Brod, Garvin; Shing, Yee Lee

Specifying the role of the ventromedial prefrontal cortex in memory formation

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
Neuropsychologia 111 (2018), S. 8-15

Specifying the role of the ventromedial prefrontal cortex in memory formation

Garvin Brod[a,b] & Yee Lee Shing[a,c]

[a]Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany; [b]German Institute for International Educational Research, Frankfurt am Main, Germany; [c]Division of Psychology, University of Stirling, Stirling, UK

1 Correspondence concerning this article should be addressed to Garvin Brod or Yee Lee
2 Shing: garvin.brod@dipf.de; yee.shing@stir.ac.uk
3

Abstract

Recent neuroimaging research suggests that the ventromedial prefrontal cortex (vmPFC) plays an important role for successful memory formation that takes place in the context of activated prior knowledge. These findings led to the notion that the vmPFC integrates new information into existing knowledge structures. However, a considerable number of neuroimaging studies that have investigated memory formation in the context of prior knowledge have not found vmPFC involvement. To resolve this inconsistency, we propose a distinction between knowledge-relevance (the degree to which new information can be linked to prior knowledge) and knowledge-congruency (the perceived match between prior knowledge and the to-be-encoded information). We hypothesized that the vmPFC contributes to successful memory formation only when perceived knowledge-congruency is high, independent of knowledge-relevance. We tested this hypothesis in a design that varied both congruency and relevance during memory encoding, which was performed in the MR scanner. As predicted, the results showed that vmPFC contributions to memory formation vary as a function of knowledge-congruency, but not as a function of knowledge-relevance. Our finding contributes to elucidating the seemingly inconsistent findings in the literature and helps to specify the role of the vmPFC in memory formation.

Introduction

In recent years, cognitive neuroscience research on memory has become increasingly interested in the role of the ventromedial prefrontal cortex (vmPFC) in all stages of memory processing. Starting with the observation that vmPFC lesions can lead to confabulation (Moscovitch, 1989; Moscovitch & Melo, 1997), a role for the vmPFC in retrieval monitoring was proposed in which the vmPFC provides a "feeling of rightness" for memory cues during retrieval (Moscovitch & Winocur, 2002). Following this account, vmPFC contributions are not necessary for memory retrieval, but a lack of them leads to the erroneous retrieval of inappropriate associations. On the contrary, vmPFC contributions can also increase erroneous retrieval in a situation in which memories have to be rejected that fit well into an activated knowledge structure (also called schema, Berkers et al., 2016; Warren, Jones, Duff, & Tranel, 2014). This double-edged role of the vmPFC can best be illustrated by its contribution to the so-called congruency effect, which denotes a memory advantage for knowledge-congruent as opposed to knowledge-incongruent new information. The congruency effect can be interpreted as an estimate of the influence of prior knowledge on episodic memory. vmPFC patients do not show this effect (Spalding, Jones, Duff, Tranel, & Warren, 2015). In line with this lesion data, recent functional magnetic resonance imaging (fMRI) studies have shown that the vmPFC displays enhanced activation for successfully retrieved knowledge-congruent as compared to knowledge-incongruent information (Brod, Lindenberger, Werkle-Bergner, & Shing, 2015; van Kesteren, Rijpkema, Ruiter, & Fernández, 2010).

Concerning the role of the vmPFC in memory formation, results from a patient study (Ghosh, Moscovitch, Melo Colella, & Gilboa, 2014) suggest that vmPFC lesions lead to deficient knowledge representation and activation, which is a prerequisite for knowledge-mediated memory formation. fMRI studies have found enhanced vmPFC activation for successfully encoded information (van Kesteren et al., 2013; 2014) as well as for successful

inference performance during knowledge-related memory encoding (Schlichting & Preston, 2016; Zeithamova, Dominick, & Preston, 2012). Consequently, it has been argued that the role of the vmPFC during memory encoding is to support the integration of new information into existing knowledge structures (Gilboa & Marlatte, 2017; Schlichting & Preston, 2015; van Kesteren, Ruiter, Fernández, & Henson, 2012). Based on findings in animals, it has been suggested that the mPFC is suited for this role because of its direct anatomical connections to the hippocampus (Nieuwenhuis & Takashima, 2011).

Despite this seemingly clear picture, it has to be acknowledged that a considerable number of studies that have used memory tasks for which prior knowledge should be activated and used have not found vmPFC activation that was predictive of later memory (Bein, Reggev, & Maril, 2014; Brod, Lindenberger, Wagner, & Shing, 2016; van Buuren et al., 2014; Webb, Turney, & Dennis, 2016). Conversely, other studies that have found differential vmPFC involvement in successful memory encoding did not use conditions that clearly differed in prior knowledge activation (e.g., Benoit, Szpunar, & Schacter, 2014; Reggev, Bein, & Maril, 2016). Therefore, the proposed relationship between prior knowledge-related memory processing and vmPFC activation is likely more complicated than initially believed, and there may be several boundary conditions that determine whether or not the vmPFC is involved.

We (Brod, Werkle-Bergner, & Shing, 2013) have speculated before that the vmPFC might be involved only when there is a strong congruency dimension in the task, and not when information is encoded against the backdrop of prior knowledge. In other words, we proposed that knowledge-congruency can be distinguished from knowledge-relevance. Knowledge-relevance describes the degree to which the to-be-remembered information can be linked to a pre-existing semantic network, and, thus, the degree to which prior knowledge can be used to enable elaborative (i.e., semantic) encoding. By knowledge-congruency, we mean

the degree to which the information evokes a sense of fit to the particular, activated knowledge structures (similar to the "feeling of rightness" notion in memory retrieval by Moscovitch & Winocur, 2002). Following this terminology, examples of common memory tasks containing a knowledge-relevance but no knowledge-congruency dimension include object–place associations in familiar vs. unfamiliar task environments or high vs. low expertise conditions. Conversely, memory tasks containing a knowledge-congruency but not a knowledge-relevance dimension include object–place associations in a familiar task environment in which an object can be expected vs. not expected to occur at a particular location or event memory for rule-consistent vs. rule-violating chess moves. In short, the congruency dimension comes into play in the context of expectancies that are confirmed or violated, whereas the relevance dimension comes into play whenever stimuli make varying levels of connection to prior knowledge. The two dimensions are not proposed to be mutually exclusive, i.e., there are situations in which the proposed congruency and relevance dimensions are positively correlated.

