to a cross-over pattern when Words and Pictures swapped target/distractor assignment.

In a second analysis, we also calculated independent measures of discrimination and response bias (see Snodgrass & Corwin, 1988), in order to assess whether unintentional recognition of distractors primarily affected participant response biases rather than their ability to discriminate between old versus new targets. The *Pr* discrimination measure is calculated by subtracting each individual's false alarm rate from their hit rate on a recognition task, and thus provides a measure of discrimination between old and new items that is

corrected for response biases. For Experiment 1 therefore, New Word false alarm rates were subtracted from Old Word hit rates separately for when distractor Pictures were old versus new. For Experiment 2, New Picture false alarm rates were subtracted from Old Picture hit rates separately for when distractor Words were old versus new. The *Br* response bias measure is calculated by dividing each individual's false alarm rate by 1-*Pr*. Values of *Br* that are above 0.5 indicate a tendency to guess "Old" rather than "New" when uncertain (a positive response bias), whereas values below 0.5 indicate the opposite tendency. For Experiment 1 therefore, New Word false alarm rates were divided by the *Pr* measure calculated in the previous step, separately for when distractor Pictures were old versus new. For Experiment 2, New Picture false alarm rates were divided by the *Pr* measure calculated in the previous step, separately for when distractor Words were old versus new. These measures are presented in Table 2.

Table 2. Mean discrimination performance (*Pr*) and response bias (*Br*) for target recognition decisions in both experiments.

	Experiment 1, Word targets and Pictures distractors		Experiment 2, Picture targets and Words Distractors	
	Mean *Pr* (SD)	Mean *Br* (SD)	Mean *Pr* (SD)	Mean *Br* (SD)
Old Distractors	.83(.13)	.57(.23)	.93(.06)	.35(.18)
New Distractors	.83(.11)	.44(.20)	.93(.06)	.32(.19)

Two-way ANOVAs with the factors Experiment (1 vs. 2) x Distractor memory status (Old vs. New) on *Pr* and *Br* revealed a significant main effect of Experiment on discrimination ($F(1,46)=14.69$, $p<0.001$, $\eta p^2 = .24$), with significantly higher discrimination in Experiment 2 ($M=.93$, $SEM=.02$) than Experiment 1 ($M=.83$, $SEM=.02$). There was no effect of Distractor memory status and no interaction (both Fs < 0.15). For response bias, there was also a significant main effect of Experiment ($F(1,46)=12.28$, $p=0.001$, $\eta p^2 = .21$) since participants showed a significantly more positive response bias in Experiment 1 ($M=.50$, $SEM=.03$) than in Experiment 2 ($M=.34$, $SEM=.03$). Participants were also significantly more likely to respond "Old" when the distractors were Old ($M=.46$, $SEM=.03$) than New ($M=.38$, $SEM=.03$; $F(1,46)=6.59$, $p=0.014$, $\eta p^2 = .13$). However, the Experiment x Distractor memory status interaction was not significant ($F(1,46)=1.85$, $p=0.18$, $\eta p^2 = .04$).

In both experiments, accuracy on the WM task was negatively (but non-significantly) correlated with the size of the congruency accuracy effect (congruent minus incongruent conditions) on the target recognition task (Experiment 1 $r_s= -.25$, $p=.24$; Experiment 2 $r_s= -.24$, $p=.26$). This finding shows that recognition biases were not simply related to how much participants complied with instructions to divide attention over both tasks, since this account would predict that recognition biases should increase as accuracy on the WM task increased (i.e. a positive correlation). To the contrary, participants who showed larger distraction-induced recognition biases also performed more poorly on the WM task.

In sum, target recognition accuracy and RTs showed qualitatively different patterns in both experiments, albeit in different ways. In Experiment 1 where words were targets and pictures distractors, accuracy was highest on congruent (where targets and distractors had the same memory status, i.e. were both either old or new) compared to incongruent (where

targets and distractors had opposite memory status) trials. RTs however were longer for old distractors than new, and longer for new targets than old targets. This pattern suggests that distractor effects on performance in Experiment 1 cannot be accounted for by a simple speed-accuracy trade-off.

