Elevated Cortisol at Retrieval Suppresses False Memories in Parallel with Correct Memories

Susanne Diekelmann\textsuperscript{1}, Ines Wilhelm\textsuperscript{1}, Ullrich Wagner\textsuperscript{1,2}, and Jan Born\textsuperscript{1}

Abstract

Retrieving a memory is a reconstructive process in which encoded representations can be changed and distorted. This process sometimes leads to the generation of "false memories," that is, when people remember events that, in fact, never happened. Such false memories typically represent a kind of "gist" being extracted from single encountered events. The stress hormone cortisol is known to substantially impair memory retrieval. Here, in a double-blind, placebo-controlled crossover design, we tested the effect of an intravenous cortisol infusion before retrieval testing on the occurrence of false memories and on recall of correct memories using a modified Deese–Roediger–McDermott (DRM) paradigm. Subjects studied sets of abstract shapes, with each set being derived from one prototype that was not presented during learning. At retrieval taking place 9 hr after learning, subjects were presented with studied shapes, nonstudied shapes, and the prototypes, and had to indicate whether or not each shape had been presented at learning. Cortisol administration distinctly reduced susceptibility to false memories (i.e., false recognition of prototypes) and, in parallel, impaired retrieval of correct memories (i.e., correct recognition of studied shapes). Response bias as well as confidence ratings and remember/know/guess judgments were not affected. Our results support gist-based theories of false memory generation, assuming a simultaneous storage of the gist and specific details of an event. Cortisol, by a general impairing influence on retrieval operations, decreases, in parallel, retrieval of false (i.e., gist) and correct (i.e., specific) memories for the event.

INTRODUCTION

Memory is not a literal record of the world but can be changed by newly acquired information and preexisting knowledge in the brain (Schacter, Norman, & Koutstal, 1998; Loftus, Feldman, & Dashiel, 1995; Bartlett, 1932). One of the most intriguing features of human memory is the ability to extract the general meaning or the "gist" of single encountered events in an adaptive process. After learning single exemplars of one category, people subsequently vividly remembered having encountered the prototype of that category, which actually was never presented during learning (i.e., they generated a false memory; Roediger & McDermott, 1995; Bransford & Franks, 1971; Posner & Keele, 1968). Generally, the term "false memory" refers to instances in which people claim to remember events that, in fact, never happened. A classical approach to the study of false memory generation is the Deese–Roediger–McDermott (DRM) paradigm, in which subjects learn lists of highly associated words like "white," "night," "cat," and "dark," and so forth, while the common theme or gist word of the list, in this example "black," is not presented during learning (Roediger & McDermott, 1995; Deese, 1959). On a later retrieval test, participants frequently and with high confidence falsely remember having encountered the gist word (Roediger, McDermott, Pisoni, & Gallo, 2004; Seamon et al., 2002; Thapar & McDermott, 2001; Toglia, Neuschatz, & Goodwin, 1999; McDermott, 1996; Payne, Elie, Blackwell, & Neuschatz, 1996). Such "false" memories are highly undesirable in situations where it is essential to rely on veridical memory, for instance, in eyewitness testimony (Loftus, 2003), but it can also be useful to remember the gist or general concept of what had been experienced instead of specific details. It has been proposed that false memories might be unwanted by-products of the human memory system that acts to adaptively change memory representations and extract general knowledge, that is, the gist, from single learned exemplars (Schacter, 1999). Along this line, so-called gist-based (or schema-based) views of false memory generation assume that all the single learned exemplars share common features with the gist whereby the gist, through the overlapping features of the encoded exemplars, becomes simultaneously activated and stored in parallel with the actually encoded exemplars, with both types of memories sharing common mechanisms (Brainerd & Reyna, 2001; Reyna & Brainerd, 1998; McClelland, McNaughton, & O’Reilly, 1995; Bransford & Franks, 1971; Posner & Keele, 1968). However, evidence for a common

\textsuperscript{1}University of Lübeck, Lübeck, Germany, \textsuperscript{2}Bangor University, Gwynedd, UK

© 2010 Massachusetts Institute of Technology

Journal of Cognitive Neuroscience X.Y, pp. 1–10
neurophysiological mechanism producing both false and veridical memories is scarce.

