NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 33 (2006) 1188 – 1193

Neural correlates of memory confidence

Steffen Moritz,[a,*,1] Jan Gläscher,[b,1] Tobias Sommer,[b] Christian Büchel,[b] and Dieter F. Braus[c]

[a]*Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany*
[b]*Institute for Systems Neuroscience, NeuroImage Nord, University Medical Center Hamburg-Eppendorf, Germany*
[c]*Division of Psychiatry, NeuroImage Nord, University Medical Center Hamburg-Eppendorf, Germany*

Received 23 March 2006; revised 3 August 2006; accepted 4 August 2006
Available online 9 October 2006

The present study aimed to shed light on the neural underpinnings of high vs. low memory confidence. To dissociate memory confidence from accuracy, the Deese–Roediger McDermott (DRM) paradigm was employed, which – compared to other memory paradigms – elicits a rather evenly distributed number of high-confident responses across all possible combinations of memory response types (i.e., hits, false alarms, correct rejections, and misses). In the standard DRM procedure, subjects are first presented with thematically interrelated word lists at encoding, which at recognition are intermixed with related and unrelated distractor items. The signature of a false memory or DRM effect is an increased number of high-confident false memories, particularly for strongly related lure items. For the present study, 17 female subjects were administered a verbal DRM task, whereas neural activation was indexed by fMRI. The behavioral analyses confirmed the expected false memory effect: subjects made more high-confident old responses (both hits and false alarms) the closer the items were related to the central list theme. Across all four memory response types, an increase in confidence at recognition was associated with bilateral activation in the anterior and posterior cingulate cortex along with medial temporal regions. In contrast, increments in doubt were solely related to activation in the superior posterior parietal cortex. To conclude, the study provides some evidence for dissociable systems for confidence and doubt.

Introduction

Memory retrieval is by no means an all-or-nothing mechanism with a memory episode either being successfully recollected or not (Koriat et al., 2000). Rather, memory recollection is modulated by the degree of subjective confidence that an event or stimulus has been encountered previously. Thus, memory confidence serves as an important adaptive cognitive tool (Koriat and Goldsmith, 1996; Koriat et al., 2001): whereas conviction facilitates decisive actions, doubt cautions a subject to withhold a response and to prolong the search process. Notwithstanding that memory confidence is not an optimal indicator for memory accuracy, numerous studies have confirmed that correct responses are usually accompanied by higher confidence ratings than are incorrect ones (see Keren, 1991; Moritz et al., 2003b, 2005). As a consequence, the impact of a correct selection is enhanced, whereas incorrect responses receive a "not trustworthy" tag thereby attenuating potential consequences of a wrong decision. If memory retrieval was just a binary all-or-nothing process (i.e., unmodulated by confidence), correct and incorrect responses would receive the same weight, which in the case of errors might have severe negative implications.

The investigation of memory confidence has been stimulated by research on eye-witness testimony and psychiatric disorders. For example, patients with obsessive–compulsive disorder (OCD) and schizophrenia (Koren et al., 2005; Moritz et al., 2005) display disruptions in the assessment of memory confidence. OCD patients appear to have a decreased memory confidence (Zitterl et al., 2001) despite rather uncompromised memory accuracy (Moritz et al., 2003a). Conversely, schizophrenia patients have been repeatedly found to be over-confident in memory errors while being at the same time under-confident in correct responses (Moritz et al., 2003b, 2005).

The neural pattern underlying memory confidence is yet poorly understood (Chua et al., 2006). Henson and coworkers (2000) detected activation in several prefrontal and parietal regions when subjects made low-confident vs. high-confident responses. No significant results were reported for the reverse analysis. In contrast, a recent study (Chua et al., 2006) on novel face recognition found no activation for low-confident vs. high-confident responses, whereas the opposite contrast was associated with activation of the anterior as well as posterior cingulate and medial temporal lobe. In general, regions associated with high confidence judgments mapped anatomically with limbic structures ("circuit of Papez"). Studies on *remember–know* judgments (i.e., vivid recollection vs. familiarity) are also relevant to this aspect of metamemory because *remember* judgments unlike *know* judgments are usually accompanied by high-confident responses (Moritz and Woodward, 2006; Yonelinas, 2001). Recollection, as measured by *remember* or source judgments, has been linked with activation in the posterior cingulate as well as medial temporal regions (Eldridge et al., 2000; Henson et

* Corresponding author. Fax: +49 40 42803 6565.
E-mail address: moritz@uke.uni-hamburg.de (S. Moritz).
[1] Both first authors have equally contributed in conducting the study.
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.08.003

al., 1999; Sommer et al., 2005; Yonelinas et al., 2005). More lateral areas, including the anterior and dorsolateral prefrontal cortex, have been recently linked to increases in familiarity confidence (Yonelinas et al., 2005). Feeling of knowing (FOK) represents another index of metamemory, which is defined as the feeling that one has some information in memory that is currently not retrievable but could be recollected either at a later time-point or when provided with cues. FOK has been linked to the prefrontal cortex (Kikyo et al., 2002; Schnyer et al., 2004, 2005), although there is some preliminary evidence for involvement of the parietal cortex as well (Maril et al., 2005).

