The neural mechanism of fluency-based memory illusions: the role of fluency context

Carlos Alexandre Gomes, Axel Mecklinger, and Hubert Zimmer

Department of Psychology, Saarland University, D-66123 Saarbrücken, Germany

Recognition memory judgments can be influenced by a variety of signals including fluency. Here, we investigated whether the neural correlates of memory illusions (i.e., misattribution of fluency to prior study) can be modulated by fluency context. Using a masked priming/recognition memory paradigm, we found memory illusions for low confidence decisions. When fluency varied randomly across trials, we found reductions in perirhinal cortex (PrC) activity for primed trials, as well as a (pre)cuneus-PrC (BA 35) connectivity. When the fluency context was unchanging, there was increased PrC activity for primed trials, with the (pre)cuneus showing greater connectivity with PrC (BA 36). Thus, our results tentatively suggest two neural mechanisms via which fluency can lead to memory illusions.

A long-held theory of recognition memory is that it can be supported by recollection (the retrieval of detailed contextual information associated with an event), as well as by familiarity (an acontextual sense that something has been previously experienced) (Yonelinas 2002). In recent years, the idea that recognition memory judgments can be influenced by how fluent items are perceived has gained much support. Several studies have found that participants exhibit a tendency to judge an item “old” if fluency associated with that item has been previously enhanced (Jacoby and Whitehouse 1989; Whittlesea et al. 1990; Whittlesea and Williams 2000). This experience of perceived greater fluency has been demonstrated even for items which had not actually been studied before, producing what some researchers have termed “illusions of memory” (Whittlesea et al. 1990; Whittlesea 1993).

Converging electrophysiological and functional magnetic resonance imaging (fMRI) evidence suggests that, first, fluency-based memory is associated with multiple neural networks, each of which has the potential to contribute to recognition memory judgments (Woollams et al. 2008; Park and Donaldson 2016; Li et al. 2017), and second, that activity changes in the perirhinal cortex (PrC) (Dew and Cabeza 2013; Wang et al. 2014, 2016) and its connectivity with the (pre)cuneus (Dew and Cabeza 2013) reflect fluent processing.

Recently, we conducted an fMRI study in which we manipulated the fluency context, such that a group of participants was presented with primed and unprimed words intermixed with each other (random context [RC]), whereas a second group was presented with primed and unprimed words in separate blocks (blocked context [BC]) (Gomes et al. 2017). Priming increased high confidence (HC) recognition memory performance in the RC (but not BC) condition, suggesting that a fluency context that allowed variations in fluency on a trial-by-trial level (i.e., RC) was determinant for the ability of priming to support objective recognition memory (see also Leynes and Zish 2012; Brueett and Leynes 2015).

Whereas our previous paper (Gomes et al. 2017) looked at the effect of priming on recognition memory accuracy, here, we examined whether the neural correlates of fluency-based memory illusions may be also sensitive to alterations of fluency context. For that purpose, we used the same data set as in Gomes et al. (2017) but restricted our analyses to FAs, as this category consists of unstudied words that are (incorrectly) categorized “old,” such that any increase in FAs due to priming could be used as a proxy for fluency-based memory illusions (see Dew and Cabeza 2013 for a similar rationale). We were also interested to test whether priming would be obtained in the BC condition (Leynes and Zish did not report fluency effects for FAs separately, so it is unclear whether memory illusions occurred in the BC condition).

For the RC, we hypothesized that the (pre)cuneus and the right PrC would be involved in fluency-based memory misattributions. The (pre)cuneus has been consistently implicated in the production of both true and false memories (Cabeza et al. 2001; Gonsalves et al. 2004; Slotnick and Schacter 2004; Garoff-Eaton et al. 2006; Atkins and Reuter-Lorenz 2011; Wang et al. 2016), suggesting that activity in this region during recognition memory tests may reflect perceived oldness. The PrC has been shown to deactivate during conceptually driven priming (Voss et al. 2009; Wang et al., 2010, 2014; Gomes et al. 2016), and at least two studies have directly implicated the PrC in the illusory memory effect (Dew and Cabeza 2013; Wang et al. 2016).