In the current study, we sought to include both the knowledge-congruency and the knowledge-relevance dimension in the same memory encoding task to be able to delineate vmPFC contributions to prior knowledge-related memory encoding more precisely. We present new analyses of a previously published data set (Brod et al., 2016) that examined how real-life gains in knowledge affect the neural correlates of episodic encoding, as measured by fMRI. Final year medical students were tested on an episodic memory task related to medical knowledge before and after their final exam. For the current purpose, we only analyzed data from the first measurement occasion. In the memory task, participants had to memorize either face–diagnosis (high knowledge-relevance) or face–name (low knowledge-relevance) pairs. Common names and familiar diagnoses (determined in pilot studies) were used along with unfamiliar Caucasian faces. The design of the memory task was inspired by previous research

showing that remembering face–name associations is much more difficult than remembering face–personal feature associations, because common names are arbitrary (except for allowing inferences about gender and, sometimes, nationality) and, thus, lack clear semantic associations (e.g., Cohen, 1990; McWeeny, Young, Hay, & Ellis, 1987). On the other hand, personal features (such as, in our case, a known medical diagnosis given to a person) are linked to a rich semantic network, which facilitates elaborative, semantic encoding (cf. Cohen, 1990; McWeeny et al., 1987). Thus, the diagnoses and names used in our study are assumed to differ in the extent to which they evoke a schema that can be applied to elaborate on a given face (e.g., a person with chronic obstructive pulmonary disease (COPD) will likely have slightly blue lips and look pale vs. a Michael may have blond hair). In sum, while we do not imply that prior knowledge cannot be leveraged at all for remembering face–name pairs, based on previous research we assume that it can be elaborated less effectively than for remembering face–diagnosis pairs which evoke a rich semantic network in medical exam candidates. Importantly, we additionally examined subjective congruency ratings during encoding, which were not explicitly modeled in previous analyses (see Brod et al., 2016). This gave us leverage to examine both the knowledge-congruency and the knowledge-relevance dimension within the same memory encoding task.

We hypothesized that vmPFC activation would distinguish between knowledge-congruent and knowledge-incongruent associations, but not between high and low knowledge-relevance associations. In particular we hypothesized a higher vmPFC activation for congruent as compared to incongruent information, and an enhanced vmPFC contribution to successful memory encoding of congruent information. In contrast, we expected the vmPFC to not display differential activity nor to contribute differentially to successful memory encoding for high vs. low knowledge-relevance associations. We tested these hypotheses in two parallel analyses. In one set of analyses, we compared vmPFC activation

for congruent and incongruent events as well as for events of high vs. low knowledge-relevance separately. Next, we tested whether vmPFC regions detected in these contrasts overlapped with regions contributing to successful memory formation (i.e., remembered > forgotten contrast). In the other set of analyses, we extracted % signal change from an independently defined vmPFC cluster and submitted these values to a repeated-measures ANOVA to directly test whether the vmPFC involvement in successful memory formation differs as a function of knowledge-congruency and/or knowledge-relevance. This full factorial analysis was performed on a subset of the full sample to ensure there were sufficient number of trials within each cell of the factor levels (see Participants).

Materials and Methods

Participants

Complete data from forty-nine medical students (29 female, age range = 23–29 years, mean age = 25.6 years) were collected in the initial study (reported in Brod et al., 2016). Participants were recruited from Berlin universities and were paid 76 Euro for their participation. All participants were right-handed, had no history of psychiatric or neurological disorders, and gave written informed consent. The current analyses were performed on data from the first measurement occasion of Brod et al. (2016), but go beyond the previously published data in that they also take into account participants' congruency ratings during encoding. This was outside the scope of the earlier analyses, which focused on longitudinal changes in knowledge and how these relate to changes in brain activation patterns. However, due to the added factor of congruency rating in the current analysis, which led to eight instead of four within-subject conditions, twenty-four participants had to be excluded for the second (full factorial) set of analyses because they did not provide enough (>5) valid trials per block in every condition. Thus, data of twenty-five participants (19 female, age range = 23–29

years, mean age = 26.0 years) were analyzed for the current full factorial analysis. Ethics approval was obtained from the ethics committee of the German Psychological Society (DGPs).

Task and Procedure

The encoding phase was performed after the structural scans and took 20 minutes in total (for a graphical depiction of the task, see Figure 1). Before entering the MRI scanner, participants were instructed to memorize face–word pairs, in which half of the words were diagnoses and the other half were first names. They were told that there would be a memory test later, but no details were given concerning the nature of the memory test. They were further instructed to try to memorize both the face–diagnosis and face–name pairs equally well. A total of 140 medical diagnoses and 140 common German first names were used together with 140 neutral face pictures. Each face was pseudorandomly combined with one diagnosis and one name, whereby faces and names/diagnoses were matched for gender. Two parallel stimulus lists of 140 face–word pairs each were created and counterbalanced across participants. The stimulus lists were further subdivided into two experimental blocks, each consisting of 70 trials. The face stimuli consisted of pictures of Caucasian young adults taken from the Center for Vital Longevity Face Database (Minear & Park, 2004). Face–word pairs were presented for 5 seconds each in an interleaved fashion (in pseudorandom order). Trials were separated by a variable fixation cross period of 2–5 seconds (mean: 3.5 seconds). During presentation of the face–word pairs, participants were asked to indicate whether or not the name / diagnosis fit with the face (congruency judgment), responding with their left / right index finger. Left / right response options were counterbalanced across participants.

The retrieval phase took place outside of the scanner, about 10 minutes after the end of the encoding session. Participants were instructed that they would now see all 140 faces again (in pseudorandom order) and they would see each face together with either 4 first names or 4

diagnoses, of which one name/diagnosis had been presented with the face during the encoding phase (target), whereas the other three were seen with other faces during encoding (lures). Participants indicated their choice via button press. Afterwards, they were asked to indicate their decision confidence on a scale of 1 (guess) to 4 (very sure). They were given no time limit for their responses, but were told to answer as quickly and as correctly as possible.