In Experiment 2 where pictures were targets and words distractors, there was no congruency effect and no main effect of distractors on accuracy, only an effect of target memory status whereby new targets were more accurately classified as new than old targets were correctly classified as old. RTs were however similar to Experiment 1, with longer RTs for old distractors than new, and longer RTs for new targets than old targets. This pattern suggests again a lack of a simple speed-accuracy trade-off in Experiment 2.

The analyses of discrimination and response bias measures showed that across both experiments, unintentional recognition of distractors significantly influenced only response bias but not discrimination, consistent with the view that distractors were biasing responses towards the memory status of the distractor rather than influencing participants' ability to discriminate between Old and New targets.

ERPs

Grand-average ERPs from the mid frontal (Fz) and left parietal (P3) electrode sites from both experiments are displayed in Figure 1.

Targeted analysis

FN400 old/new effects

In Experiment 1 where words were targets and pictures distractors, both Old target Words and Old distractor Pictures elicited significantly more positive FN400 amplitudes than New target Words and New distractor Pictures respectively (2x2 ANOVA; main effect of Word memory status: $F(1,23)=6.54$, $p=0.02$, $\eta p^2 = .22$; main effect of Picture memory status: $F(1,23)=28.50$, $p<0.0001$, $\eta p^2 = .55$) but there was no interaction ($F<1$, $p=0.53$).

In Experiment 2 where pictures were targets and words were distractors, Old target Pictures elicited significantly more positive FN400 amplitudes than New target Pictures (2x2 ANOVA; Picture memory status: $F(1,23)=16.78$, $p<0.001$, $\eta p^2 = .42$), but there was no significant old/new effect for distractor Words ($F(1,23)=1.72$, $p=0.20$), nor an interaction ($F<1$, $p=0.86$).

A three-way ANOVA with Experiment as the third factor revealed significant main effects of Word memory status and Picture memory status, with significantly more positive FN400s for Old than New words ($F(1,46)=8.08$, $p=0.007$, $\eta p^2 = .15$) and for Old than New Pictures ($F(1,46)=39.02$, $p<0.0001$, $\eta p^2 = .46$) across both experiments. The FN400 was also significantly more positive in Experiment 1 than Experiment 2 ($F(1,46)=6.26$, $p=0.016$, $\eta p^2 = .12$). However, there were no significant two-way or three-way interactions (Word memory status x Experiment: $F(1,46)=1.74$, $p=.19$, $\eta p^2 = .04$, all other Fs <1, ps>.43).
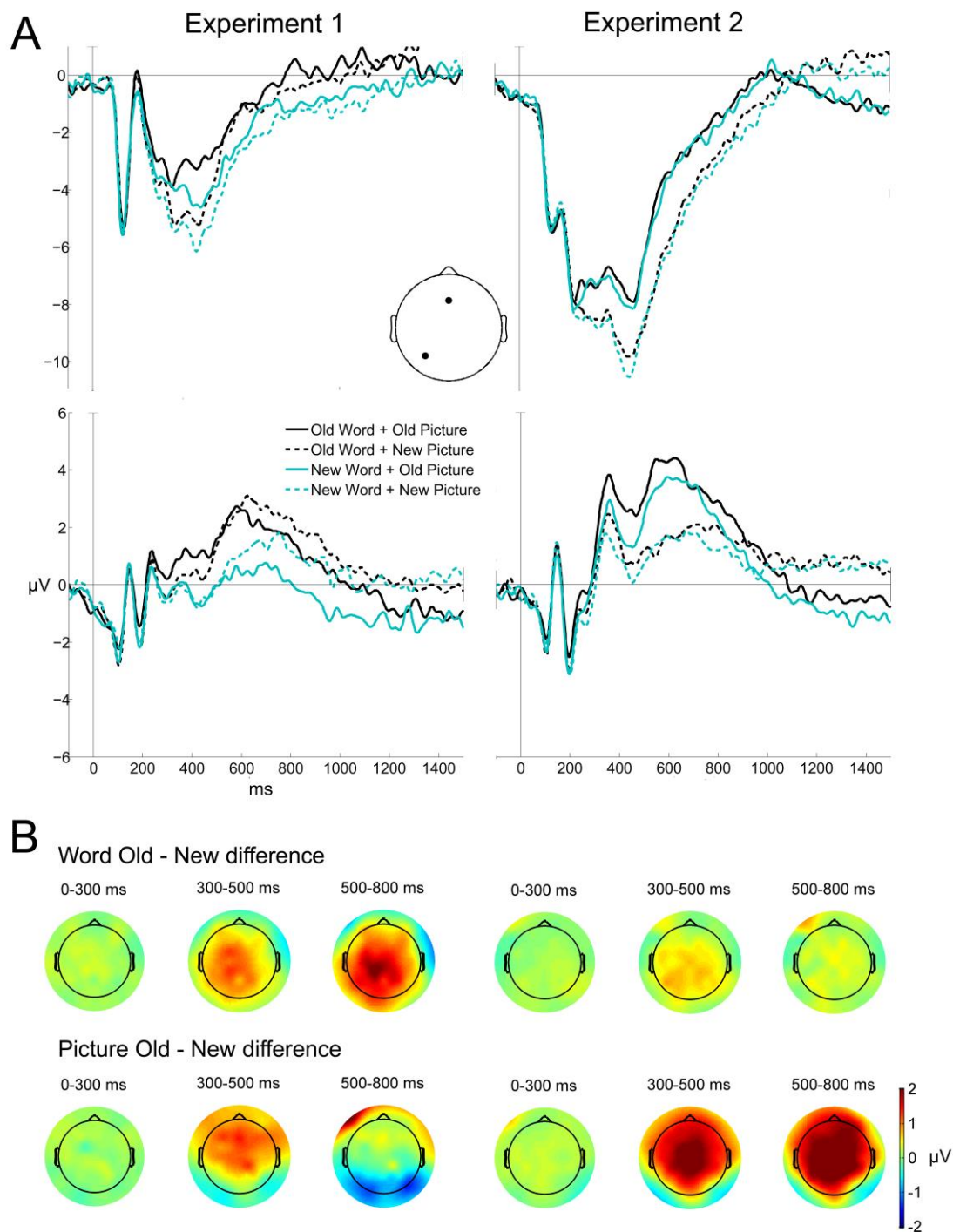
Figure 1. Grand-average ERPs and scalp topographies of old/new effects for targets and distractors in both experiments. **A**, ERPs from mid-frontal (Fz, top row) and left parietal (P3, bottom row) sites in Experiments 1 (left panel) and 2 (right panel). **B**, Scalp topographies of the mean amplitude old minus new difference for Words irrespective of Picture memory status (top row) and the old minus new difference for Pictures irrespective of Word memory status (bottom row) in Experiments 1 (left panel) and 2 (right panel). Whereas Old targets elicited both typical early (300-500ms) and late (500-800ms) positive ERP amplitudes in both experiments, Old Picture distractors (Experiment 1) only elicited an early ERP positivity, and was associated with a later enhanced negativity across posterior sites. Old Word distractors did not differ based on old/new status.