In fact, an influential alternative view on false memory generation (i.e., the so-called monitoring theories) holds that the memory for the gist word is generated already in the phase of encoding, which consequently produces a sense of familiarity when the gist word is presented at retrieval testing (Marsh & Bower, 2004; Gallo & Roediger, 2002). In this view, false memories result from a failure in retrieval monitoring, when the subject mistakenly attributes this sense of familiarity for having encountered the gist word at learning (Mitchell & Johnson, 2009; Johnson, Hashtroudi, & Lindsay, 1993). Here, to examine the different views on false memory generation, we used a pharmacological approach, that is, we administered cortisol, which is well known to reliably impair the retrieval of correct memories (Kuhlmann, Kirschbaum, & Wolf, 2005; de Quervain et al., 2005; Wolf et al., 2001; de Quervain, Roozendaal, Nitsch, McGaugh, & Hock, 2000). If false memories are due to an impaired retrieval monitoring, a cortisol-induced impairment of retrieval should enhance false memory generation, whereas reduced rates of false memory generation are expected under cortisol if, as assumed by gist-based theories, false memories are formed as a by-product of correct memories basically relying on similar neurophysiological mechanisms.

There have been several attempts to characterize the effects of cortisol on false memory generation on the basis of the investigation of stress-induced release of glucocorticoids. Smeets, Jelicic, and Merckelbach (2006), as well as Payne, Nadel, Allen, Thomas, and Jacobs (2002), introduced psychosocial stress, namely, the Trier Social Stress Test (Kirschbaum, Pirke, & Hellhammer, 1993), before encoding of DRM word lists, and retrieval was tested immediately thereafter. Whereas the occurrence of false memories was significantly increased in the Payne et al. (2002) study, Smeets, Jelicic, and Merckelbach (2006) found no change in false memory rate following stress. Yet, both studies applied stress already before encoding, which prevents a clear-cut dissociation of effects of cortisol on retrieval from those on encoding. Unlike retrieval, encoding can even be enhanced by cortisol (Abercrombie, Kalin, Thurow, Rosenkranz, & Davidson, 2003; Rimmle, Domes, Mathiak, & Hautzinger, 2003; Buchanan & Lovallo, 2001). In a recent study, Smeets, Otgaar, Candel, and Wolf (2008) introduced a physiological stressor (i.e., cold pressor stress), either before encoding, immediately after encoding (i.e., before consolidation), or before retrieval testing of DRM word lists, and found no significant effects in neither condition but even slightly decreased false memory rates. Yet, the pharmacological interpretation of these findings is still difficult because stress, beyond stimulating cortisol release, concurrently affects numerous other processes that are implicated in memory functions, such as sympatho-adrenal responses, mood, and different cognitive functions, and thus, possibly confounds the effects of cortisol (Ulrich-Lai & Herman, 2009; Lupien, Maheu, Tu, Fiocco, & Schramek, 2007; de Kloet, Joels, & Holsboer, 2005).

In the present study, we were interested in dissociating the effect of cortisol on the generation of false memories compared to correct memories at retrieval in order to determine whether false memories compared with correct memories would be affected by cortisol in the same or opposite way. We directly administered cortisol (vs. placebo) to exclude possible concurrent effects of stress on false memory formation. Substance administration took place shortly before testing retrieval and more than 7 hr after encoding to exclude effects of cortisol on encoding or consolidation.

**METHODS**

**Participants**

Twelve healthy male subjects [age = 22.67 ± 2.90 years (mean ± SD), range = 18–29 years] were recruited at the University of Lübeck to participate in the study. All subjects were nonsmokers, free of any medication, and had no history of neurological, psychiatric, or endocrine disorders. They were not allowed to ingest any caffeine or alcohol on the days of the experiments. All subjects gave written informed consent and were paid for participation in the study, which was approved by the local ethics committee of the University of Lübeck.

**Design and Procedure**

All subjects participated in two treatment conditions (cortisol and placebo) according to a double-blind crossover design, with the order of conditions balanced across subjects (Figure 1A). The two treatment conditions for each participant were separated by at least 2 weeks. Each condition started at 1030 h, with the subject learning the false memory task. Following learning, the participants left the laboratory to engage in their everyday activities and returned at 1500 h. Two venous catheters were then placed into the subject’s forearms for substance administration. Subjects were instructed not to take any naps during the retention interval in order to exclude possible effects of sleep on the generation of false memories (Diekelmann, Born, & Wagner, in press; Payne...
et al., 2009). Adherence to this instruction was ensured by a postexperimental questionnaire and constant supervision by the experimenter between 1500 and 2000 h.