Data were analyzed using R (R Core Team, 2014). A repeated-measures ANOVA was performed with condition (diagnoses / names) and congruency judgment (congruent, incongruent) as within-subjects factors to test for differences in memory (% correctly retrieved associations) as a function of knowledge-relevance (high for diagnoses, low for names) and congruency. A further repeated measures ANOVA was performed to test for differences in reaction time (RT) between the condition. This ANOVA contained the same within-subject factors as before plus the additional within-subject factor memory (remembered, forgotten).

fMRI Data Acquisition and Preprocessing

T2*-weighted echo-planar images were acquired using a 3T Siemens TIM Trio MRI scanner (direction = transverse (interleaved ascending), FOV = 216 mm, TR = 2500 ms, TE = 30 ms, number of slices = 45, slice thickness = 2.5 mm, matrix = 72 x 72, voxel size = 3 x 3 x 2.5 mm, distance factor = 20%, 2 runs with 232 volumes each, including 4 dummy volumes each). To attenuate signal dropout in orbitofrontal regions, the slice orientation was tilted upwards vertically by 15 degrees after alignment to the anterior commissure–posterior commissure plane (Weiskopf, Hutton, Josephs, & Deichmann, 2006). To estimate geometric distortion and signal loss in the EPI, an additional 53-seconds fieldmap was acquired. Structural data was acquired using a T1-weighted 3D magnetization-prepared rapid gradient echo sequence (TR 2500 ms, TE 2500 ms, sagittal orientation, spatial resolution 1 x 1 x 1 mm).

Data were preprocessed and analyzed using FEAT in FSL (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl; Smith, Jenkinson, & Woolrich, 2004). Functional data were corrected for motion (MCFLIRT), slice acquisition times (interleaved), and local field inhomogeneities (BBR / FUGUE), then high-pass filtered (80 Hz), and spatially smoothed using a 5-mm full-width half-maximum Gaussian filter. Data were first coregistered with the structural image and then spatially normalized into a common space (Montreal Neurological Institute (MNI) 152 standard-space 2 mm^3).

fMRI Analyses

Brain Activation

After preprocessing, first-level analyses were conducted using general linear modeling (GLM), separately for individual participants and runs (the two experimental blocks). Regressors were generated by convolving the impulse function related to the onset and length of encoding events with a Gamma hemodynamic response function (5 seconds boxcar function). To explore subsequent memory effects (SMEs, i.e. remembered > forgotten contrasts), encoding trials were sorted according to the retrieval data. The two runs were combined using a within-subject fixed-effects analysis and normalized into MNI space. Across-subjects analyses were carried out using a mixed-effects model in the FLAME framework in FSL. Z-statistic images were thresholded at a voxel-wise threshold of $z > 2.3$, with a FWE-corrected cluster threshold of $p < 0.05$, using FLAME1 in FSL. Based on our a priori hypothesis about differences in the vmPFC, we created an anatomical mask of the vmPFC based on FSL's Harvard-Oxford Cortical Structural Atlas, which consisted of the bilateral frontal medial cortex. In addition, exploratory whole-brain analyses were performed.

Two sets of analyses were performed. For the first set of analyses, three separate GLMs were modeled; one that distinguished high and low knowledge-relevance events, one

that distinguished congruent and incongruent events, and another one that distinguished remembered and forgotten events. The first GLM consisted of separate regressors for remembered and forgotten face–diagnosis pairs (high knowledge-relevance), respectively, as well as for remembered and forgotten face–name pairs (low knowledge-relevance), and a regressor of no interest, which contained all correctly remembered pairs that received a "guess" rating during retrieval. High and low knowledge-relevance events were then contrasted, independent of later memory. The second GLM consisted of remembered and forgotten events that were judged as congruent, remembered and forgotten events that were judged as incongruent, and the "guess" regressor of no interest. Congruent and incongruent events were contrasted, independent of later memory. The third GLM consisted of remembered and forgotten events independent of congruency/relevance and again a "guess" regressor of no interest. Remembered and forgotten events were contrasted to determine SMEs. For the across-subject analyses, we tested whether the vmPFC areas revealed in the first two GLMs (knowledge-relevance and knowledge-congruency, respectively) overlap with the vmPFC cluster identified in the third GLM (SME, remembered > forgotten). We did so by using the clusters found in the first two GLMs as a pre-thresholded mask for the SME analysis.

For the second set of analyses, one GLM was constructed that modeled all nine types of events: remembered congruent diagnoses, forgotten congruent diagnoses, remembered congruent names, forgotten congruent names, remembered incongruent diagnoses, forgotten incongruent diagnoses, remembered incongruent names, forgotten incongruent names forgotten, as well as the "guess" regressor of no interest. For the across-subject analyses, we extracted percent signal change for the eight main events of interest (against implicit baseline) from a vmPFC cluster defined based on the SME analysis of those 24 subjects whose data could only be used for the first set of analyses. This analysis approach was chosen to obtain

an unbiased cluster for the percent signal change analyses (due to difficulties in defining anatomical sub-regions in vmPFC, see Bein, Reggev, & Maril, 2014). The key interest was to directly test for interactions between memory, congruency, and relevance, in particular the significance of two interaction terms: congruency x memory and relevance x memory. Due to the rather low and differing trial counts per cell in this analysis[1], which might lead to differences in signal-to-noise ratio between conditions, we controlled for differences in trial counts by entering trial counts per cell as a covariate in a linear mixed effects analysis. The linear mixed effects analysis allowed us to deal with interdependence given our within-subject design and was performed using *lme4* (Bates, Mächler, Bolker, & Walker, 2015) in R. As fixed effects, we entered congruency, relevance, and memory as interacting regressors into the model, along with number of trials per cell and encoding RTs as covariates. Subjects were entered as random effects into the model. Furthermore, a precursory model that tested for interactions between our covariate and the other regressors revealed a significant memory x trial count interaction (i.e., more remembered trials than forgotten trials, see Footnote 1). Therefore, this interaction term was entered into the analysis as an additional fixed effect to avoid misspecification in the model. To further probe the significance of the main interaction terms of interest (congruency x memory and relevance x memory), likelihood ratio tests were performed comparing the goodness of fit between a model with the critical interaction and a model without this interaction. Statistical significance of the model difference was determined using χ^2 (chi-squared) tests with degrees of freedom equal to the difference in dimensionality of the two models (i.e., 1).

[1] High Relevance Congruent Remembered: 21.6 ± 5.5 (M ± SD); High Relevance Incongruent Remembered: 24.2 ± 6.3; Low Relevance Congruent Remembered: 23.8 ± 7.0; Low Relevance Incongruent Remembered: 14.4 ± 7.0; High Relevance Congruent Forgotten: 8.3 ± 4.1; High Relevance Incongruent Forgotten: 12.6 ± 5.3; Low Relevance Congruent Forgotten: 16.4 ± 6.0; Low Relevance Incongruent Forgotten: 15.2 ± 5.4.

Results

Memory performance

As can be seen in Figure 2, a repeated-measures ANOVA revealed (a) a main effect of knowledge-congruency ($F(1,44) = 46.82$, $p < .001$, $eta^2_G = .10$), indicating better memory performance for face–word pairs judged as congruent as compared to those that were judged as incongruent; (b) a main effect of knowledge-relevance ($F(1,44) = 70.41$, $p < .001$, $eta^2_G = .25$), indicating better memory performance for high relevance (face–diagnosis) as compared to low relevance (face–name) pairs; and (c) no interaction ($F(1,44) = 0.78$, $p = .383$, $eta^2_G = .003$).