Taken together, although the current literature strongly suggests an involvement of the posterior cingulate cortex as well as medial temporal areas in the modulation of memory confidence, a solid cortical signature of metamemory processes remains to be established. Importantly, studies differ whether or not the hippocampus and its adjacent cortices, whose involvement in episodic memory retrieval are undisputed (Squire, 1992), are also engaged in metamemory.

The present study explores memory confidence for different classes of memory responses (i.e., hits, false alarms, correct rejections, and misses). In particular, it was investigated whether memory confidence and doubt are represented in anatomically separable regions (see also Yonelinas et al., 2005): memory confidence could be modeled either as a single process (inhibition vs. excitation of the same cortical areas resulting in confidence vs. doubt) or independent processes. To meet the study purpose, the Deese–Roediger McDermott (DRM) paradigm (Deese, 1959; Roediger and McDermott, 1995) was administered. In the basic DRM procedure, lists of words (e.g., *hill*, *climb*, *valley*, *summit*, *top*, *molehill*, *peak*, *plain*, *glacier*, *goat*, *bike*, *climber*, *range*, *steep*) are consecutively presented to the participant, each converging on a so-called critical lure item (e.g., *mountain*). It has been shown (Deese, 1959; Roediger and McDermott, 1995) that healthy participants frequently falsely remember the semantically related lures to a large degree (50–80%) in a later recall or recognition trial. The DRM paradigm elicits a high number of high-confident as well as low-confident responses across the entire range of memory responses, whereas other memory paradigms typically produce only few false memories, thereby complicating the separation between accuracy with metamemory (confidence ratings).

Methods

Subjects

Seventeen right-handed healthy female participants took part in the investigation (mean age: 27.41 (SD: 7.51), range: 20–47 years). We selected a homogeneous sample with respect to gender and handedness because slight differences in brain activation for females vs. males may have added noise to the data (Cahill, 2006). Participants did not suffer from any neurological or psychiatric disorders as evidenced by a short interview. The entire session including practice trial, scanning period, and final assessment lasted approximately 2–3 h. Participants received a honorarium of 8€/h. Ethics approval was obtained from the local ethics committee.

Materials

The stimuli for the present study were derived in the course of a staged process. At first, a norming study was conducted, for which 55 healthy participants, none of whom took part in the later experiment, were requested to produce up to ten spontaneous associations for an entire set of 114 theme words. Subsequently, for each of the theme words, lists of 16 items (including the theme word) were compiled in descending order with respect to their response frequency. For example, the second word in the list (i.e., the item following the theme word) was the word most often produced in the association study. If list words shared the same association frequency, the final sequence was determined by the first author. Finally, eighteen lists were chosen which were later divided into three blocks comprising 6 lists each. The main selection criteria for word lists were minimal semantic overlap between lists (no shared associations) and suitability according to experts' opinion (e.g., good backward and forward associative strength).

To compile the items for the encoding and recognition phase, the word lists (excluding the first [theme] item) were divided into three groups of five words each (weakly related, medium related, and strongly related to the theme word). From each group one word was taken out to serve as a lure item in the recognition list. The remaining 12 words were shown during encoding. The theme word served as the so-called critical lure item. In the recognition phase all 16 items per list (12 studied, 4 non-studied items) were visually presented. Further, 12 recognition items were created per block, which were unrelated to any of the list words. Thus, the recognition list for each block (i.e., six lists) consisted of 72 old words (for each list 12 items were created: four strongly, four medium and four weakly related words of all six lists) and 36 new words (each one for every word list of the following types: critical lure, strongly related lure, medium related lure, weakly related lure as well as 12 unrelated new words).

fMRI experiment

For the fMRI experiment, an event-related design was employed which was administered in three blocks. Each block consisted of an encoding and a recognition phase, which were scanned in separate runs. During encoding, words from the six lists were visually presented each for 3 s, whereby lists were displayed in random order. As noted, each list contained 12 stimuli that were presented in descending semantic relatedness to the list theme. In order to ensure semantic processing of the stimuli, participants were asked to indicate whether each item was a noun or not via a key-press with their index or middle finger during encoding. Subjects were instructed that their recognition memory would be tested afterwards. The presentation of lists was separated by a pause of 10 s.

During the recognition phase, items were visually presented above a 6-point Likert scale: for each item, the subject was requested to move a red rectangle located at a random position to one of the six response alternatives (1 = 100% confident old, 2 = rather confident old, 3 = guessing old, 4 = guessing new, 5 = rather confident new, 6 = 100% confident new). In addition, we implemented a "response loop", so that subjects could switch between extreme response options (i.e., 1 and 6) with one button press only. This manipulation ensured that confidence ratings were not confounded with the number of button presses. Subjects were instructed that their responses only referred to the learning items that immediately preceded the recognition phase. For each recognition item, subjects were provided a response window of 4 s with the final position serving as response (i.e., no further