Parietal old/new effects

In contrast to the FN400, a typical increased parietal positivity for Old compared to New items was only found for Word targets in Experiment 1 (2x2 ANOVA; Word memory status: $F(1,23)=31.11$, $p<0.0001$, $\eta p^2 = .58$). Distractor Pictures in fact showed a difference in the opposite direction, with significantly more negative parietal amplitudes for Old compared to New distractors ($F(1,23)=5.94$, $p=0.02$, $\eta p^2 = .21$). There was no interaction between Word and Picture memory status ($F<1$, $p=0.48$).

When pictures were targets in Experiment 2 however, Old target Pictures did elicit significantly more positive parietal ERPs than New target Pictures ($F(1,23)=20.56$, $p<0.001$, $\eta p^2 = .47$), and there was also a non-significant trend for distractor Words in the same direction with more positive ERPs for Old than New distractors ($F(1,23)=3.62$, $p=.07$, $\eta p^2 = .14$). Again, there was no interaction between these factors ($F<1$, $p=0.62$).

A three-way ANOVA with Experiment as the third factor confirmed that both the Word and Picture parietal old/new effects were qualitatively different across the two Experiments, as indicated by significant interactions between Word memory status and Experiment ($F(1,46)=11.75$, $p=0.001$, $\eta p^2 = .20$) and Picture memory status and Experiment ($F(1,46)=26.48$, $p<0.0001$, $\eta p^2 = .37$). The two-way interaction between Word and Picture memory status and the three-way interaction were not significant (both Fs <1, ps$>.40$).