Results were highly similar for the subgroup of subjects used for the full factorial analysis (i.e., significant main effects of congruency and relevance, non-significant interaction between the two factors).

We also explored RTs to rule out that any interactions in RT confound the interactions observed in our full factorial fMRI analysis. A repeated-measures ANOVA revealed significant main effects of relevance ($F(1,24) = 175.98$, $p < .001$, $eta^2_G = .41$), indicating faster RTs for the low-relevance condition, and memory ($F(1,24) = 9.55$, $p = .005$, $eta^2_G = .01$), indicating faster RTs for remembered events. No main effect of congruency ($F(1,24) = .84$, $p = .37$, $eta^2_G = .001$) and no significant interactions (all $p > .25$) were observed.

fMRI Results

In the following, we will report results of two sets of analyses. In the first set of analyses, we tested whether the vmPFC distinguishes between associations judged as congruent vs. incongruent and/or associations for which medical knowledge is of high vs. low relevance and whether these areas overlap with vmPFC areas that show a SME. These analyses were performed with the full sample (n = 49). In the second set of analyses, we tested whether the vmPFC involvement in successful memory formation interacts with the vmPFC involvement in knowledge-congruency and/or knowledge-relevance processing. We did so by extracting % signal change from the vmPFC cluster showing a SME and subjecting these data to a within-subject ANOVA. The latter analysis was performed in a subgroup (n = 25) that provided enough (>5) valid trials in each of the 8 conditions.

vmPFC activation as a function of congruency, relevance, and memory

This section reports results from the first set of analyses (n = 49, anatomical vmPFC mask, for exploratory whole-brain results see Table 1). Testing for activation that was greater for the encoding of associations that were judged as congruent as compared to associations judged as incongruent revealed a cluster in the vmPFC (peak voxel: 6, 42, -16; Z = 3.8, 208 voxels, see Figure 3, in green). The opposite contrast, testing for activation that was greater for associations judged as incongruent, revealed no cluster in the vmPFC.

Testing whether the vmPFC was more strongly activated for associations for which the participants' medical knowledge was of high (i.e. face–diagnosis pairs) vs. low (i.e. face–name pairs) relevance revealed activation in a cluster in the vmPFC (peak voxel: -2, 36, -16, Z = 5.01, 121 voxels, see Figure 3, in blue). The opposite contrast, testing for brain regions that expressed higher activation for low relevance associations also revealed activation in a cluster in the vmPFC (peak voxel: 4, 52, -4; Z = 6.26, 190 voxels, see Figure 3, in yellow).

Next, we tested whether the vmPFC contributed to successful memory formation, independent of congruency and relevance. This analysis revealed a large cluster in the vmPFC (peak voxel: -4, 50, -14; Z = 4.6, 396 voxels, see Figure 3 in red; see Table 1 for a complete list of regions that displayed SME). Finally, we sought to test whether this SME cluster overlaps with the clusters that distinguished congruency and relevance, as revealed in the first set of analyses. We tested this by using the latter clusters as a pre-thresholded mask for the SME analysis. These analyses revealed an overlapping cluster with the congruent > incongruent contrast (peak voxel: -4, 48, -14; Z = 4.59, 164 voxels see Figure 3 in green), but not with the high > low relevance or low > high relevance clusters.

These results suggest that the vmPFC is indeed sensitive to differences in knowledge-congruency in that it displays enhanced activation for associations that were judged as congruent. Concerning the vmPFC's sensitivity to differences in knowledge-relevance, results were inconclusive in that neighboring clusters within the vmPFC displayed enhanced activation for both high and low knowledge-relevance associations. Most importantly, however, both of these clusters did not overlap with the cluster exhibiting a SME. In contrast, the vmPFC region that was sensitive to knowledge-congruency overlapped with the SME cluster. This suggests that the vmPFC's involvement in congruency detection might interact with its role in memory formation.

vmPFC contributions to memory formation vary as a function of knowledge-congruency, but not of knowledge-relevance

We extracted percent signal change from a vmPFC SME cluster (peak voxel: -2, 48, -14; Z = 3.13, 236 voxels) that was defined based on those 24 subjects whose data could not be used for the percent signal change analyses. The goal of the percent signal change analyses was to directly test whether the vmPFC involvement in successful memory formation differed between knowledge-congruent and knowledge-incongruent and/or high and low knowledge-

relevance associations. Descriptive results are presented in Figure 4. A linear mixed effects analysis that included trial counts and encoding RTs as covariates revealed a significant congruency x memory interaction ($\chi^2(1) = 5.81$, $p = .016$), but no relevance x memory interaction ($\chi^2(1) = .23$, $p = .64$) and no congruency x relevance x memory interaction ($\chi^2(1) = .56$, $p = .45$). To validate the significance of the detected congruency x memory interaction, we performed an additional likelihood ratio test comparing a model with the congruency x memory interaction with a model without this interaction. This comparison revealed a significant difference between the two models ($\chi^2(1) = 5.70$, $p = .017$), underlining the significance of the congruency x memory interaction. In contrast, comparing models with and without the relevance x memory interaction term revealed no significant effect ($\chi^2(1) = .22$, $p = .636$). Taken together, these findings suggest that the vmPFC contributes more to successful memory formation when perceived congruency is high than when it is low. In contrast, vmPFC's contributions to successful memory formation do not vary as a function of knowledge-relevance.

Discussion

This study tested the hypothesis that vmPFC contributions to successful memory formation vary as a function of knowledge-congruency – being strong when an individual perceives a high fit between activated prior knowledge and new information– but not as a function of knowledge-relevance.

We found evidence for our hypothesis in two sets of analyses. In the first one, we observed that a cluster in the vmPFC displayed stronger activation for associations perceived as congruent compared to associations perceived as incongruent, which suggests that the vmPFC is indeed sensitive to knowledge-congruency. Furthermore, this vmPFC cluster strongly overlapped with a vmPFC cluster that contributed to successful memory formation

(i.e., showed a SME), indicating that the vmPFC's role in congruency detection might interact with its role in memory formation. In the second analysis, we probed this interaction directly using a linear mixed effects analysis on the percent signal change data extracted from the vmPFC SME cluster of those participants whose data were not used for the second analysis. This analysis revealed a significant congruency x memory interaction in the vmPFC. No significant interactions involving the knowledge-relevance factor were found. The latter was true even though memory performance was strongly modulated by knowledge-relevance, which indicates that prior knowledge was indeed useful for memorizing in our high relevance condition. These findings indicate that vmPFC contributions to memory formation differ as a function of knowledge-congruency, but not as a function of knowledge-relevance.

