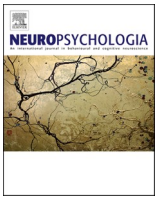
Neuropsychologia 148 (2020) 107659

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia



Usefulness of familiarity signals during recognition depends on test format: Neurocognitive evidence for a core assumption of the CLS framework

Regine Bader [a,*], Axel Mecklinger [a], Patric Meyer [b]

[a] Experimental Neuropsychology Unit, Department of Psychology, Saarland University, Saarbrücken, Germany
[b] Department of Psychology, SRH University of Applied Sciences, Heidelberg, Germany

A B S T R A C T

Familiarity-based discrimination between studied items and similar foils in yes/no recognition memory tests is relatively poor. The complementary learning systems (CLS) framework explains this with the small difference in familiarity strength between targets and foils. The framework, however, also predicts that familiarity values of targets and corresponding similar foils are directly comparable – as long as they are presented side by side in a forced-choice corresponding (FCC) test. This is because in each trial, targets tend to be more familiar than their corresponding foils. In contrast, when forced-choice displays contain non-corresponding foils (FCNC) which are similar to other studied items, familiarity values are not directly comparable (as in yes/no-tasks). In a recognition memory task with pictures of objects, we found that the putative ERP correlate of familiarity, the mid-frontal old/new effect for targets vs. foils, was significantly larger in FCC compared to FCNC displays. Moreover, single-trial target-foil amplitude differences predicted the accuracy of the recognition judgment. This study supports the assumption of the CLS framework that the test format can influence the diagnostic reliability of familiarity. Moreover, it implies that the mid-frontal old/new effect does not reflect the difference in the familiarity signal between studied and non-studied items but the task-adequate assessment of this signal.

1. Introduction

It is well established that recognition memory is generally supported by two distinct processes. While familiarity is a mere feeling of having something encountered before, recollection involves memory for details of a prior encounter (Yonelinas, 2002). Generally, the capability to recognize events on the basis of familiarity is impressive. However, if studied and non-studied items are too similar, this ability can break down in standard yes/no testing situations where items are presented one by one (Morcom, 2015). A computational explanation for this is given by the complementary learning systems (CLS) framework (Norman & O'Reilly, 2003): Recollection is assumed to rely on the integrity of the hippocampus, which assigns pattern-separated (i.e. non-overlapping) representations to each single episode, even when events are similar. In contrast, familiarity signals are assumed to be created by the medial temporal lobe cortex, which assigns overlapping representations to similar events. Thus, although studied items are more familiar than non-studied similar foils, the difference in familiarity strength between these two item classes is on average relatively small and their familiarity strength distributions are highly overlapping (see Fig. 1). This renders familiarity-based recognition unreliable (Migo et al., 2009) and usually leads to high error rates in standard yes/no (YN) tasks, where test items are presented one at a time and a global decision criterion across all test trials can be assumed. However, the CLS predicts better performance when studied targets and corresponding similar foils are presented together on a forced-choice (FC) test display. In those cases, the familiarity values of these two items can be directly compared which permits trial-unique decision criteria. Although the overall difference between the familiarity distributions does not change for FC tests, the high co-variation of the familiarity values for studied items and similar lures renders the small within-trial differences in familiarity reliable.

To gain support for this CLS assumption, Holdstock et al. (2002) tested patient Y.R. who had a selective hippocampal lesion, which impaired recollection but spared familiarity. As predicted, using a picture recognition test involving similar foils, the patient performed within the range of healthy controls when tested in a FC but not in a YN test (Holdstock et al., 2002). While other studies showed no benefit from FC tests for hippocampal patients (Bayley et al., 2008; Jeneson et al., 2010), one study with older adults, for whom a disproportional deficit in recollection is assumed, showed an increase in familiarity-based responses in a FC compared to a YN test when recognition memory for similar faces was tested (Bastin and van der Linden, 2003). Moreover, a recent study (Migo et al., 2009) that investigated test format effects on familiarity-based recognition in healthy younger individuals contrasted three conditions: YN, forced-choice with targets next to corresponding

* Corresponding author.
E-mail address: regine.bader@mx.uni-saarland.de (R. Bader).

https://doi.org/10.1016/j.neuropsychologia.2020.107659
Received 2 April 2020; Received in revised form 24 September 2020; Accepted 11 October 2020
Available online 16 October 2020
0028-3932/

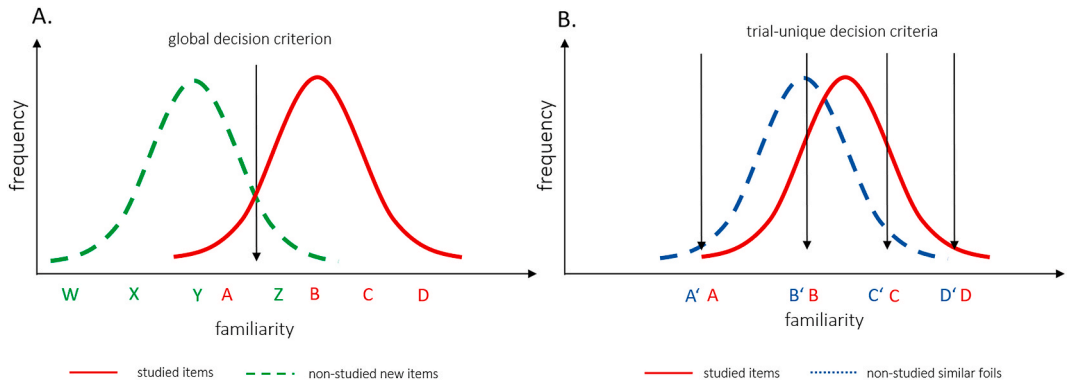


Fig. 1. Familiarity distributions of studied and non-studied items during test assuming equal variances. A. When studied items (A,B,C,D) and non-studied new items (W,X,Y,Z) are dissimilar, familiarity distributions are only partly overlapping and a global decision criterion (as assumed in yes-no-tests) can be used. B. When studied items (A,B,C,D) and non-studied foils (A′,B′,C′,D′) are similar, familiarity differences due to study exposure are smaller than overall variance leading to strongly overlapping distributions and thus only the use of trial-specific decision criteria as in forced-choice corresponding tests is useful.

similar foils (FCC), and forced-choice with targets next to foils which were similar to other studied items (forced-choice non-corresponding, FCNC). The FCNC condition served as control condition because it was comparable with FCC tests (Fig. 2A), but did not allow for direct comparison of familiarity strength values. Supporting the view that familiarity is more useful in the FCC condition, instructions to use only familiarity reduced performance compared to standard instructions in the FCNC and YN conditions, but not in the FCC condition.