FN400 and parietal old/new effects differences dependent on target status

To confirm that the FN400 and parietal old/new effects were qualitatively different across experiments, we calculated old minus new difference measures for targets and distractors for both effects (average difference between 300-500ms at Fz for the FN400, and average difference between 500-800ms at P3 for the left parietal effect), as displayed in Figure 2. These difference measures were analysed in a 2x2x2 mixed ANOVA with factors Stimulus Type (Word/Picture), ERP Effect (FN400/Parietal old/new effect) and Experiment (1/2). The ANOVA confirmed a significant three way interaction ($F(1,46)=17.91$, $p<0.001$, $\eta p^2 = .28$) which was followed-up with separate Stimulus Type x ERP effect ANOVAs within each experiment.

In Experiment 1, the FN400 and parietal old/new effects showed a qualitatively different pattern for the two stimulus types (interaction: $F(1,23)=37.23$, $p<0.00001$, $\eta p^2 = .62$). For target Words, the parietal old-new difference was significantly larger than the FN400 old/new effect ($t(23)=2.22$, $p=0.04$, $d = 0.49$), whereas for distractor Pictures, the parietal old-new difference was significantly smaller (in fact negative) than the FN400 old-new difference ($t(23)=5.97$, $p<0.00001$, $d = 1.62$).

In Experiment 2, there was only a significant main effect of stimulus type ($F(1,23)=11.49$, $p=0.003$, $\eta p^2 = .33$) whereby the old-new difference was significantly larger on average for Pictures (M=3.71µV, SEM=0.78) than Words (M=0.67µV, SEM=0.32), and there was no interaction with ERP effect ($F<1$, $p=.96$).

There were no significant correlations between the size of the ERP FN400 or parietal old/new effects and individual differences in WM accuracy, nor between the ERP effects and the size of the congruency accuracy effect (congruent minus incongruent conditions) on the recognition task.
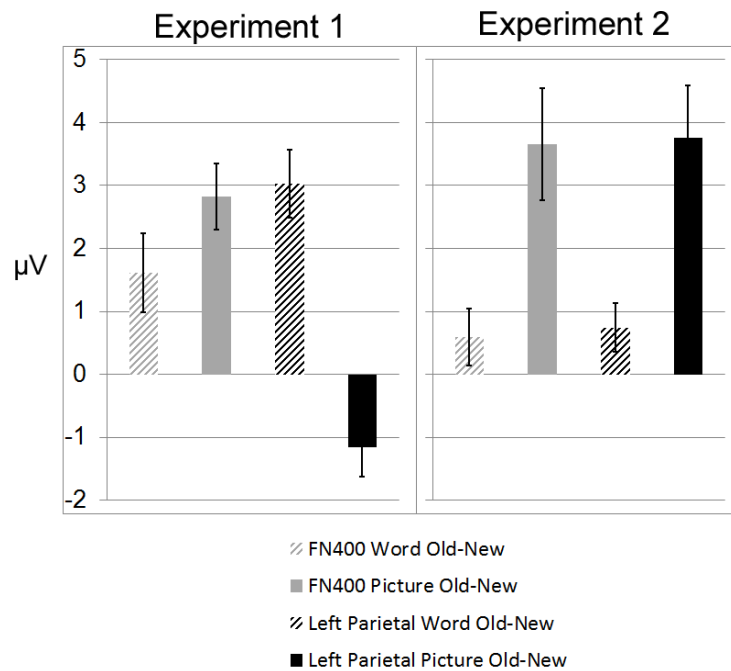
Figure 2. Mean FN400 and left parietal old minus new ERP differences for Words and Pictures in both experiments. Error bars represent one SEM.

In sum, the targeted ERP analysis revealed qualitatively different old/new effects depending on the target/distractor status of Words and Pictures. When participants were asked to recognise Words and ignore Pictures in Experiment 1, both target Words and distractor Pictures were associated with typical early FN400 old/new effects, but only target Words were associated with a later parietal old/new effect, whilst distractor Pictures showed a reversal of typical amplitudes with more *negative* parietal ERPs for old compared to new distractors.

In contrasts, when participants were asked to recognise Pictures and ignore Words in Experiment 2, this reversal in amplitude for distractors between the FN400 and the parietal old/new effect was no longer present. Target Pictures were associated with very large FN400 and parietal old/new effects, but Word distractors did not elicit significant FN400 nor parietal old/new effects (although the latter was a numerical trend in the same direction as for targets).