Our results contribute to a better understanding of the role of the vmPFC in memory formation. They suggest that the vmPFC's involvement in memory encoding is not modulated by prior knowledge of the stimulus material per se, but that its contributions are modulated by the perceived congruency between prior knowledge and the to-be-encoded information. These findings emphasize the subjective nature of congruency, which can be high even when overall knowledge-relevance is low (such as when associating names with faces). They also provide empirical support for our claim that knowledge-relevance and knowledge-congruency can be distinguished and might help to explain why a number of published experiments that examined prior knowledge effects on memory encoding have not found vmPFC activation (Bein, Reggev, & Maril, 2014; Brod, Lindenberger, Wagner, & Shing, 2016; van Buuren et al., 2014). All of these studies contrasted high and low knowledge-relevance associations (in the case of Bein et al., 2014, semantically related and unrelated word pairs), which did not involve a congruency dimension. We, thus, propose an amendment to the existing models of the vmPFC's role in memory encoding (Gilboa & Marlatte, 2017; Schlichting & Preston, 2015; van Kesteren et al., 2012). We suggest that the vmPFC's contributions to memory

encoding are dependent on the subjectively perceived congruency between prior knowledge and new information (i.e., stronger when congruency is high), but that they seem not to be dependent on how well the new information can be linked to a pre-existing semantic network. This claim resonates well with the idea of the vmPFC's role in memory retrieval as providing a "feeling of rightness", which was based on work with confabulating patients (Moscovitch & Winocur, 2002). It is also in line with the vmPFC's role in self-referential processing (Macrae, Moran, Heatherton, Banfield, & Kelley, 2004; Northoff & Bermpohl, 2004) and in providing affective value information in decision making, such as the correctness of a prediction (Kumaran, Summerfield, Hassabis, & Maguire, 2009; Roy, Shohamy, & Wager, 2012). All of these different lines of research highlight the subjective dimension of vmPFC recruitment, and we believe that this common role of the vmPFC extends to the memory domain.

Several limitations of our study and of the proposed model revision have to be discussed. First, even though our proposed distinction between knowledge-congruency and knowledge-relevance is able to explain why several recent memory studies have not observed vmPFC involvement despite being knowledge-related, it is challenged by one study that found differential vmPFC involvement although its conditions did not seem to differ in knowledge-congruency. In this study (van Kesteren et al., 2014), students of biology and education had to encode new facts that were related to either biology or education. Successful encoding of facts from their own discipline (i.e. of high knowledge-relevance) led to enhanced vmPFC activation as compared to facts from the other discipline. Although the strength of the activation difference was modest (27 voxels), this finding seems difficult to reconcile with our model. One could speculate that, even though the two conditions did not differ in congruency per se, the participants generally perceived higher congruency for facts related to their own subject as compared to the other one. Evidence for this speculation comes

414 from data of the encoding task, in which the participants had to indicate whether they will
415 remember the fact or not. For their own subject, participants indeed more often expected to
416 remember the new fact as compared to for the other subject (cf. van Kesteren et al., 2014).
417 This points to a more general issue, which is that a congruency decision may also entail a
418 difficulty decision because associations that are easier to encode may be deemed congruent.
419 This leads to a second limitation of our model, which is that knowledge-congruency and
420 knowledge-relevance are often not completely independent. Nevertheless, our data suggest
421 that knowledge-congruency and associated vmPFC activation can be high even though overall
422 knowledge-relevance is low. This suggests that the subjective congruency dimension can be
423 independent of the experimental condition manipulation. A further concern is that the
424 reported lack of a relevance x memory interaction in the vmPFC has to be interpreted with
425 caution due to its null-effect nature. This finding does not preclude the possibility that the
426 vmPFC is sensitive to differences in knowledge-relevance. In fact, two clusters in the vmPFC
427 were sensitive to differences in knowledge-relevance, albeit in opposite directions (i.e.,
428 greater activation for high vs. low in one cluster, and vice versa for the other cluster).
429 Critically, however, their involvement was not predictive of successful memory formation.

430 Future studies are necessary to determine whether making an explicit decision is
431 actually necessary for the vmPFC to be involved. Our study, along with most of the studies
432 reported thus far, included explicit congruency judgments performed by the participants and
433 sorted trials based on these judgments. Knowledge-relevance, on the other hand, was content-
434 based (diagnoses vs. names) and defined by the experimenters. Nevertheless, making a
435 decision that something is congruent could trigger reward-related processes that have been
436 shown to lead to vmPFC activation as well (Rushworth, Noonan, Boorman, Walton, &
437 Behrens, 2011), as has been shown for information rated as self-related (Gutchess, Kensinger,
438 & Schacter, 2007). Thus, it is currently unclear whether a task in which there is a clear

congruency dimension would be enough to trigger vmPFC activation even when the participants are not asked to give a response. Further studies are also needed to determine whether vmPFC contributions to memory encoding differ by sub-region. As an example, a study on memory-based decision-making has reported distinctive contributions of subcallosal vmPFC and posterior orbitofrontal cortex to monitoring and control processes, respectively (Hebscher, Barkan-Abramski, Goldsmith, Aharon-Peretz, & Gilboa, 2016, for a proposal on sub-regional organization of the vmPFC, see Hebscher & Gilboa, 2016).

To conclude, we have shown that the vmPFC contributions to memory encoding differ by knowledge-congruency, but not by knowledge-relevance. We reported evidence for a theoretical distinction according to which the vmPFC is not involved in memory encoding in the context of prior knowledge per se, but that its contributions are modulated by the perceived congruency between prior knowledge and the to-be-encoded information. We believe that this revision to the emerging model of the vmPFC's role in knowledge-based memory encoding can be helpful to advance research in the field because it is easily falsifiable and it allows to derive clear hypotheses about when the vmPFC can be expected to be involved in memory encoding.

Funding

This work was funded by the Max Planck Society and a Heinz Maier Leibnitz prize awarded to YLS by the German Research Foundation. Garvin Brod was supported by a doctoral fellowship of the International Max Planck Research School on the Life Course (LIFE; www.imprs-life.mpg.de), and a scholarship of the German Academic Exchange Service (DAAD).