**False Memory Task**

To induce false memories, we used a nonverbal version of the DRM paradigm (Roediger & McDermott, 1995; Deese, 1959), which was developed by Slotnick and Schacter (2004) (for a detailed description of the materials, see Slotnick & Schacter, 2004). We applied the nonverbal version instead of the standard verbal DRM paradigm to prevent subjects from using deliberate mnemonic strategies to memorize the learning material, because such strategies could affect not only the correct retrieval of memory but also the generation of false memories (McCabe & Smith, 2006; Libby & Neisser, 2001). The materials comprised two parallel versions that were counterbalanced across subjects and treatment conditions. During learning, subjects studied 16 sets of abstract shapes, each set consisting of 10 shapes (2 shown here) derived from one prototype that was not presented during learning. At retrieval testing (right), subjects were presented with studied shapes, new shapes, and prototypes, and had to indicate for each shape whether or not it had been presented during learning.

Blood Sampling and Biochemical Analyses

Blood was sampled every 30 to 45 min between 1630 h and 2100 h (Figure 2). Hormone concentrations were determined in serum samples that were stored at −80°C until assay. Serum cortisol concentrations were assessed via Immulite [DPC Biemann, Bad Nauheim, Germany, intra- and interassay coefficients of variation (CV) < 10%]. Additionally, adrenocorticotropic hormone (ACTH) was assessed in plasma via Lumitest [Brahms Diagnostica, Henningsdorf, Germany, interassay CV 2.8%, intra-assay CV 1.6%]. ACTH
Figure 2. Recognition performance. (A) Enhanced cortisol levels at retrieval reduced the occurrence of false memories (old responses to prototypes) and impaired correct memory retrieval (old responses to studied shapes; means ± SEM are shown). Individual recognition data for each subject in the placebo and cortisol conditions are depicted for false memories (upper left panel) and correct memories (upper right panel), respectively. Cortisol enhancement did not affect the overall response bias. *p < .05. (B) The occurrence of false memories was correlated with correct memory retrieval following cortisol infusion (filled dots) as well as under placebo conditions (empty dots; r = .54 in the cortisol condition (lower regression line) and .64 in the placebo condition (upper regression line)).
(released from the pituitary) is the major secretagogue of adrenal cortisol and becomes suppressed as a consequence of the inhibitory feedback influence that cortisol exerts on the hypothalamo-pituitary-adrenal system.

**Statistical Analysis**

Memory results were analyzed using the signal detection measure $d'$ as a bias-corrected measure of recognition performance in order to take into account interindividual differences in the baseline propensity to accept items (Snodgrass & Corwin, 1988). The calculation of $d'$ was based on three recognition memory measures: hit rate = old responses to studied shapes; false alarm rate = old responses to new shapes; and false memory rate = old responses to prototypes. We calculated $d'$ separately for correct memories [$z$(hit rate) $- z$(false alarm rate)] and false memories [$z$(false memory rate) $- z$(false alarm rate)]. In this procedure, false memory rate is treated as a “hit rate” in order to provide a false memory measure corrected for response bias (Seamon et al., 2002). Additionally, to exclude that cortisol administration affected response bias per se, we calculated the bias index $C$, again separately for correct memories [$-0.5 z$(hit rate) $+ z$(false alarm rate)] and false memories [$-0.5 z$(false memory rate) $+ z$(false alarm rate)].

Statistical analyses were performed using ANOVAs for repeated measures with the factors “cortisol/placebo” and “false/correct memories” for the memory measures. As the ability to remember whether shapes were presented at the left or right side of the screen (i.e., spatial source memory) did not significantly differ between the cortisol and placebo conditions for false memories ($63.48 \pm 5.88\%$ vs. $68.54 \pm 5.18\%$, $p > .40$) or correct memories ($64.55 \pm 4.88\%$ vs. $73.05 \pm 1.99\%$, $p = .10$; chance level = 50%), “old left” and “old right” responses were collapsed for the present analyses. ANOVA performed on cortisol and ACTH concentrations included a “time” factor in addition to the “cortisol/placebo” factor. Post hoc $t$ tests were used to specify significant main effects and interactions. The Greenhouse–Geisser correction of degrees of freedom was applied where appropriate. The level of significance was set to $p = .05$. For analyses of confidence ratings and remember/know/guess judgments, multiple $t$ tests were applied with the level of significance adjusted according to the Bonferroni correction.

**RESULTS**

**False Memories and Correct Memories**

Cortisol infusion before retrieval testing profoundly reduced the susceptibility to false memories. The probability of falsely recognizing nonstudied prototypes was $0.87 \pm 0.10$ in the placebo condition and $0.49 \pm 0.13$ following cortisol treatment ($t(11) = -2.81, p = .017$; Figure 2A). Thus, false memories were reduced by 44% when cortisol was enhanced during retrieval testing compared to placebo.