Whole-head PLS results

In Experiment 1, the non-rotated Task-PLS analysis found significant effects of both distractor Picture old/new status ($p=0.011$, accounting for 30% of cross-block covariance) and target Word old/new status ($p<0.001$, 58% of cross-block covariance), but no interaction ($p=0.218$, 12% of cross-block covariance). The electrode salience to bootstrapped standard error ratios for significant contrasts are shown in Figure 3, where it can be seen that the whole-head exploratory PLS analysis revealed similar findings to the targeted analysis. In Experiment 1, target Old Words elicited reliably more positive ERPs than New Words, and this effect peaked between about 400-700ms onwards with a central and left parietal distribution. Distractor Old Pictures elicited reliably more positive ERPs than New Pictures across frontal and central sites between around 300-500ms, which was followed by a

sustained negativity (maximal ~500-1000ms post-stimulus) for Old compared to New pictures across parietal and occipital sites.

In Experiment 2, only the effect of target Picture old/new status was significant ($p<0.001$, 89% of cross-block covariance), and there was no effect of distractor Word old/new status ($p=0.255$, 8% of cross-block covariance), nor was the interaction significant ($p=0.961$, 3% of cross-block covariance). The electrode salience to bootstrapped standard error ratios for the significant Picture main effect contrast are shown in Figure 3, again showing similar effects as the targeted statistical analysis. Target Old Pictures elicited more positive ERPs than New Pictures, and this effect was highly reliable across central and parietal sites, peaking between around 300-700ms. Towards the end of the epoch, the distribution of this positivity had a right frontal distribution. No such right frontal Old>New effect was observed for either Words or Pictures in Experiment 1.
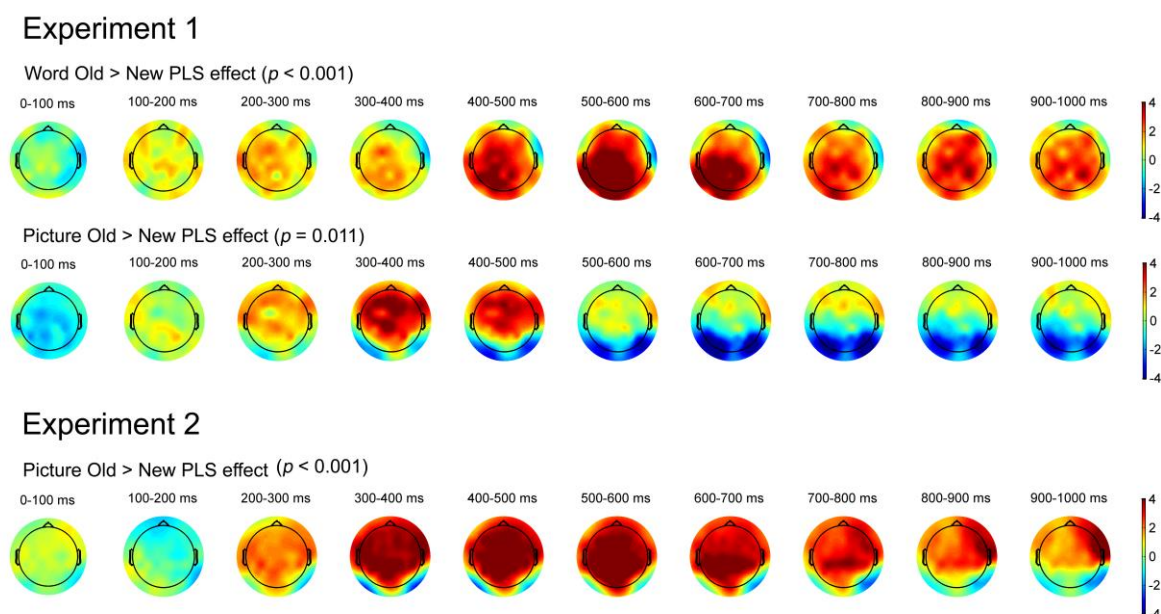
Experiment 1

Word Old > New PLS effect (p < 0.001)

Picture Old > New PLS effect (p = 0.011)

Experiment 2

Picture Old > New PLS effect (p < 0.001)

Figure 3. Topographic distributions of the electrode salience to bootstrapped standard error ratios for significant contrasts in the whole-head PLS analysis. The bootstrap ratios are approximately equivalent to z-scores; values >1.96 indicate electrodes and time-points that reliably show more positive ERP amplitude for old than new items, and values <−1.96 indicate electrodes and time-points that reliably show more negative ERP amplitude for old than new items.