Acknowledgments

This study was conducted within the "Cognitive and Neural Dynamics of Memory across the Lifespan" (CONMEM) project at the Center for Lifespan Psychology, Max Planck Institute for Human Development. GB was supported by a fellowship of the International Max Planck Research School on the Life Course (LIFE; **www.imprs-life.mpg.de**). We thank Ulman Lindenberger and Anthony Wagner for their comments and support for this project.

References

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. http://doi.org/10.18637/jss.v067.i01

Bein, O., Reggev, N., & Maril, A. (2014). Prior knowledge influences on hippocampus and medial prefrontal cortex interactions in subsequent memory. *Neuropsychologia*, *64*, 320–330. http://doi.org/10.1016/j.neuropsychologia.2014.09.046

Benoit, R. G., Szpunar, K. K., & Schacter, D. L. (2014). Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. *Proceedings of the National Academy of Sciences*, *111*(46), 16550-16555. http://doi.org/10.1073/pnas.1419274111

Berkers, R. M. W. J., van der Linden, M., de Almeida, R. F., Müller, N. C. J., Bovy, L., Dresler, M., … Fernández, G. (2016). Transient medial prefrontal perturbation reduces false memory formation. *Cortex*. http://doi.org/10.1016/j.cortex.2016.12.015

Brod, G., Lindenberger, U., Wagner, A. D., & Shing, Y. L. (2016). Knowledge Acquisition during Exam Preparation Improves Memory and Modulates Memory Formation. *Journal of Neuroscience*, *36*(31), 8103–8111. http://doi.org/10.1523/JNEUROSCI.0045-16.2016

Brod, G., Lindenberger, U., Werkle-Bergner, M., & Shing, Y. L. (2015). Differences in the neural signature of remembering schema-congruent and schema-incongruent events. *NeuroImage*, *117*, 358–366. http://doi.org/10.1016/j.neuroimage.2015.05.086

Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The Influence of Prior Knowledge on Memory: A Developmental Cognitive Neuroscience Perspective. *Frontiers in Behavioral Neuroscience*, *7*(October), 1–13. http://doi.org/10.3389/fnbeh.2013.00139

Cohen, G. (1990). Why is it difficult to put names to faces? *British Journal of Psychology*. *81,* 287–297. http://doi.org/10.1111/j.2044-8295.1990.tb02362.x

Ghosh, V. E., Moscovitch, M., Melo Colella, B., & Gilboa, A. (2014). Schema Representation in Patients with Ventromedial PFC Lesions. *Journal of Neuroscience*, *34*(36), 12057–12070. http://doi.org/10.1523/JNEUROSCI.0740-14.2014

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. http://doi.org/10.1016/j.tics.2017.04.013

Hebscher, M., Barkan-Abramski, M., Goldsmith, M., Aharon-Peretz, J., & Gilboa, A. (2016). Memory, decision-making, and the ventromedial prefrontal cortex (vmPFC): the roles of subcallosal and posterior orbitofrontal cortices in monitoring and control processes. *Cerebral Cortex*, *26*(12), 4590-4601. https://doi.org/10.1093/cercor/bhv220

Hebscher, M., & Gilboa, A. (2016). A boost of confidence: The role of the ventromedial prefrontal cortex in memory, decision-making, and schemas. *Neuropsychologia*, *90*, 46-58. https://doi.org/10.1016/j.neuropsychologia.2016.05.003

Kumaran, D., Summerfield, J. J., Hassabis, D., & Maguire, E. a. (2009). Tracking the emergence of conceptual knowledge during human decision making. *Neuron*, *63*(6), 889–901. http://doi.org/10.1016/j.neuron.2009.07.030

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. http://doi.org/10.3758/BF03210951

Macrae, C. N., Moran, J. M., Heatherton, T. F., Banfield, J. F., & Kelley, W. M. (2004). Medial prefrontal activity predicts memory for self. *Cerebral Cortex*, *14*(6), 647–654. http://doi.org/10.1093/cercor/bhh025

Maril, A., Avital, R., Reggev, N., Zuckerman, M., Sadeh, T., Sira, L. Ben, & Livneh, N. (2011). Event congruency and episodic encoding: A developmental fMRI study. *Neuropsychologia*, *49*(11), 3036–3045.

http://doi.org/10.1016/j.neuropsychologia.2011.07.004
McWeeny, K. H., Young, A. W., Hay, D. C., & Ellis, A. W. (1987). Putting names to faces. *British Journal of Psychology*, *78*, 143–149. http://doi.org/10.1111/j.2044-8295.1987.tb02235.x
Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630–633. http://doi.org/10.3758/BF03206543
Moscovitch, M. (1989). Confabulation and the Frontal Systems: Strategic versus Associative Retrieval in Neuropsychological Theories of Memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: essays in honour of Endel Tulving* (pp. 133–161). Hillsdale, NJ: Lawrence Erlbaum.
Moscovitch, M., & Melo, B. (1997). Strategic retrieval and the frontal lobes: Evidence from confabulation and amnesia. *Neuropsychologia*, *35*(7), 1017–1034. http://doi.org/10.1016/S0028-3932(97)00028-6
Moscovitch, M., & Winocur, G. (2002). The frontal cortex and working with memory. In D. T. Stuss & R. T. Knight (Eds.), *Principles of Frontal Lobe Function* (pp. 188–209). New York, NY: Oxford University Press.
Nieuwenhuis, I. L. C., & Takashima, A. (2011). The role of the ventromedial prefrontal cortex in memory consolidation. *Behavioural Brain Research*, *218*(2), 325–334. http://doi.org/10.1016/j.bbr.2010.12.009
Northoff, G., & Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences*, *8*(3), 102–107. http://doi.org/10.1016/j.tics.2004.01.004
R Core Team. (2014). R: A Language and Environment for statitical computing. Vienna, Austria: R Foundation for Statistical Computing.
Reggev, N., Bein, O., & Maril, A. (2016). Distinct neural suppression and encoding effects for conceptual novelty and familiarity. *Journal of cognitive neuroscience, 28*(10). 1455–1470. http://doi.org/10.1162/jocn_a_00994
Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*(3), 147–156. http://doi.org/10.1016/j.tics.2012.01.005
Rushworth, M. F. S., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron*, *70*(6), 1054–1069. http://doi.org/10.1016/j.neuron.2011.05.014
Schlichting, M. L., & Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8. http://doi.org/10.1016/j.cobeha.2014.07.005
Schlichting, M. L., & Preston, A. R. (2016). Hippocampal–medial prefrontal circuit supports memory updating during learning and post-encoding rest. *Neurobiology of Learning and Memory*, *134*(Part A), 91–106. http://doi.org/10.1016/j.nlm.2015.11.005
Smith, S., Jenkinson, M., & Woolrich, M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23*, 208–219.
Spalding, K. N., Jones, S. H., Duff, M. C., Tranel, D., & Warren, D. E. (2015). Investigating the Neural Correlates of Schemas: Ventromedial Prefrontal Cortex Is Necessary for Normal Schematic Influence on Memory. *Journal of Neuroscience*, *35*(47), 15746–15751. http://doi.org/10.1523/JNEUROSCI.2767-15.2015
van Buuren, M., Kroes, M. C. W., Wagner, I. C., Genzel, L., Morris, R. G. M., & Fernández, G. (2014). Initial investigation of the effects of an experimentally learned schema on spatial associative memory in humans. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(50), 16662–70. http://doi.org/10.1523/JNEUROSCI.2365-14.2014

van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia*, *51*(12), 2352–9. http://doi.org/10.1016/j.neuropsychologia.2013.05.027