Hypotheses regarding the BC condition are more difficult to formulate, given that most fluency-related studies have used a paradigm in which fluent and control trials are randomly intermixed (as in our RC experiment). Nevertheless, if fluency-based misattributions also occur in the BC condition, then there is the possibility that (pre)cuneus-PrC connectivity may also be present in the BC.

In this experiment, 48 participants—24 in the RC and 24 in the BC—were recruited (mean age: 23; 26 females). The stimuli consisted of 600 low-frequency German words. At study, participants judged whether words (200 in total) contained exactly two vowels or more/less than two vowels, followed by a fixation cross of jittered duration. At test, they judged whether a target word presented for 2 sec was either old or new. This target word was always preceded by a prime, which was sandwiched between two masks and was either the same (primed trials) or a different (unprimed trials) word than the target word. There was a total of 200 primed prime/target and 200 unprimed prime/target trials, half of which had been presented in the study phase. After the recognition decision, participants selected among four choices how confident they were about each response. The percentage of each response category was calculated for the RC and BC conditions, and the percentage of items judged as “old” was compared between conditions.
were that the target word was old/new (see Fig. 1 and Gomes et al. 2017 for a more detailed description). At the end of the trial, a fixation cross of jittered duration was presented. After completion of the experiment, all participants performed a prime detection test outside the scanner. They were all at chance level at detecting the primes.

The behavioral analysis focused on proportion of trials as the dependent variable (for the descriptive statistics of all response categories split by prime status, confidence, and context, see Gomes et al. 2017). Because we were interested in memory misattributions, only FAs were analyzed in this paper. We collapsed the first two rating choices into a low confidence (LC) category, and the third and fourth choices into a HC condition (Gomes et al. 2017). The experimental design consisted of prime status (primed, unprimed) and confidence (LC, HC) as a within-subject factors and context (RC, BC) as a between-subject factor. The data were analyzed using (mixed) repeated-measures analysis of variance (ANOVA) and t-tests. A Huynh-Feldt correction was applied to the degrees of freedom of those tests for which the assumption of sphericity was violated. The α level was set, for all statistical tests, at 0.05.

The MRI data were collected using a 3T Siemens Skyra MRI scanner. Functional images were acquired using a T2*-weighted echo-planar imaging (EPI) sequence, providing blood oxygen level-dependent (BOLD) contrast (TR = 1.8 sec, TE = 35 ms, flip angle = 75°, FOV = 240 × 240 mm², and matrix size = 96 × 96, voxel resolution = 2.5 × 2.5 × 3.75 mm³). Each image volume consisted of 32 contiguous axial slices, positioned parallel to the AC-PC axis and acquired in ascending order. A high-resolution structural T1-weighted image was also acquired from each participant (voxel resolution = 0.9375 × 1.1733 × 0.90 mm³).

The fMRI data were analyzed using the Statistical Parametric Mapping (SPM8, http://www.fil.ion.ucl.ac.uk/spm/) toolbox. Preprocessing of images included spatial realignment to correct for head movement, slice timing correction to account for slice acquisition order, coregistration of the structural image to the mean functional image, segmentation of the structural image, normalization of the structural and functional images to Montreal Neurological Institute (MNI) space, and spatial smoothing (full-width at half-maximum [FWHM] = 6 mm).

Using our previous participant-specific general linear model (GLM), statistical parametric maps (SPMs) of the t-statistic were generated for the contrasts between primed and unprimed FAs, implicitly masked by the main effect of priming, and cluster-wise corrected for multiple comparisons at P < 0.05 (10-voxel extent). For ROI analysis, the signal associated with each contrast was computed and averaged across the relevant ROI and subsequently submitted to ANOVAs and t-tests.