Discussion

We investigated the neurocognitive underpinnings of distraction effects on recognition memory in order to determine whether intentional recognition of target stimuli and unintentional recognition of distracting stimuli in the same environment would be associated with similar or different underlying brain mechanisms. The results showed that unintentional recognition was associated with a distinct ERP old/new pattern that differed substantially from ERPs during intentional recognition. Unintentional distractor recognition was only associated with the FN400 ERP correlate of familiarity and not the left parietal ERP correlate

of recollection, as the latter was uniquely elicited by intentional target recognition. The results thus revealed a clear dissociation between these two well-established ERP markers of recognition processes (e.g. Rugg et al., 1998; Rugg & Curran, 2007; Wilding & Ranganath, 2012), in line with dual-process models that consider familiarity and recollection as functionally and neurally independent retrieval processes (Yonelinas & Jacoby, 2012). The findings are consistent with the view that unintentional recognition is driven by relatively automatic familiarity rather than recollection (Anderson et al., 2011), and with previous evidence that recollection can be voluntarily suppressed when it is unwanted (Bergström et al., 2007; 2009a; 2009b; 2013a; Hanslmayr et al., 2009; Hu et al., 2015; Mecklinger et al., 2009).

As predicted, unintentional recognition of distracting pictures had a biasing effect on the accuracy of participants' word recognition judgements, so that accuracy was highest when the memory status of the picture was congruent rather than incongruent with the word memory status. Word distractors however did not affect the accuracy of picture recognition judgements, in line with previous research that found less consistent effects of word distractors on intentional picture recognition, than vice versa (Anderson et al., 2011). Likewise, ERP evidence of distractor recognition was only found for pictures and not words, suggesting that the difference in bias between word and picture distractors was related to their likelihood of eliciting unintentional recognition (i.e. the actual recovery of stored memory information), rather than the extent to which people engaged post-retrieval processing to discount unintentional memory signals for words versus pictures (see Rugg & Wilding, 2000). Consistent with this account, previous research has found that irrelevant old contexts can sometimes elicit familiarity-related ERP FN400 effects whilst participants make recognition judgements about superimposed objects (Tsivilis, Otten, & Rugg, 2001), but that cueing participants to selectively attend to the objects reduces the context effect on ERPs (Ecker, Zimmer, Groh-Bordin, & Mecklinger, 2007). Our findings extend on this prior research by linking the FN400 effect for irrelevant distractors with biased recognition judgements to targets, providing evidence that distractor-induced behavioural biases are driven by familiarity-related brain processes.

Pictures may have been more likely to elicit unintentional recognition than words in our study because they were perceptually more salient and thus more likely to attract attention. However, other lines of research suggest that word processing is often highly automatic and can interfere greatly with participants' ability to make accurate judgements, for example in the classic Stroop task where word reading interferes with colour naming (Stroop, 1935). One possibility for this discrepancy may be that in our experiments, participants conducted a verbal working memory task that may have interfered more with their word processing than their picture processing (cf. Fernandes & Moscovitch, 2000), thus making unintentional recognition of word distractors less likely. However, previous research using a very similar paradigm to ours (Anderson, et al., 2011) also found larger, more consistent effects of picture than word distractors even without a simultaneous divided attention task, suggesting that the nature of our working memory task cannot be the sole reason for the difference between material types. Future research should clarify the factors that determine whether a stimulus elicits unintentional recognition, which likely includes perceptual salience as well as other factors such as memorability and distinctiveness (see Anderson et al., 2011; Ste-Marie & Jacoby, 1993; for further discussion). It is also important to assess whether unintentional recognition is sensitive to domain overlap with concurrent, task-related processing (Fernandes & Moscovitch, 2000).