As neuropsychological and behavioral studies have previously produced mixed results, it becomes clear that neurocognitive evidence in healthy young participants is essential, but still missing. Here, we explored the effects of test format on familiarity and recollection using event-related potentials (ERPs). Typically, ERPs to old items are more positive-going than those to new items. Familiarity has been associated with the mid-frontal old/new effect which is most-pronounced between 300 and 500 ms post-stimulus (but see Paller et al., 2007, for an alternative view) but can also be temporally extended (Rugg et al., 1998; Tsivilis et al., 2001). Recollection on the other hand has been linked to the later occurring (500–800 ms) late parietal old/new effect (Rugg and Curran, 2007). While the former varies with familiarity strength (Woodruff et al., 2006; Yu and Rugg, 2010) and is not affected by speeded response requirements (Mecklinger et al., 2010), the latter is sensitive to the amount of remembered study details (Vilberg et al., 2006; Wilding and Rugg, 1996). As the two effects have also been doubly dissociated in a variety of studies (e.g., Curran and Doyle, 2011; Jäger et al., 2006; Opitz and Cornell, 2006), they can be regarded as reliable neural correlates of familiarity and recollection and can be employed as independent estimates of these processes.

Here, we administered two study-test-cycles, one with FCC test displays and one with FCNC test displays. In the intentional study phases, participants were required to complete a size judgment task (Is the depicted object smaller or larger than a shoebox?) for black and white pictures. As ERPs in the test phase were recorded separately for targets and foils, the stimuli had to be presented sequentially. This variant of the forced-choice display has also been used in previous ERP studies (Rosburg et al., 2011; Schwikert and Curran, 2014; Voss and Paller, 2009) in which ERP correlates of familiarity and recollection have been observed. Importantly, we are confident that this mandatory change in the procedure compared to behavioral paradigms does not affect the assumptions regarding the underlying processes. Critically, these assumptions do not rely on the simultaneous presentation of the two pictures but generally on the kind of comparison that has to be made in each trial.

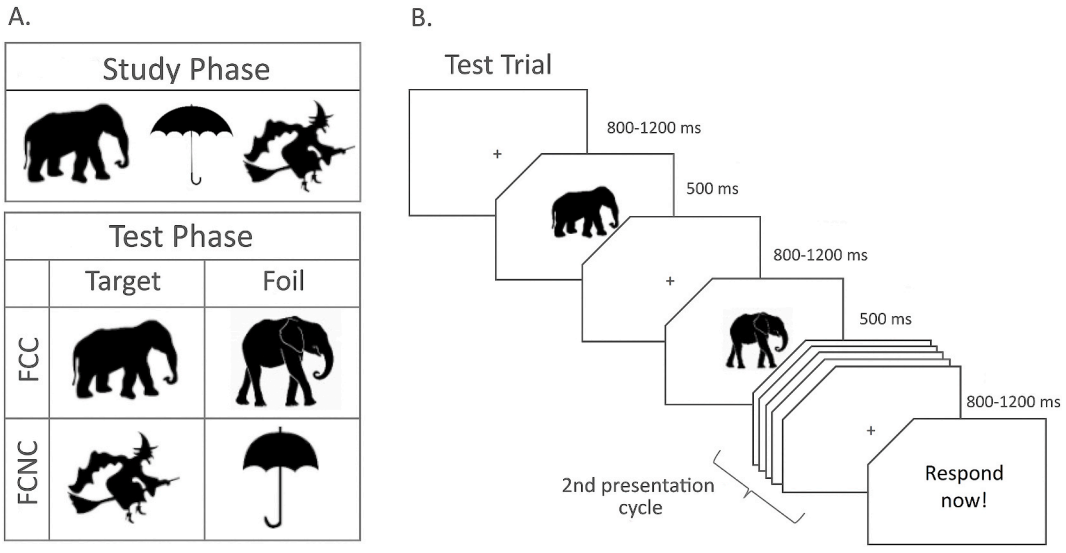


Fig. 2. Experimental materials and trial procedure. A. Example target-foil combinations in forced-choice corresponding (FCC) and forced-choice non-corresponding (FCNC) conditions. B. Schematic illustration of a test trial in the FCC condition. Presentation was identical in the FCNC condition except for the target-foil combination. The first four displays of the trial were repeated in the 2nd presentation cycle. Targets were at the 1st position in half of the trials for each participant. Black and white images are taken from the internet and are reprinted under the creative commons license CC0 1.0.

This comparison should be the same for parallel and sequential presentation. In standard FC procedures, participants presumably switch back and forth between the two simultaneously presented pictures. Thus, in order to enhance the comparability between the sequential and the simultaneous presentation format and to also enable back and forth switching between pictures, we adapted the procedure used by Voss and colleagues (Voss and Paller, 2009). The authors also examined ERP old/new effects in a forced-choice recognition test and repeated the target-foil sequence within a trial which allowed participants to look at each picture twice (Fig. 2B). Apart from enhancing the ecological validity of the procedure, repeating stimuli also allowed for a valid analysis of the condition differences. This is because these differences manifest themselves only once both pictures have been presented as subjects can make the critical comparison only after they have also seen the second stimulus of each picture pair. Data from our pilot study suggested that the changes in the ERPs reflecting differences in visual processing (repeating a similar picture vs. presenting two different pictures in succession) are largest in the second picture of the first cycle. This holds a relatively high risk of overshadowing any more subtle differences in the second picture of the first cycle. Therefore, our analyses focused on the second presentation of the two pictures. In line with the CLS framework, we predicted that the mid-frontal old/new effect would be larger in the FCC than FCNC condition. In a second step, we tested the CLS assumption that within-trial differences in the FCC condition can be reliably used to guide recognition judgments. For this purpose, we used a logistic regression approach to assess whether the amplitude difference between the target and the foil for each single trial can predict the accuracy of a subject's response (see Noh et al., 2018; Ratcliff et al., 2016, for other studies using single-trial approaches).

2. Materials & methods

2.1. Participants

Thirty-two students from Saarland University (16 female, mean age of 24.4 [19–35] years) participated in the experiment. Two additional subjects had to be excluded because they did not perform above chance (p[hits] > 0.5) in the recognition test as revealed by a binomial test ($p > .05$) or could not contribute enough artefact-free trials (at least 13 per condition). All participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision and no known neurological problems (self--report). They gave informed consent and were reimbursed with eight Euros/hour or course credit. The study was approved by the ethics committee of the Faculty of Human and Business Sciences at Saarland University and adhered to the Declaration of Helsinki.

2.2. Stimuli & procedure

Visual stimuli were 352 pairs of black and white images (silhouettes or icons) collected from the internet. Each pair consisted of two very similar versions of one object. The experiment was divided into two study-test blocks, one for each condition. The order of the blocks was counterbalanced. In each study phase, participants were presented with 176 single images. In the FCNC condition, 88 of these images were presented as targets in the test phase together with a foil which was similar to one of the remaining 88 images form the study phase. Thus, in the FCNC condition, only one version of each pair appeared during the test as judgments on one image of a similar pair could influence subsequent judgments on the other one. In the FCC condition, 88 images were presented in the test phase together with the corresponding similar foil. The remaining 88 images in the study phase served as filler items to equalize block length in both conditions because in the FCNC condition 176 study pairs were needed to obtain 88 test trials (see Fig. 2A).