For functional connectivity analysis, denoising was further applied to the fMRI time series using the aCompCor approach (Behzadi et al. 2007), which regresses out the five principal components of white matter and CSF. Main effects of task and the motion parameters estimated during realignment were also added as covariates. We created 8-mm spheres for regions of interest on the basis of either significant activations in the current study (e.g., (pre) cuneus) or previous research (e.g., right PrC). Subsequently, we calculated the Fisher z-transformed correlation coefficient value between a seed and target regions for the primed > unprimed FAs contrast and for both the RC and BC conditions. For both conditions, our main connectivity analysis was that between the (pre) cuneus (seed) and PrC (target), since this connection has been observed in at least one fMRI study on fluency-based memory decisions (e.g., Dew and Cabeza 2013). We also examined functional connectivity among other ROIs that were significant in the primed versus unprimed FAs contrast. For the RC, these ROIs were the right middle occipital gyrus, left inferior occipital and fusiform gyrus. For the BC, the ROIs were the bilateral putamen, right middle frontal gyrus and middle temporal gyrus. Functional connectivity analyses were conducted using CONN (www.nitrc.org/projects/conn).

Behaviorally, a 2 Context × 2 Confidence × 2 Prime Status mixed repeated-measures ANOVA on FAs yielded a significant main effect of confidence, F(1,40) = 10.72, P < 0.01, with a greater proportion of LC (0.15) relative to HC (0.10) FAs, and a significant main effect of prime status, F(1,40) = 4.62, P < 0.05, with a greater proportion of primed (0.13) relative to unprimed (0.12) FAs. The confidence by prime status interaction also reached significance, F(1,40) = 4.78, P < 0.05, which indicated that, for both the RC and BC conditions there was a greater incidence of primed (0.16) relative to unprimed (0.14) FAs for LC ratings (RC: l(20) = 1.73, P < 0.05, d = 0.50; BC: l(20) = 2.32, P < 0.05, d = 0.51), whereas no differences were found between primed (0.10) and unprimed (0.10) FAs for HC ratings (Gomes et al. 2017). No other effects were significant.

Regarding the fMRI analysis, we initially conducted the primed versus unprimed FAs contrast, separately for the RC and BC conditions, which revealed a significant cluster in the left and right (pre)cuneus, respectively (Fig. 2A). We extracted the parameter estimates from an 8-mm sphere around the peak coordinate of these clusters, and submitted the data to a 2 Context × 2 Confidence × Prime Status mixed repeated-measures ANOVA. This analysis only revealed a significant three-way interaction, F(1,39) = 8.34, P < 0.01. Separate ANOVAs revealed that, for HC data, the left (pre)cuneous showed increased response for primed versus unprimed FAs in the RC, whereas the right

![Figure 1. Experimental design. At study, participants were shown low-frequency German words and asked to decide whether a word had either exactly two vowels or more/less than two vowels. At test, a prime word was presented very briefly and sandwiched between masks (random array of seven symbols). A target word was then shown, to which participants made old/new (Alt/Neu) recognition memory decisions. In half of the trials, prime and target were the same word (primed trials, e.g., prime: fugen, target: FUGEN), whereas in the other half they were a different word (unprimed trials, e.g., prime: zutun, target: ARTELEN). After the old/new memory judgment, they decided how confident they were by selecting one out of four rating options that varied in confidence strength: Raten (Guess), Nicht sicher (Not sure), Sicher (Sure), Wissen (Know).](image-url)
Confluent clusters identified using an 8-mm sphere centered around the peak coordinate of the PrC in response to primed FAs, an ROI analysis was conducted for primed relative to unprimed FAs in the RC experiment, whereas the reverse was apparent in the BC (both appeared to be reduced relative to unprimed LC FAs in the RC experiment). Although the three-way interaction was not significant, the opposite effect in the BC experiment. Given that the (pre)cuneus-PrC connectivity was of primary interest, we conducted a 2 Prime Status × 2 PrC (BA35, BA36) × 2 Context mixed repeated-measures ANOVA using the PrC ROIs from the previous analysis. There was only a significant three-way interaction (primed > unprimed trials) between the (pre)cuneus and the PrC BA 35 in the RC, $t_{(1,36)} = 4.84$, $P < 0.05$. As seen in Figure 3 (bottom), this was due to increased connectivity (primed > unprimed trials) between the (pre)cuneus and the PrC BA 35 in the RC, $t_{(1,36)} = 4.13$, $P < 0.05$; see Fig. 2B).

For the BC, there was reduced activation for primed versus unprimed FAs in the bilateral putamen, right middle frontal gyrus, and left middle temporal gyrus (see Fig. 2C, bottom). There were no interactions with confidence in these regions in either context.