When old pictures were intentionally recognised as targets, they showed a typical pattern with both early fronto-central and late parietal positive ERP amplitudes compared to new target pictures. However, when the same pictures were unintentionally recognised as distractors, they still elicited a large initial early positivity, but later ERPs across parietal and occipital electrode sites were reversed so that old distractor pictures showed more *negative* ERP amplitudes than new distractor pictures. Left parietal amplitudes in memory tests are typically positively correlated with the amount of information that is recollected (Vilberg, Moosavi, & Rugg, 2006; Wilding, 2000), but it is unlikely that new distractors would elicit more recollection than old distractors. Furthermore, the PLS analysis showed that the topography of the distractor old<new effect was more posterior than the target old>new effect. Instead, the enhanced negativity for old distractors likely corresponds to the Late Posterior Negativity (LPN) that originates in the Precuneus (Bergström, Henson, Taylor, Simons, 2013b). The LPN is thought to index evaluation of retrieved information or monitoring of responses (Johansson & Mecklinger, 2003). In the current paradigm, such retrieval monitoring processes may be recruited when old items trigger automatic familiarity in order to determine the source of the memory signal. Thus, early unintentional recognition may be followed by later, intentionally engaged monitoring that enables people to counteract the biasing influence of distracting recognition (c.f. Hu et al., 2015).

Together with previous research, our findings have implications for eyewitness memory tests where recognition of a suspect is tested in the context of multiple distractors that may elicit feelings of familiarity and thereby bias recognition responses to the suspect. For example, standard line-ups typically present the suspect simultaneously with other "filler" non-suspects (cf. Steblay, Dysart, & Wells, 2011), and some types of facial composite creation systems involve asking the eyewitness to recognise the face that most resembles the suspect amongst several alternatives (Frowd 2012; Solomon, Gibson, & Mist, 2013). Although our study did not investigate face recognition biases due to distractor familiarity, other studies have shown that presenting unfamiliar faces on a familiar background can cause people to falsely attribute memory signals from the background to the face (Deffler, Brown, & Marsh, 2014; Gruppuso, Lindsay, & Masson, 2007). Similar misattributions of familiarity may also occur between multiple faces that are simultaneously presented (Bower & Karlin, 1974), and such biases may be more likely in populations that are particularly susceptible to distraction, such as those with impaired attentional control (Engle, 2002), which likely includes older adults (Anderson et al., 2011; Campbell, Hasher, & Thomas, 2010; De Fockert, Ramchurn, Van Velzen, Bergström, & Bunce, 2009). Future research should determine the extent of recognition biases in eyewitness memory tests.

In conclusion, our findings show that unintentional and intentional recognition are dissociated by the well-established ERP correlates of familiarity and recollection. Unintentional recognition is driven by a rapid automatic familiarity process, while intentional recognition also involves a slower recollection process that is under voluntary control. These results are relevant to how recognition memory works in real world environments where we are surrounded with multiple stimuli that range in familiarity. Distraction-induced biases may render recognition memory in the real world less accurate than implied by typical experiments in the scientific literature.

References

Anderson, B. A., Jacoby, L. L., Thomas, R. C., & Balota, D. A. (2011). The effects of age and divided attention on spontaneous recognition. *Memory & Cognition, 39,* 725-735.

Bergström, Z. M., Velmans, M., De Fockert, J., & Richardson-Klavehn, A. (2007). ERP evidence for successful voluntary avoidance of conscious recollection. *Brain Research, 1151*, 19–133.

Bergström, Z. M., De Fockert, J., & Richardson-Klavehn, A. (2009a). Event-related potential evidence that automatic recollection can be voluntarily avoided. *Journal of Cognitive Neuroscience, 21,* 1280–1301.

Bergström, Z. M., De Fockert, J., & Richardson-Klavehn, A. (2009b). ERP and behavioural evidence for direct suppression of unwanted memories. *Neuroimage, 21,* 1280–1301.

Bergström, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013a). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological Psychology, 94,* 1-11.

Bergström, Z. M., Henson, R. N., Taylor, J. T., & Simons, J. S. (2013b). Multimodal imaging reveals the spatiotemporal dynamics of recollection. *Neuroimage, 68,* 141-153.

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. *Technical report C-1*, Gainesville, FL. The Center for Research in Psychophysiology, University of Florida.

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology, 103,* 751–757.

Campbell, K. L., Hasher, L., & Thomas, R. C. (2010). Hyperbinding: A unique age effect. *Psychological Science, 21,* 399–405.

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition, 28,* 923–938.

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behavioral Research Methods, 43,* 468-477.