In the beginning of the experiment, participants were told that there would be several study-test blocks. Before the first block, there was a short study-test practice block, which included both, FCC and FCNC test trials, presented in the same order as the experimental blocks, to familiarize the participants with the complete procedure and to make the testing procedures in both blocks as equal as possible. Moreover, participants were explicitly made aware of the two types of test displays. In order to control for differential encoding strategies, subjects were told whether foils were similar to the target or to another previously studied image only before the test phase and not at the beginning of each block. As subjects were told that they would see several rather than specifically 'two blocks', they did not expect to see the other condition in the second block.

During the study phase, participants had the task to memorize the images as accurately as possible and to decide whether the depicted object was smaller or larger than a shoebox. They were not explicitly told which test format would follow. A study trial started with a 500 ms fixation cross before the image, which was presented for 3000 ms. Afterwards a question mark prompted participants to make the shoebox decision for which they had a maximum of 1500 ms. After a 500 ms blank screen the next trial started. After every 22 trials, participants could make a self-paced short break.

Each test trial comprised two presentation sequences of target and foil. To avoid EEG oscillations time-locked to stimulus presentation, each sequence started with a jittered fixation cross (800–1200 ms) followed by the presentation of the first image for 500 ms. After another jittered fixation cross (800–1200 ms), the second image was shown for 500 ms. For each participant, the target was the first image within this sequence for half of the trials and the second image for the other half. After the second presentation of the sequence, a jittered fixation cross (800–1200 ms) appeared followed by a prompt ("Jetzt antworten!"/ "Respond now!") to indicate whether the target was the first or the second image within the sequence. Participants had a maximum of 1000 ms to respond. After a 1000 ms blank screen the next trial began. After every 22 trials, participants could make a self-paced break.

2.3. EEG data acquisition & processing

BrainVision Recorder 1.0 (Brain Products) was used to record EEG continuously from 59 scalp sites according to the extended 10–20 system (Jasper, 1958). The EEG was amplified with electrode AFz as ground electrode and the left mastoid electrode as reference using a 16-bit BrainAmp Amplifier (Brain Products). Data were digitized using a sampling rate of 500 Hz and an on-line analog band-pass filter of 0.016–250 Hz. Data were stored using an on-line digital low-pass filter of 100 Hz. Impedances were kept below 5 kΩ. Four additional electrodes were placed on the outer canthi and above and below the right eye to record electrooculographic (EOG) activity. BrainVision Analyzer 2.1 (Brain Products) was used for offline data processing which started with visually discarding excessive artifacts to improve independent component analysis (ICA) performance which was used for the correction of EOG and cardiac artifacts. The EEG was first filtered using a 0.05–30 Hz Butterworth filter (order: 4) and the ICA with a classic restricted infomax algorithm was employed. After re-referencing to the average of both mastoid electrodes, the data were segmented into −200 to 1000 ms epochs relative to image onset for each of the four image presentations within a trial. The epochs were baseline-corrected and artifacts were rejected by identifying segments including voltage steps greater than 30 μV/ms, voltage differences greater than 100 μV within a 200 ms interval or greater absolute amplitudes than ± 70 μV. Finally, the data were checked manually for remaining artifacts (especially excessive alpha waves). For graphical illustration, waveforms were exported and we used the ggplot2 package of the software R (Wickham, 2009) to plot the ERP waveforms. Brain Vision Analyzer 2.1. (Brain Products) was used to create topographic maps.

2.4. Experimental design and statistical analysis

Inferential statistics were conducted using the software R (R Core Team, 2017) in RStudio (RStudio Team, 2016), especially the packages tidyverse (Wickham et al., 2019), lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017) and ez (Lawrence, 2016). Significance level was set to $\alpha = 0.05$. Behavioral data were analyzed with ANOVAs with the between-subjects factor condition order (FCC first, FCNC first) and the within-subjects factor condition (FCC, FCNC). For the ERP data, only data from the second presentation cycle were entered in the analyses as conditions do not differ in the first picture of the first cycle and overall differences in visual processing (presenting a similar picture vs. a dissimilar picture) might overshadow differences in old/new effects in the second picture. In the second cycle, all pictures are repetitions and therefore these overall differences are reduced rendering the comparison between conditions more specific to differences in memory processes. Based on previous literature (e.g., Küper et al., 2012), mean amplitudes from 300 to 500 ms were extracted for the mid-frontal old/new effect and from 500 to 650 ms for the late parietal old/new effect. The latter time window was slightly shorter compared to other ERP recognition memory studies because of the offset potential being evident in the ERP around 700 ms. Amplitudes were pooled to be analyzed in a fronto-central (F1, Fz, F2, FC1, FCz, FC2) and a centro-parietal cluster (CP1, CPz, CP2, P1, Pz, P2). To examine the effect of test format on the old/new effects, only ERPs from correctly answered trials were used. Mean trial numbers (range) were 35.3 (26–43) for FCC hits in the first position, 34.1 (20–43) for FCC foils in the first position, 34.1 (20–43) for FCC hits in the second position, 34.84 (24–43) for FCC foils in the second position, 30.75 (12–41) for FCNC hits in the first position, 32.41 (17–42) for FCNC foils in the first position, 31.7 (17–43) for FCNC hits in the second position, and 31.4 (14–41) for FCNC foils in the second position. The $2 \times 2 \times 2 \times 2$ repeated-measures ANOVAs included the between-subjects factor condition order (FCC first, FCNC first) and the within-subjects factors condition (FCC, FCNC), item type (target, foil), and picture position (first, second). Only effects involving the factors of condition or item type are reported. Significant interactions were followed-up with *t* tests, for which in case of unplanned comparisons *p*-values were adjusted according to the Bonferroni-Holm procedure (Holm, 1979). In case we tested directional hypotheses using tests that allowed one-tailed testing (i.e. whenever we directly tested targets vs. foils in the FCC condition), we report the *p* values of the one-tailed test. Partial eta square (η_p^2) and Cohen's d_{av} with the average of the two standard deviations as the denominator are provided as measures of effect size.

In order to test the hypothesis that accuracy of a participant's response in a single trial can be predicted based on the ERP signature of familiarity strength in a single trial, we used multi-level binary logistic regression instead of standard binary logistic regression analyses to