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(47), 15888–94. http://doi.org/10.1523/JNEUROSCI.2674-10.2010

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., Morris, R. G. M., & Fernández, G. (2014). Building on prior knowledge: Schema-dependent encoding processes relate to academic performance. *Journal of Cognitive Neuroscience*, *26*(10), 2250–2261. http://doi.org/10.1162/jocn_a_00630

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. http://doi.org/10.1016/j.tins.2012.02.001

Warren, D. E., Jones, S. H., Duff, M. C., & Tranel, D. (2014). False Recall Is Reduced by Damage to the Ventromedial Prefrontal Cortex: Implications for Understanding the Neural Correlates of Schematic Memory. *Journal of Neuroscience*, *34*(22), 7677–7682. http://doi.org/10.1523/JNEUROSCI.0119-14.2014

Webb, C. E., Turney, I. C., & Dennis, N. A. (2016). What's the gist? The influence of schemas on the neural correlates underlying true and false memories. *Neuropsychologia*, *93*(September), 61–75. http://doi.org/10.1016/j.neuropsychologia.2016.09.023

Weiskopf, N., Hutton, C., Josephs, O., & Deichmann, R. (2006). Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3 T and 1.5 T. *Neuroimage*, *33*(2), 493–504.

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168–79. http://doi.org/10.1016/j.neuron.2012.05.010

Tables

Table 1. Regions exhibiting stronger activation for high vs. low and low vs. high knowledge-relevance pairs as well as for subsequently remembered vs. forgotten pairs. To better capture the involved brain regions, local maxima are presented in addition to cluster maxima for very large clusters.

Region	x	y	z	Z-Max	# voxels
High vs. Low Knowledge-Relevance					
Left Inferior Temporal Gyrus	-46	-54	-16	8.23	42790
Left Temporooccipital Fusiform Cortex	-40	-46	-18	8.18	"
Left Lateral Occipital Cortex	-48	-68	-14	8.11	"
Left Inferior Temporal Gyrus	-42	-52	-14	6.72	"
Left Superior Frontal Gyrus	-54	-52	-12	8.07	"
Left Inferior Temporal Gyrus	-52	-56	-12	7.99	"
Paracingulate Gyrus / Superior Frontal Gyrus	-6	16	48	8.34	3629
Insular Cortex	32	26	2	7.38	717
Right Middle / Inferior Frontal Gyrus	48	14	32	4.6	667
Low vs. High Knowledge-Relevance					
Right Supramarginal / Angular Gyrus	60	-42	38	7.24	50504
Paracingulate Gyrus	2	48	2	3,31	"
Right Supramarginal Gyrus	54	-40	30	7.17	"
Cingulate Gyrus	-2	38	6	6.99	"
Right Supramarginal Gyrus	62	-32	36	6.89	"
Cingulate Gyrus	-2	36	12	6.83	"
Subsequent Memory Effect (Rem > Forg)					
Right Lateral Occipital Cortex	42	-72	-6	4.65	4093
Left Temporooccipital Fusiform Cortex	-40	-56	-14	4.61	2906
Left Inferior Frontal Gyrus / Frontal Pole	-54	32	14	4.86	2715
Frontal Pole	-8	54	42	4.62	2560
Left Amygdala / Hippocampus	-18	-6	-14	4.9	1009
Left Lateral Occipital Cortex	-48	-70	36	4.11	796
Right Amygdala / Hippocampus	20	-6	-16	4.77	630
Bilateral Ventromedial Prefrontal Cortex	-4	50	-14	4.61	575
Right Inferior Frontal Gyrus	56	34	12	3.96	571
Congruent vs. Incongruent					
Bilateral Ventromedial Prefrontal Cortex	2	62	16	4.19	580
Bilateral Caudate	-8	16	0	4.35	401
Incongruent vs. Congruent					
Right Middle Frontal Gyrus	48	28	36	3.81	555

Figures

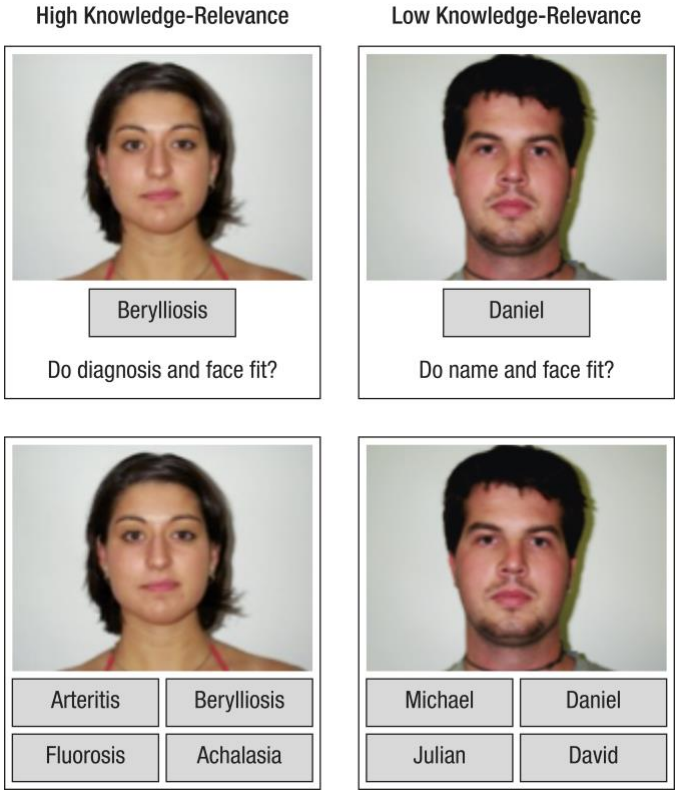

Figure 1. Memory task. Participants were instructed to memorize face–word pairs in the MRI (upper part) and to indicate whether the face fits the word or not (congruency judgment). Half of the words were diagnoses (high knowledge-relevance, left example) and half were first names (low knowledge-relevance, right example). Retrieval took place outside of the scanner (lower part). All of the studied faces were presented again, together with four first names or four diagnoses, of which only one had been presented with the face during the encoding phase. Participants had to indicate the word with which the face was presented during encoding. The three lures were names or diagnoses that had been paired with other faces during the encoding phase.