Because whole-brain analysis revealed no activity change in the PrC in response to primed FAs, an ROI analysis was conducted using an 8-mm sphere centered around the peak coordinate of the PrC clusters identified in Dew and Cabeza’s study (BA 35 [33, −26, −23] and BA 36 [37, −15, −27]). For the PrC BA 35, a 2 Context × 2 Confidence × 2 Prime Status repeated-measures ANOVA revealed a significant Context × Prime Status interaction, $F_{(1,36)} = 4.50, P < 0.05$, (all other $Fs < 3.64, Ps > 0.10$), indicating decreased activity for primed relative to unprimed FAs in the RC experiment, and the opposite effect in the BC experiment. Although the three-way interaction was not significant, PrC activity for primed LC FAs appeared to be reduced relative to unprimed LC FAs in the RC experiment, whereas the reverse was apparent in the BC (both $Ps < 0.06$, see Fig. 2D). In contrast, these differences did not approach significance for the HC condition (both $Ps > 0.10$). Even though the analysis of the PrC BA 36 did not reveal any significant effects, there was a numerical trend consistent with the PrC BA 35 results (see Fig. 2D).

For the connectivity analyses, we decided to focus on LC trials since the behavioral and neural effects were mostly present in this condition. Given that the (pre)cuneus-PrC connectivity was of primary interest, we conducted a 2 Prime Status × 2 PrC (BA35, BA36) × 2 Context mixed repeated-measures ANOVA using the PrC ROIs from the previous analysis. There was only a significant three-way interaction (primed > unprimed trials) between the (pre)cuneus and the PrC BA 35 in the RC, $t_{(1,36)} = 4.13$, $P < 0.05$. As seen in Figure 3 (bottom), this was due to increased connectivity (primed > unprimed trials) between the (pre)cuneus and the PrC BA 35 in the RC, $t_{(1,36)} = 4.84$, $P < 0.05$. For the sake of exploration, we also examined functional connectivity among the ROIs identified in the priming contrast, noting that these analyses were uncorrected for multiple comparisons. There was increased connectivity between the right middle occipital gyrus seed and the left fusiform and inferior occipital gyri targets only in the RC (both $ts > 2.22, Ps < 0.05, ds > 0.52$), which was greater for primed relative to unprimed LC FAs. Regarding the BC, there was only increased connectivity between the right (pre)cuneus and the bilateral putamen, $t_{(1,36)} = 1.97$, $P < 0.05, d = 0.44$, which, surprisingly, was also significant in the RC condition, $t_{(1,36)} = 3.35, P < 0.01, d = 0.79$. Figure 3 (top) summarizes the results of the combined connectivity analyses.

The present study aimed at elucidating whether the neural correlates of the illusory memory effect (i.e., misattributions of enhanced fluency to prior study) might also be influenced by changes in fluency context. Our results suggest that the memory illusory effect may only occur for LC decisions and could indicate two distinct neural mechanisms via which fluency signals may be interpreted as evidence of prior study. When primed and unprimed words were randomly interspersed (RC condition), the brain mechanism responsible for memory misattributions relied on a strong precuneus-PrC connectivity, as well as visual cortex, communication. Although Dew and Cabeza (2013) had already linked (pre)cuneus-PrC connectivity with fluency-related misattributions