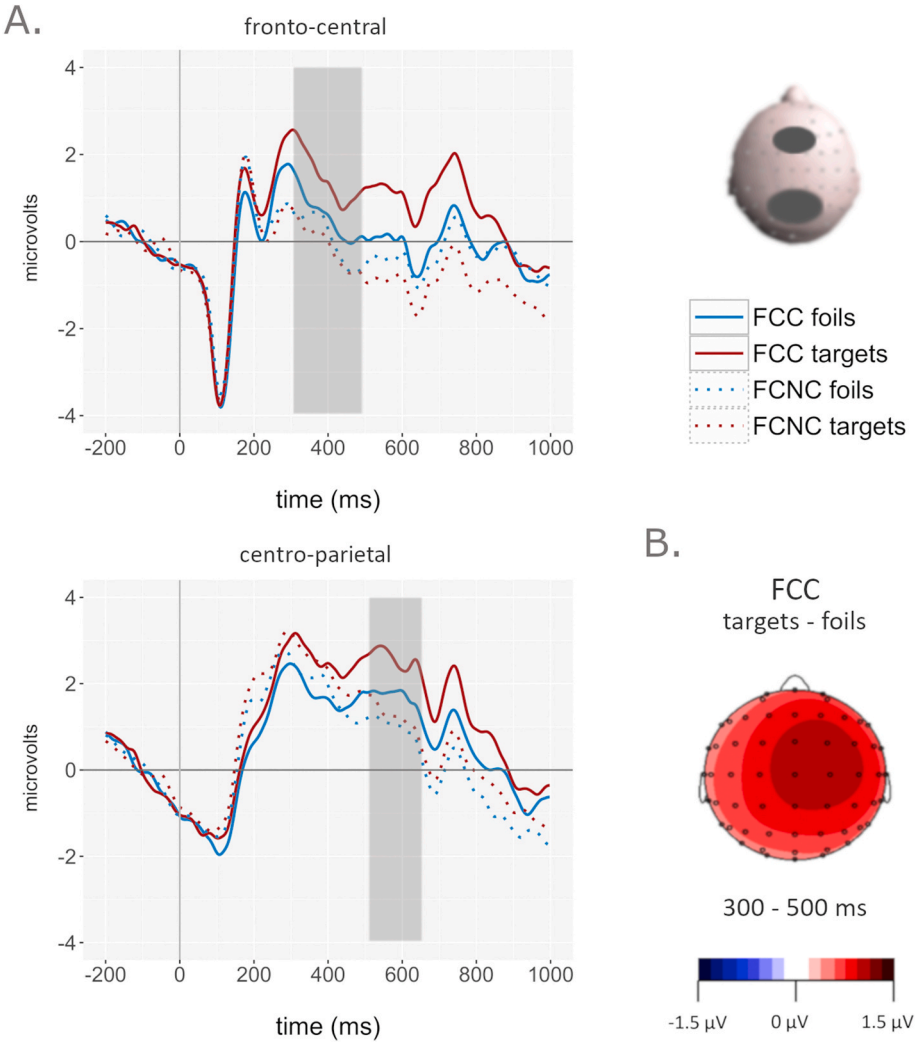


Fig. 3. ERP results for the second presentation cycle. A. ERP waveforms at the fronto-central and the centro-parietal electrode cluster for all four conditions. Shaded areas indicate analyses time windows. B. Topographic distribution of the target vs. foil difference in the FCC condition for the early time window.

account for dependencies in the data within subjects. For each trial, difference scores between target and foil amplitudes in the fronto-central ROI were calculated for the early time window (see Rosburg et al., 2011, for a similar approach). These difference scores were calculated for the second cycle. In addition to reducing differences in overall visual processing, this was done as targets and foils should be taken from the same cycle and the difference between test format does not take effect at the first picture of the first cycle. These difference scores were entered into two different multi-level models per condition. For one model, we included the target-foil difference scores as a predictor and permitted random intercepts across subjects (random intercepts only model). In the other model, we allowed also the predictor to vary across subjects (random intercepts/random slopes model). We then compared whether the random intercepts/random slopes model reliably improved the fit of the data as compared to the random intercepts only model based on a χ^2 test of the change in −2 log likelihood. In case of no improvement, we kept the simpler model. Significance of single effects was assessed based on the significance test for the predictor (z statistic). To test whether the target-foil difference score better predicted the accuracy of a response in the FCC than the FCNC condition, we used the data of both conditions and included the difference score, condition and the interaction term of Condition x Difference Score in the model. Before this, we centered the difference score variable within subjects and used centered values for condition (−1 = FCC, 1 = FCNC).

3. Results

3.1. Behavioral results

The 2 × 2 ANOVA with the between-subjects factor condition order and the within-subjects factor condition on percent correct revealed that the main effect of condition order was not significant, $F(1,30) = 0.039$, $p = .845$, $MSE = 0.023$, $\eta_p^2 = 0.001$, and that overall accuracy was significantly higher in the FCC condition ($M = .82$, $SD = .09$) than in the FCNC condition ($M = .75$, $SD = .13$), $F(1,30) = 18.19$, $p < .001$, $MSE = 0.004$, $\eta_p^2 = 0.377$. The interaction was not significant, $F(1,30) = 0.004$, $p = .950$, $MSE = 0.004$, $\eta_p^2 < 0.001$. An ANOVA with the same factors on mean reaction times yielded no significant main effect of condition order, $F(1,30) = 0.11$, $p = .747$, $MSE = 9590$, $\eta_p^2 = 0.004$, but a marginally significant main effect of condition, $F(1,30) = 4.08$, $p = .052$, $MSE = 1257$, $\eta_p^2 = 0.120$, which was qualified by a significant interaction of the two factors, $F(1,30) = 9.56$, $p = .004$, $MSE = 1257$, $\eta_p^2 = 0.242$. This interaction was due to significantly shorter reaction times in the FCC condition ($M = 305$, $SD = 71$) than in the FCNC condition ($M = 351$, $SD = 85$) when the FCNC condition was administered first, $t(15) = 3.38$, $p = .004$, $d_{av} = 0.58$. When the FCC condition was administered first, reaction times did not differ significantly between the FCC condition ($M = 341$, $SD = 64$) and the FCNC condition ($M = 331$, $SD = 72$), $t(31) = 0.82$, $p = .426$, $d_{av} = 0.14$.

3.2. ERP results

Fig. 3 shows ERP waveforms collapsed across picture positions. As can be seen, ERPs in the early time window at frontal electrodes are generally more positive in the FCC condition than in the FCNC condition. Moreover, targets elicit more positive-going waveforms than foil items in the FCC condition while no such difference is observable for the FCNC condition. As can be seen in Fig. 4B, this is especially evident during the first picture position of the second cycle. A similar pattern in the waveforms is observable for the late time window and parietal recording sites.

3.3. Time window 300–500 ms

A 2 × 2 × 2 × 2 ANOVA with the between-subjects factor condition order and the within-subjects factors condition, item type, and picture position was run on mean amplitudes measured over the fronto-central electrode cluster during the second presentation cycle. As there were no interactions ($ps > .23$) with the factor condition order, we report here the results of the 3-way-ANOVA without this factor. This analysis yielded a significant main effect of condition, $F(1,31) = 14.45$, $p = .001$, $MSE = 4.69$, $\eta_p^2 = 0.32$, a significant main effect of item type, $F(1,31) = 6.22$, $p = .018$, $MSE = 1.12$, $\eta_p^2 = 0.17$, a significant Condition × Picture Position interaction, $F(1,31) = 16.74$, $p < .001$, $MSE = 3.03$, $\eta_p^2 = 0.35$, a marginally significant Item Type × Picture Position interaction, $F(1,31) = 4.09$, $p = .052$, $MSE = 2.35$, $\eta_p^2 = 0.12$, and as predicted a significant Condition × Item Type interaction, $F(1,31) = 5.86$, $p = .022$, $MSE = 3.27$, $\eta_p^2 = 0.16$. The latter interactions were not qualified by a Condition × Item Type × Picture Position interaction, $F(1,31) = 1.27$, $p = .268$, $MSE = 2.63$, $\eta_p^2 = 0.04$. The main effect of picture position, $F(1,31) = 0.002$, $p = .967$, $MSE = 2.83$, $\eta_p^2 < 0.001$, was also not significant.