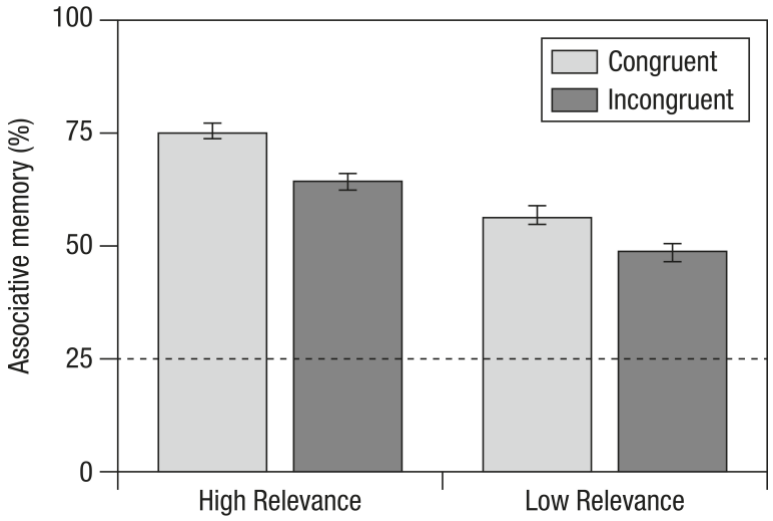

Figure 2. Memory performance was higher for associations that were rated as congruent and that had high knowledge-relevance (i.e., face–diagnosis pairs), with no interaction between congruency and relevance. Chance level was 25%. Error bars are within-subject standard errors (Loftus & Masson, 1994).

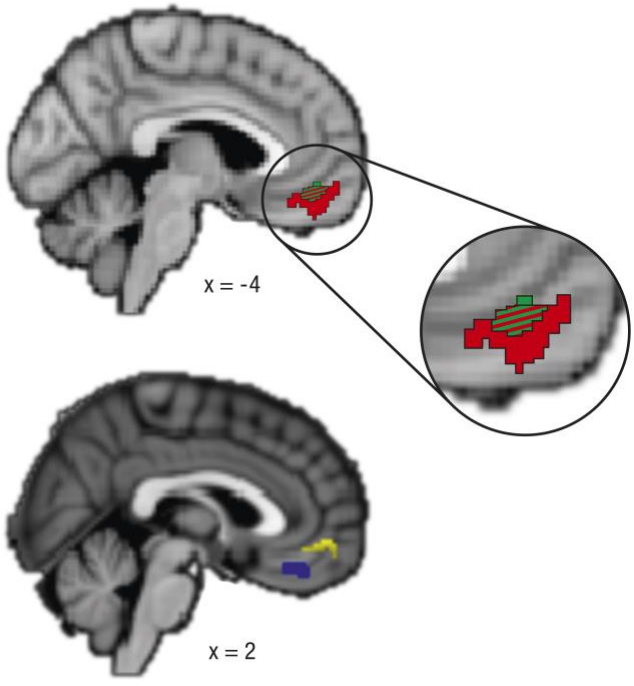

Figure 3. Effects of memory, congruency, and relevance within our vmPFC anatomical mask. Upper part: the vmPFC was more strongly activated for associations that were judged as

congruent as compared to associations judged as incongruent (peak voxel: 6, 42, -16; Z = 3.8, 208 voxels, in green). This cluster overlaps (overlap = 208 voxels, striped) with the vmPFC cluster distinguishing associations that were later remembered vs. forgotten (i.e. SME) (peak voxel: -4, 50, -14; Z = 4.6, 396 voxels, in red). Lower part: Nearby regions of the vmPFC displayed more activation for associations for which the participants' medical knowledge was of high vs. low relevance (peak voxel: -2, 36, -16, Z = 5.01, 121 voxels, in blue) and of low vs. high relevance (peak voxel: 4, 52, -4; Z = 6.26, 190 voxels, in yellow).

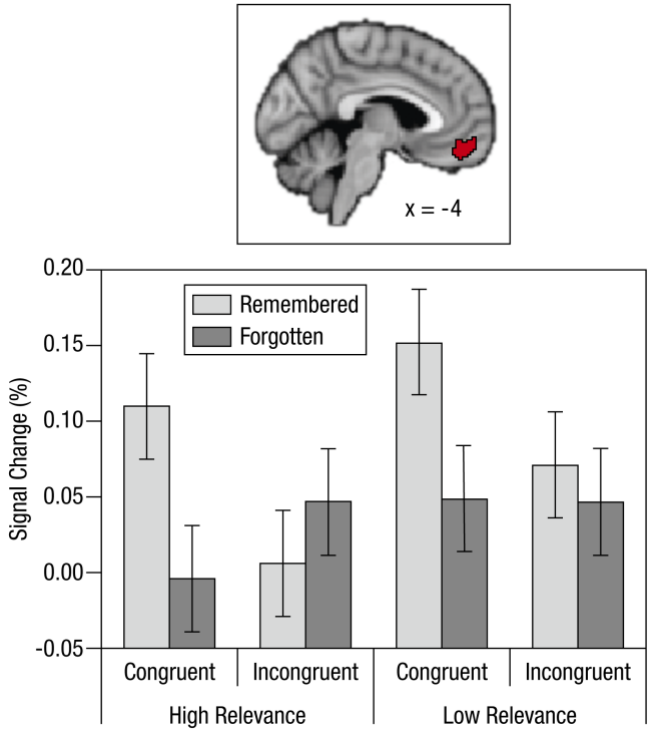

Figure 4. Congruency x memory interaction in the vmPFC. Signal change (%) was extracted from a vmPFC SME cluster (peak voxel: -2, 48, -14; Z = 3.13, 236 voxels, in red) that was defined in an independent sample. A linear mixed effects analysis revealed a significant congruency x memory interaction ($\chi^2(1) = 5.81$, $p = .016$), but no relevance x memory interaction ($\chi^2(1) = .23$, $p = .64$). Error bars are within-subject standard errors (Loftus & Masson, 1994).