Deffler, S. A., Brown, A. S., & Marsh, E. J. (2014). Judging the familiarity of strangers: Does the context matter? *Psychonomic Bulletin and Review, 22,* 1041-1047.

De Fockert, J.W, Rees, G., Frith, C., & Lavie, N. (2001). The Role of Working Memory in Visual Selective Attention. *Science, 291,* 1803-1806.

De Fockert, J. W., Ramchurn, A., Van Velzen, J., Bergström, Z. M., & Bunce, D. (2009). Behavioral and ERP evidence of greater distractor processing in old age. *Brain Research, 1282,* 67-73.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of singletrial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177.

Ecker, U. K. H., Zimmer, H. D., Groh-Bordin, C., & Mecklinger, A. (2007). Context effects on familiarity are familiarity effects of context—An electrophysiological study. *International Journal of Psychophysiology, 64,* 146–156.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19-23.

Frowd, C. D. (2012). Facial recall and computer composites. In C. Wilkinson & C. Rynn (Eds.), *Craniofacial identification.* Cambridge University Press. Cambridge, UK.

Fernandes, M.A., & Moscovitch, M. (2000). Divided attention and memory: Evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General, 129,* 155-76.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Science, 16,* 129–135.

Gruppuso, V., Lindsay, D. S., & Masson, M. E. (2007). I'd know that face anywhere! *Psychonomic Bulletin and Review, 14,* 1085–1089.

Hanslmayr, S., Leipold, P., Pastötter, B., & Bäuml, K.-H. (2009). Anticipatory signatures of voluntary memory suppression. *Journal of Neuroscience, 29,* 2742–2747.

Hayama, H. R., Johnson, J. D., & Rugg, M. D. (2008). The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: specific or non-specific? *Neuropsychologia, 46,* 1211–1223.

Hockley, W. E., Bancroft, T. D., & Bryant, E. (2012) Associative and familiarity-based effects of environmental context on memory. *Canadian Journal of Experimental Psychology, 66,* 81–89.

Hu, X., Bergström, Z. M., Bodenhausen, G.V., & Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences: Evidence from electrophysiology and an implicit autobiographical memory test. *Psychological Science, 26,* 1098-1106.

Johansson, M., & Mecklinger, A. (2003). The late posterior negativity in ERP studies of episodic memory: action monitoring and retrieval of attribute conjunctions. *Biological Psychology, 64,* 91–117.

Lang, P. J., Bradley, M. M., & Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-8.* University of Florida, Gainesville, FL.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252 –271.

McIntosh, A. R., & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: applications and advances. *Neuroimage, 23(Suppl. 1),* S250–S263.

Mecklinger, A., Parra, M., & Waldhauser, G. T. (2009). ERP correlates of intentional forgetting. *Brain Research, 1255,* 132–147.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General, 128,* 403–415.

Paller, K. A., Voss, J. L., & Boehm, S. G. (2007). Validating neural correlates of familiarity. *Trends in Cognitive Science, 11,* 243–250.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Science, 11,* 251–257.

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allen, K (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature, 392,* 595–598.

Rugg, M. D., Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Science, 4*, 108–115.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34.

Solomon, C. J., Gibson, S. J., & Mist, J. J. (2013). Interactive evolutionary generation of facial composites for locating suspects in criminal investigations. *Applied Soft Computing, 13,* 3298-3306.

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: a meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17,* 99–139.

Ste-Marie, D. M., & Jacoby, L. L. (1993). Spontaneous versus directed recognition: The relativity of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 777–788.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Tsivilis, D., Otten, L. J., & Rugg, M. D. (2001). Context effects on the neural correlates of recognition memory: an electrophysiological study. *Neuron, 31,* 497–505.

Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research, 1122,* 161–170.

Wilding, E. L. (2000). In what way does the parietal ERP old/new effect index recollection? *International Journal of Psychophysiology, 35,* 81–87.

Wilding, E. L., & Ranganath, C. (2012). Electrophysiological correlates of episodic memory processes. In: *The Oxford Handbook of Event-Related Potential Components* (S. Luck and E. Kappenman, eds.), pp. 373-396. Oxford: Oxford University Press.

Yonelinas, A. P., Jacoby, L. J. (2012). The process-dissociation approach two decades later: Convergence, boundary conditions, and new directions. *Memory & Cognition, 40,* 663-680.