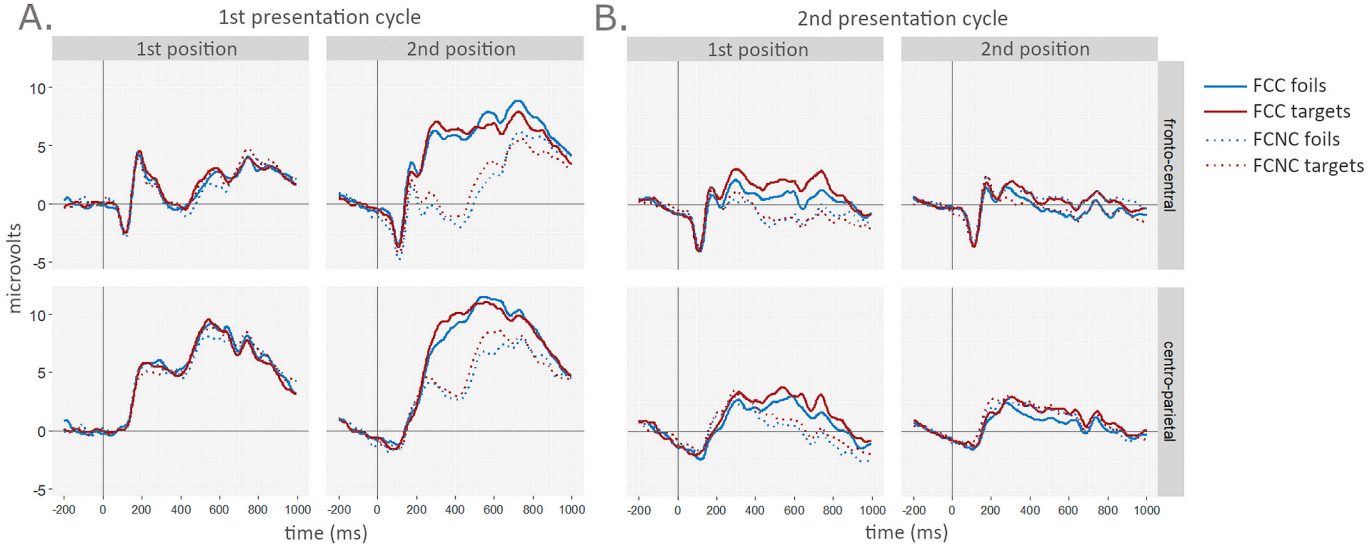


Fig. 4. ERP waveforms at fronto-central (upper panel) and centro-parietal (lower panel) electrode clusters for targets and foils in both conditions. A. First presentation cycle, separately for each picture position. B. Second presentation cycle, separately for each picture position.

To follow-up the significant Condition × Item Type interaction, amplitudes at fronto-central sites to target and foils were compared for each condition separately, collapsed across both picture positions. As planned comparisons revealed, in the FCC condition, targets elicited significantly more positive-going waveforms than foils, $t(31) = 3.24$, $p = .001$, $d_{av} = 0.27$, one-tailed, while the difference was reversed, but not significant in the FCNC condition, $t(31) = 0.86$, $p = .398$, $d_{av} = 0.08$. To follow-up the significant Condition × Picture Position interaction, we compared post hoc the two conditions for each picture position separately, collapsed across item types. For the first position, mean amplitudes in the FCC condition were significantly larger than in the FCNC condition, $t(31) = 5.78$, $p < .001$, $d_{av} = 0.66$. In contrast, there was no significant difference between conditions for the second picture position, $t(31) = 0.39$, $p = .702$, $d_{av} = 0.04$. To sum up, in the early time window in the fronto-central ROI, analyses of the second cycle revealed an old/new effect in the FCC condition, but not in the FCNC condition. Moreover, across both item types, there was a significant condition difference in the first but not the second position.

3.4. Time window 500–650 ms

The 2 × 2 × 2 × 2 ANOVA with the between-subjects factor condition order and the within-subjects factors condition, item type, and picture position on the centro-parietal electrode cluster for the second cycle again revealed no interaction involving condition order ($ps > .09$). Therefore, we report the ANOVA dropping the between-subjects factor. There were significant main effects of condition, $F(1,31) = 16.25$, $p < .001$, $MSE = 4.00$, $\eta_p^2 = 0.34$, and item type, $F(1,31) = 5.28$, $p = .028$, $MSE = 3.47$, $\eta_p^2 = 0.15$, as well as a significant interaction of Condition × Picture Position, $F(1,31) = 36.25$, $p < .001$, $MSE = 2.80$, $\eta_p^2 = 0.54$. However, all other interaction effects were not significant ($ps \geq .200$).

Due to the significant Condition × Picture Position interaction, we post hoc compared conditions for each picture position separately, collapsed across item types. For the first position, waveforms in the FCC condition were significantly more positive-going than in the FCNC condition, $t(31) = 6.64$, $p < .001$, $d_{av} = 0.77$, whereas there was no difference for the second position, $t(31) = 0.81$, $p = .423$, $d_{av} = 0.10$. To sum up, in the late time window a reliable old/new effect was evident across both conditions. Moreover, a condition difference was observed for the first but not for the second position.

3.5. Control analysis

As anticipated and visible in Fig. 4, the mere difference in the visual experience within the first presentation cycle between conditions (repetition of similar pictures vs. presenting two different pictures) led to large differences in the ERPs between conditions during the first cycle which is why we focused our analyses on the second cycle. However, we ran a control analysis to be assured that the condition differences in the first cycle did not influence the differential old/new effects in the second cycle. If this was the case, we would expect that the size of the mean condition difference in the second picture of the first cycle (FCC–FCNC) is closely related to the mean target-foil difference for the second cycle in the FCC condition. To test this hypothesis, we normalized amplitudes in a way that each difference score represents the effect size for each subject, controlling for overall differences in amplitude size and variability, and calculated the across-subject correlation between these difference scores which was not significant, $r = 0.11$, $p = .543$.

Table 1

	FCC	FCNC	Both conditions
Intercept	1.72 (0.13)	1.33 (0.14)	1.73 (0.13)
Random effects	0.42	0.55	0.48
Difference score	.009 (.005)	.004 (.004)	0.009 (0.005)
Condition			−0.42 (0.07)
Difference score * condition			−0.005 (0.006)

Coefficients (standard error) of the final random intercepts models for the early time window.