**Figure 2.** fMRI results. (A) Location of the (pre)cuneus cluster detected in the primed versus unprimed FAs in both the RC ($[−6, −82, 37]$, red) and BC ($[18, −79, 22]$, green). (B) Mean parameter estimates within the (pre)cuneus ROIs in both types of context. (C) The left fusiform gyrus ($[−18, −97, −11]$) and bilateral inferior/middle occipital gyrus (left: $[−24, −88, 1]$, right: $[24, −94, 7]$) also showed sensitivity to the primed versus unprimed FAs contrast in the RC (top), whereas the right middle frontal gyrus ($[30, 2, 46]$), bilateral putamen (left: $[−18, 14, 4]$, right: $[21, 17, 2]$), and middle temporal gyrus ($[−42, −58, 7]$) not shown) also showed sensitivity to the primed versus unprimed FAs contrast in the BC (bottom). (D) Mean parameter estimates within the PrC ROIs (left: BA 35, right: BA 36) for both the RC (red) and BC (green). P: primed, UP: unprimed, HC: high confidence, LC: low confidence.
using a paradigm not dissimilar to ours, we additionally showed that this connection seems true only when decision confidence is particularly low. It is noteworthy that some research suggests that PrC activity during objective memory is detected particularly when decisions rely on weak memory traces (Davachi et al. 2003; Ranganath et al. 2004; Song et al. 2011). Our results extend this hypothesis by suggesting that both the right PrC and (pre)cuneus also contribute to memory illusions during weak confidence judgments. When primed and unprimed trials were presented in separate blocks (BC condition), we also observed (pre)cuneus-PrC connectivity but in a different PrC location (BA 36). Furthermore, the PrC (both BA 35 and 36) exhibited reduced activity for primed relative to unprimed LC trials in the RC, but increased activity for primed items in the BC (although only significantly so in BA 35). There is recent evidence that the PrC increases in activity with life-time experience with object concepts, but it decreases in activity with accrued experimental familiarity (Duke et al. 2017). Indeed, masked priming during the BC condition recruited many of the same brain regions that ostensibly support semantic priming (Gomes et al. 2017). A speculative possibility could be that, given that fluency from masked primes was unchanging in the BC condition, conceptual processing of words predominated, leading to increased activity in the PrC.

Memory illusions were also accompanied by enhanced connectivity between the (pre)cuneus-putamen in both RC and BC conditions. The putamen involvement in this experiment could relate to facilitated access (due to masked priming) to sublexical/lexical information, a function commonly associated with this region (Kotz et al. 2002; Oberhuber et al. 2013). This processing facilitation, combined with (pre)cuneus input, may have induced a feeling that the words had been experienced before, leading to memory illusions.

At least two caveats should be mentioned. First, the behavioral masked priming effect was rather small (around 2% difference in magnitude), which could raise concerns about replicability. The literature on masked priming is, however, unequivocal in this respect—masked priming is a robust phenomenon notwithstanding the small experimental effects (often as small as ours) (Kinoshita and Lupker 2003). In any case, future experiments could use manipulations aimed to increase priming levels (e.g., using multiple prime repetitions). Second, contrary to Dew and Cabeza, we did not observe PrC-related changes in activity at the whole-brain level in either fluency context. This region is well-known to suffer from significant signal dropout due to air-tissue interfaces in the vicinity of the medial temporal lobe, which could have induced strong susceptibility artifacts (Olman et al. 2009). There was evidence, however, of PrC involvement when an ROI analysis was conducted. In fact, the primed and unprimed LC parameter estimates for FAs were strikingly similar to those obtained by Dew and Cabeza (see their Figure 3).

To our knowledge, our study is the first to demonstrate that the memory illusion effect may rely on a distinct neural mechanism, depending on what kind of fluency context was manipulated at test. In particular, the PrC decreased in activity for primed versus unprimed trials in the RC, and the BA 35 (but not BA 36) subregion showed strong connectivity with the (pre)cuneus. In contrast, the PrC appeared to increase in activity for primed versus unprimed trials in the BC, with the (pre)cuneus showing greater connectivity with BA 36 (but not BA 35). Furthermore, there seemed to be differences in visual cortex connectivity between RC and BC, but these analyses were post-hoc and should be interpreted with caution. Future research will be needed to verify that these results hold when different fluency context manipulations are used.

Acknowledgments
This work was supported by grant IRTG-1457 from the Deutsche Forschungsgemeinschaft (DFG). We thank Simon Kohl, Sandra Dürrenbächer, and Benjamin Sturm for their assistance with data collection.

References
Atkins AS, Reuter-Lorenz PA. 2011. Neural mechanisms of semantic interference and false recognition in short-term memory. Neuroimage 56: 1726–1734. doi:10.1016/j.neuroimage.2011.02.048
Behzadi Y, Restom K, Liu J, Liu TT. 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37: 90–101. doi:10.1016/j.neuroimage.2007.04.042
Bruett H, Leynes PA. 2015. Event-related potentials indicate that fluency can be interpreted as familiarity. Neuropsychologia 78: 41–50. doi:10.1016/j.neuropsychologia.2015.09.035