3.6. Multi-level logistic regression model

The final models using the mid-frontal old/new effect as predictor are summarized in Tab. 1. First, we analyzed whether the single-trial target-foil difference score predicts the accuracy of a response separately for both conditions. In the FCC condition, we compared the random intercepts/random slopes model with the random intercepts only model and found that it was not significantly better in predicting response accuracy, $\chi^2(2) = 0.12$, $p = .942$. Thus, we kept the random intercepts only model, in which the individual predictor difference score was significant, $z = 1.87$, $p = .031$, one-tailed. In the FCNC condition, the random intercepts/random slopes model was not better than the random intercepts only model, $\chi^2(2) = 1.63$, $p = .442$. In contrast to the FCC condition, the predictor difference score was not significant in the latter model, $z = 0.94$, $p = .345$, two-tailed. Thus, the target-foil amplitude difference successfully predicts the accuracy of a response only in the FCC condition, but not in the FCNC condition.

Second, we tested whether the target-foil difference score better predicted the accuracy of a response in the FCC than the FCNC condition. For this purpose, the centered difference score, centered values of condition and the interaction term of these variables were entered into the model. The random slopes/random intercepts model did not fit the data better than the random intercepts only model, $\chi^2(9) = 12.01$, $p = .213$. Thus, we kept the random intercepts only model, however, the interaction term was not significant, $z = -0.81$, $p = .421$, suggesting that prediction in the FCC condition was not significantly better than in the FCNC condition.

4. Discussion

This study set out to provide the (to our knowledge) first neurocognitive evidence in healthy subjects for one core prediction of the complementary learning systems framework (CLS) (Norman & O'Reilly, 2003). The framework assumes that familiarity has better diagnostic reliability in forced-choice corresponding (FCC) tests, in which the familiarity strength of two items can be directly compared, than in other test formats where no direct comparison is possible. As predicted, we showed that the mid-frontal old/new effect, the putative electrophysiological correlate of familiarity-based recognition memory (Rugg and Curran, 2007), is larger in an FCC test format than in a forced-choice non-corresponding test format (FCNC). The difference is that in FCC formats, the targets and similar foils are presented within the same trial whereas in FCNC formats the targets are presented together with foils which are similar to other targets from the study phase. According to the CLS framework, the medial temporal lobe cortex (MTLC) generates familiarity signals by assigning highly overlapping representations to similar inputs which results in small differences in familiarity strength between studied targets and similar foils. As these familiarity values are highly correlated, the direct comparison in FCC tests allows reliable recognition judgments even when only small differences in familiarity strength are accessible. In contrast, in FCNC tests, which are similar to standard yes/no formats, familiarity strength values must be compared to a global decision criterion which is problematic when familiarity distributions of targets and foils strongly overlap.

Our study is in line with other studies that show an increase in the accuracy of familiarity-based judgments for FCC tests (Bastin and van der Linden, 2003; Holdstock et al., 2002; Migo et al., 2009). Notably, we are aware of only one further study that investigated test format effects in healthy young participants (Migo et al., 2009) in which participants were instructed to exclusively rely on familiarity. Such a procedure poses high metacognitive demands on the subjects' insights into the

nature of familiarity and recollection and on their ability to suppress recollection. Thus, our study is the only study which dissociated the FCC and FCNC test formats without relying on subjects' meta-memory abilities. Moreover, since other existing studies with patients have revealed contrary results (Bayley et al., 2008; Jeneson et al., 2010), evidence from healthy participants is especially important.

Applying a logistic regression model, we also showed that for the 300–500 ms time window, the target-foil difference wave at fronto-central electrodes in a single trial was related to the accuracy of the response in this trial. Our single-trial analyses revealed that this was only the case for the FCC condition, i.e. when familiarity is supposed to be highly useful for recognition decisions. However, we did not find a significant condition by difference score interaction and so it remains unclear how specific the mid-frontal old/new effect's predictive value really is for the FCC condition. Thus, these results should be taken as preliminary and are meant to stimulate further research rather than delivering conclusive answers.

From a methodological perspective, analyzing old/new difference scores on a single-trial basis opens plenty of possibilities for new research questions. More specifically, this single-trial perspective can provide insights into the importance of a given neural signature for the outcome of a trial (such as the nature or speed of the response) over and above the general presence of this signature when averaged across all subjects (see Ratcliff et al., 2016, for a similar argument). In the current study, the single-trial analysis speaks to an important aspect of the CLS framework assumption as the model does not only assume an *overall* greater usefulness of familiarity for the FCC condition, but also states that *within-trial differences* in the FCC condition can be reliably used to guide recognition judgments. This claim is supported by the significant relationship between the within-trial difference in the mid-frontal old/new effect and the subject's response. A study by Ratcliff et al. (2016) used a similar approach to multivariate pattern analysis in order to fit single trial EEG data and found that only late parietal, but not early frontal EEG activity was predictive of recognition memory decisions. At first glance, this seems to be at odds with the current results. However, instead we suggest that the use of familiarity and recollection depends on the actual test situation. Here, we created conditions (that is an FCC display) in which familiarity had a better diagnostic reliability than in standard yes/no tasks as employed by the Ratcliff et al. study. Accordingly, using data from a source memory task, Noh et al. (2018) extracted an EEG classifier with a spatio-temporal distribution reminiscent of the mid-frontal old/new effect that best distinguished between hits without source judgments and correct rejections. This implicates that this component reflects a diagnostic familiarity signal.

As old/new differences at parietal electrodes from 500 ms onwards are normally associated with recollection, we also analyzed the time window from 500 to 650 ms. The old/new difference in this time window was not moderated by condition and displayed a similar topographical distribution as in the earlier time window. Prolonged frontal old/new effects that extend beyond 500 ms are not unusual and have been reported in a variety of studies before (Mecklinger et al., 2010; Schloerscheidt and Rugg, 2004; Tsivilis et al., 2001; Yu and Rugg, 2010). Rather, it is worth noting that we did not observe the typical late parietal portion of the old/new effects in this paradigm. However, interpreting this as a complete lack of recollective processing would certainly be exaggerating given the relatively high performance levels in both conditions. One explanation for the absence of the late parietal old/new effect in the FCNC condition might be pronounced recall-to-reject processing (Rotello et al., 2000), i.e. recall of item details of the originally studied picture upon the presentation of foils. If recollection takes place for targets and foils, differences between targets and foils are alleviated, thereby disguising the late parietal effect. Supporting this interpretation, Migo et al. (2009) found evidence for reliance on recall-to-reject in a remember/know variant of their experiment in which participants were asked to verbalize their decision process. A second possibility is that a late posterior negativity (LPN, see Mecklinger et al., 2016, for a review) to hits has masked the late parietal old/new effect. The LPN is assumed to reflect processes initiated to reconstruct prior study episodes, in particular in situations with high overlap of memory bound attributes, as for example when studied and non-studied pictures are highly similar. This is especially true for the FCNC condition, in which the foil resembles not the target in the actual trial but instead another studied picture. In support of this view, our ERP waveforms were more negative going in the FCNC condition compared to the FCC condition at central-parietal sites in both cycles. Finally, it is possible that recollective processing was spread across all four pictures of the trial sequence and the intervals between the pictures. Consequentially, recollection was presumably less time-locked to stimulus onset and therefore not observable in the ERPs. Note that temporal smearing is less likely for the mid-frontal old/new effect as familiarity is assumed to be elicited fast, more automatically, and with less temporal jitter.

Although not part of our predictions, performance was better in the FCC condition than in the FCNC condition. Most obviously, more reliable familiarity signals in the FCC condition than in the FCNC condition might have improved memory performance. Given the unusual topographical characteristics of the effect in the later time window, it is difficult to draw conclusions regarding the contribution of recollection to the behavioral difference. Importantly, greater amounts of recollection in the FCC condition would not challenge our main conclusions as the main aim of this study was to test the CLS predictions concerning familiarity.

Repeating pictures within a trial in the test phase seems to be both a methodological strength and caveat of this study. As outlined in the Introduction, it was necessary to present the pictures sequentially in order to obtain separate ERPs for targets and foils. As a consequence, the two conditions differed only after each picture had been presented at least once. In support of this view, Fig. 4A demonstrates that there was indeed no condition difference during the first picture position of the first cycle. Moreover, as also apparent in Fig. 4A, repetition of similar pictures in the FCC condition led to a positive shift in the waveforms during the second picture in the first cycle, presumably due to repetition priming (Penney et al, 2001, 2003). Consistent with this repetition priming view, this ERP difference between first and second presentation in the first cycle was virtually absent for the dissimilar pictures in the FCNC condition. In order to reduce a potentially confounding influence of these condition differences on the critical Condition × Item Type interaction, we focused the analysis on the second presentation cycle. Clearly, a within-trial repetition also bears the risk that differential processing in the first cycle affects processing in the second cycle and therefore ERPs have to be interpreted carefully. However, our control analysis revealed that there was no correlation between the condition difference during the first cycle and the old/new effect in the FCC condition in the second cycle. Therefore, we feel confident to conclude that the larger mid-frontal familiarity effect in the FCC condition was not an artefact of differential processing during the first cycle.

Our results also have implications for discussions on the functional significance of the mid-frontal old/new effect. Mirroring the imprecision in the definition of the familiarity process itself (e.g., feeling of "knowing" or recognition without recollection of details), the exact functional significance of the mid-frontal old/new effect remains elusive. One suggestion was that the mid-frontal old/new effect merely reflects differences in conceptual fluency between studied and non-studied items (Paller et al., 2007). However, our results add to other findings (e.g. Bader and Mecklinger, 2017; Bridger et al., 2012) strongly speaking against this explanation. We observed the mid-frontal old/new effect only in the FCC but not in the FCNC condition. That was the case despite the fact that across all items in a test list, differences in conceptual fluency between targets and foils were equated between conditions as foils in the FCNC condition were also similar to another picture from the study list. More precisely, since we assume that differences between the familiarity distributions of targets and foils are of the same size in the two test display conditions, the current results

suggest that the mid-frontal old/new effect reflects a task-adequate and fast *assessment* of the familiarity signal, not the signal associated with the pure familiarity strength value itself. This is in line with the assumption that familiarity is multiply determined. In previous studies (Bader et al., 2010; Bridger et al., 2014; Wiegand et al., 2010), we showed that early ERP old/new effects with parietal maxima are likely associated with *absolute* familiarity which signals the strength of the memory representation at a given time point. In contrast, more frontally distributed old/new effects, as revealed in the present FCC condition, were linked to the *relative* increment of the familiarity strength value for an item due to a specific recent encounter. Thus, relative familiarity is not an integral characteristic of a stimulus but the by-product of an assessment process. This is also supported by findings that the mid-frontal old/new effect is specifically tied to explicit recognition memory tasks, in which a discrimination between old and new items is task-relevant, i.e. if familiarity strength has to be assessed to guide recognition judgments (Ecker and Zimmer, 2009; Guillaume and Tiberghien, 2013; Küper et al., 2012). The mid-frontal old/new effect is usually not present in tasks requiring non-mnemonic judgments such as judgments of lifetime exposures (Yang et al., 2019). In these tasks, the mid-frontal familiarity effect is replaced by a more posterior effect, resembling the N400, an ERP index of semantic processing (see Mecklinger and Bader, 2020, for a review).

In this context, the question also arises whether the test display affects processing in the MTLC as suggested by the CLS (Norman & O'Reilly, 2003) or by other brain structures involved in familiarity judgements. Indeed, as we have discussed previously (Bader and Mecklinger, 2017), it seems more likely that the comparison of familiarity strength values, i.e. the assessment of the relative increment in familiarity and the requirement to make explicit recognition judgements, is mediated by the prefrontal cortex (PFC). The MTLC on the other hand might be more involved in the generation of the familiarity signal itself, i.e. absolute familiarity. Thus, even though inferences from scalp distributions of ERP effects on underlying brain systems have to be made with caution, we think that differences in the mid-frontal old/new effect between FCC and FCNC displays are most likely due to differences in prefrontal activity related to processes responsible for familiarity-based episodic decision making. Such a relationship between ventro-lateral PFC activity and the mid-frontal old/new effect was recently demonstrated by an EEG-informed fMRI study (Hoppstädter et al., 2015).

In conclusion, showing that the usefulness of a familiarity signal in a recognition memory task depends on the test format, we provide evidence in favor of the CLS model for recognition memory. Moreover, the current results suggest that the mid-frontal old/new effect does not reflect the mean difference in absolute familiarity strength between old and new items but instead reflects the assessment of the familiarity signal to pursue episodic recognition memory judgments.

Author contributions

Regine Bader, Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. Axel Mecklinger, Conceptualization, Methodology, Writing - review & editing. Patric Meyer, Conceptualization, Methodology, Writing - review & editing.

Acknowledgements

The authors thank Alexander Hauck, Kristin Pfaff, and Lisa Riedel for assistance with data collection and analyses, as well as Lisa K. Kuhn for her help with language editing.

Data availability

Data is available on: https://osf.io/n2w83/?view_only=d28941463fce4420a47ac6cad3804acd.

References

Bader, R., Mecklinger, A., 2017. Separating event-related potential effects for conceptual fluency and episodic familiarity. J. Cognit. Neurosci. 29 (8), 1402–1414.
Bader, R., Mecklinger, A., Hoppstädter, M., Meyer, P., 2010. Recognition memory for one-trial-unitized word pairs: evidence from event-related potentials. Neuroimage 50 (2), 772–781.
Bastin, C., van der Linden, M., 2003. The contribution of recollection and familiarity to recognition memory: a study of the effects of test format and aging. Neuropsychology 17 (1), 14–24.
Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software 67 (1), 1–48.
Bayley, P.J., Wixted, J.T., Hopkins, R.O., Squire, L.R., 2008. Yes/no recognition, forced-choice recognition, and the human hippocampus. J. Cognit. Neurosci. 20 (3), 505–512.
Bridger, E.K., Bader, R., Kriukova, O., Unger, K., Mecklinger, A., 2012. The FN400 is functionally distinct from the N400. Neuroimage 63 (3), 1334–1342.
Bridger, E.K., Bader, R., Mecklinger, A., 2014. More ways than one: ERPs reveal multiple familiarity signals in the word frequency mirror effect. Neuropsychologia 57, 179–190.
Curran, T., Doyle, J., 2011. Picture superiority doubly dissociates the ERP correlates of recollection and familiarity. J. Cognit. Neurosci. 23 (5), 1247–1262.
Ecker, U.K.H., Zimmer, H.D., 2009. ERP evidence for flexible adjustment of retrieval orientation and its influence on familiarity. J. Cognit. Neurosci. 21 (10), 1907–1919.
Guillaume, F., Tiberghien, G., 2013. Impact of intention on the ERP correlates of face recognition. Brain Cognit. 81 (1), 73–81.
Holdstock, J.S., Mayes, A.R., Roberts, N., Cezayirli, E., Isaac, C.L., O'Reilly, R.C., Norman, K.A., 2002. Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? Hippocampus 12, 341–351.
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6 (2), 65–70.
Hoppstädter, M., Baeuchl, C., Diener, C., Flor, H., Meyer, P., 2015. Simultaneous EEG–fMRI reveals brain networks underlying recognition memory ERP old/new effects. Neuroimage 116, 112–122.
Jäger, T., Mecklinger, A., Kipp, K.H., 2006. Intra- and inter-item associations doubly dissociate the electrophysiological correlates of familiarity and recollection. Neuron 52 (3), 535–545.
Jasper, H., 1958. Report of the committee on methods of clinical examination in electroencephalography. Electroencephalogr. Clin. Neurophysiol. 10, 370–375.
Jeneson, A., Kirwan, C.B., Hopkins, R.O., Wixted, J.T., Squire, L.R., 2010. Recognition memory and the hippocampus: a test of the hippocampal contribution to recollection and familiarity. Learn. Mem. 17 (1), 63–70.
Küper, K., Groh-Bordin, C., Zimmer, H.D., Ecker, U.K.H., 2012. Electrophysiological correlates of exemplar-specific processes in implicit and explicit memory. Cognit. Affect Behav. Neurosci. 12 (1), 52–64.
Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Software 82 (13).
Lawrence, M.A., 2016. Ez: Easy Analysis and Visualization of Factorial Experiments. R Package Version 4.4-0. https://CRAN.R-project.org/package=ez.
Mecklinger, A., Bader, R., 2020. From fluency to recognition decisions: a broader view of familiarity-based remembering. Neuropsychologia, p. 107527.
Mecklinger, A., Brunnemann, N., Kipp, K., 2010. Two processes for recognition memory in children of early school age: an event-related potential study. J. Cognit. Neurosci. 23 (2), 435–446.
Mecklinger, A., Rosburg, T., Johansson, M., 2016. Reconstructing the past: the late posterior negativity (LPN) in episodic memory studies. Neurosci. Biobehav. Rev. 68, 621–638.
Migo, E., Montaldi, D., Norman, K.A., Quamme, J., Mayes, A., 2009. The contribution of familiarity to recognition memory is a function of test format when using similar foils. Q. J. Exp. Psychol. 62 (6), 1198–1215.
Morcom, A.M., 2015. Resisting false recognition: an ERP study of lure discrimination. Brain Res. 1624, 336–348.
Noh, E., Liao, K., Mollison, M.V., Curran, T., Sa, V. R. de, 2018. Single-trial EEG analysis predicts memory retrieval and reveals source-dependent differences. Front. Hum. Neurosci. 12, 258.
Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev. 110 (4), 611–646.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 9 (1), 97–113.
Opitz, B., Cornell, S., 2006. Contribution of familiarity and recollection to associative recognition memory: insights from event-related potentials. J. Cognit. Neurosci. 18 (9), 1595–1605.
Paller, K.A., Voss, J.L., Boehm, S.G., 2007. Validating neural correlates of familiarity. Trends Cognit. Sci. 11 (6), 243–250.
Penney, T.B., Maess, B., Busch, N., Derrfuss, J., Mecklinger, A., 2003. Cortical activity reduction with stimulus repetition: a whole-head MEG analysis. Cognit. Brain Res. 16 (2), 226–231.
Penney, T.B., Mecklinger, A., Nessler, D., 2001. Repetition related ERP effects in a visual object target detection task. Cognit. Brain Res. 10 (3), 239–250.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
Ratcliff, R., Sederberg, P.B., Smith, T.A., Childers, R., 2016. A single trial analysis of EEG in recognition memory: tracking the neural correlates of memory strength. Neuropsychologia 93, 128–141.

Rosburg, T., Mecklinger, A., Frings, C., 2011. When the brain decides: a familiarity-based approach to the recognition heuristic as evidenced by event-related brain potentials. Psychol. Sci. 22 (12), 1527–1534.
Rotello, C.M., Macmillan, N.A., Van Tassel, G., 2000. Recall-to-reject in recognition: evidence from ROC curves. J. Mem. Lang. 43 (1), 67–88.
RStudio Team, 2016. RStudio. *Integrated development environment for R*. RStudio, Inc. http://www.rstudio.com/.
Rugg, M.D., Curran, T., 2007. Event-related potentials and recognition memory. Trends Cognit. Sci. 11 (6), 251–257.
Rugg, M.D., Mark, R.E., Walla, P., Schloerscheidt, A.M., Birch, C.S., Allan, K., 1998. Dissociation of the neural correlates of implicit and explicit memory. Nature 392, 595–598.
Schloerscheidt, A.M., Rugg, M.D., 2004. The impact of change in stimulus format on the electrophysiological indices of recognition. Neuropsychologia 42, 451–466.
Schwikert, S.R., Curran, T., 2014. Familiarity and recollection in heuristic decision making. J. Exp. Psychol. Gen. 143 (6), 2341.
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 Res. 1122 (1), 161–170.
Voss, J.L., Paller, K.A., 2009. An electrophysiological signature of unconscious recognition memory. Nat. Neurosci. 12, 349–355.
Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., 2019. Welcome to the tidyverse. Journal of Open Source Software 4 (43), 1686.
Wiegand, I., Bader, R., Mecklinger, A., 2010. Multiple ways to the prior occurrence of an event: an electrophysiological dissociation of experimental and conceptually driven familiarity in recognition memory. Brain Res. 1360, 106–118.
Wilding, E.L., Rugg, M.D., 1996. An event-related potential study of recognition memory with and without retrieval of source. Brain 119 (3), 889–905.
Woodruff, C.C., Hayama, H.R., Rugg, M.D., 2006. Electrophysiological dissociation of the neural correlates of recollection and familiarity. Brain Res. 1100, 125–135.
Yang, H., Laforge, G., Stojanoski, B., Nichols, E.S., McRae, K., Köhler, S., 2019. Late positive complex in event-related potentials tracks memory signals when they are decision relevant. Sci. Rep. 9 (1), 9469.
Yonelinas, A.P., 2002. The nature of recollection and familiarity: a review of 30 years of research. J. Mem. Lang. 46 (3), 441–517.
Yu, S.S., Rugg, M.D., 2010. Dissociation of the electrophysiological correlates of familiarity strength and item repetition. Brain Res. 1320, 74–84.