As expected from previous reports, retrieval of correct memories (of studied items) was also reduced when cortisol was infused before retrieval testing. The correct recognition score was $0.72 \pm 0.08$ in the placebo condition and $0.40 \pm 0.14$ after cortisol administration ($t(11) = -2.44, p = .033$; Figure 2A), resulting in a similar 44% reduction. Indeed, there was no evidence for a differential influence of cortisol on false and correct memories ($p > .60$, for the interaction Cortisol/Placebo × False/Correct memories, $p = .009$, for the main effect “cortisol/placebo” across both types of memory). The response bias $C$ was not affected by cortisol administration and was comparable for false memories and correct memories (all $p > .15$; Figure 2A). Above all, the occurrence of false memories was positively correlated with correct memory retrieval when averaged across the treatment conditions ($r = .66, p = .02$), as well as in both conditions separately (cortisol: $r = .54, p = .07$; placebo: $r = .64, p = .026$; Figure 2B).

Raw measures of recognition not accounting for response biases, namely, false memory rate, hit rate, and false alarm rate, are displayed in Table 1. Only false memory rate tended to be reduced with enhanced cortisol levels ($p = .065$), suggesting that the parallel reduction of false and correct memories after cortisol in the bias-corrected measure $d'_{adj}$ partly relied on an increased false alarm rate, in addition to the decreases in false memory rate and hit rate, which per se did not reach significance (hit rate: $p = .15$; false alarm rate: $p = .37$).

<table>
<thead>
<tr>
<th>Table 1. Recognition Memory Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion ($P$)</td>
</tr>
<tr>
<td>False memory rate</td>
</tr>
<tr>
<td>False memory rate</td>
</tr>
<tr>
<td>Hit rate</td>
</tr>
<tr>
<td>False alarm rate</td>
</tr>
</tbody>
</table>

| Recognition Index ($d'$)                |
| False memories                          | $0.49 \pm 0.13$ | $0.87 \pm 0.10$ | .017* |
| Correct memories                        | $0.40 \pm 0.14$ | $0.72 \pm 0.08$ | .033* |

| Response Bias Index ($C$)                |
| False memories                          | $0.27 \pm 0.13$ | $0.35 \pm 0.11$ | .421 |
| Correct memories                        | $0.23 \pm 0.14$ | $0.27 \pm 0.10$ | .713 |

Recognition is indicated by the mean proportion ($P$) of old responses to prototypes (= false memory rate), studied shapes (= hit rate), and new shapes (= false alarm rate) as well as the recognition index $d'$ for false memories (false memory rate with reference to false alarm rate) and correct memories (hit rate with reference to false alarm rate). The response bias index $C$ is indicated separately for false memories and correct memories. $p$ Values are given for the comparison between the effects of cortisol and placebo. Means ± SEM are shown.

* $p < .05$.  

Diekelmann et al. 5
Confidence ratings, as well as remember, know, and guess judgments, did not differ between the cortisol and placebo conditions (see Table 2 for detailed results). Independent of cortisol enhancement, subjects were more confident on hits and false memories than on false alarms, and guess judgments were more frequent for false alarms than for hits and false memories (all \(p \leq .001\)).

### Cortisol and ACTH Concentrations

Cortisol levels were distinctly enhanced following cortisol infusion (\(p < .001\), for cortisol/placebo main effect and interaction with time). During false memory retrieval testing (i.e., 1930–2000 h), cortisol concentrations were 3- to 4-fold higher compared to the placebo condition (\(p < .001\); Figure 3). Additionally, and consistent with the well-known inhibitory feedback cortisol exerts on the hypothalamo-pituitary–adrenal system, cortisol infusion significantly decreased ACTH concentrations (main effect cortisol/placebo, \(p < .001\); Cortisol/Placebo × Time interaction, \(p = .004\); at retrieval testing, \(p < .002\); Figure 3). Although this decrease confirms effective (feedback) inhibition of ACTH by cortisol, it can be excluded as a mediator of the observed effects on memory retrieval, as ACTH per se has no strong effects on memory functions (Born, Fehm, & Voigt, 1986).

### DISCUSSION

Our results demonstrate that elevated cortisol levels at retrieval distinctly reduce the generation of false memories. In parallel, cortisol impaired retrieval of correct memories.

#### Table 2. Confidence Ratings and Remember/Know/Guess Judgments

<table>
<thead>
<tr>
<th></th>
<th>Cortisol</th>
<th>Placebo</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Confidence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False memories</td>
<td>2.67 ± 0.12</td>
<td>2.78 ± 0.09</td>
<td>.216</td>
</tr>
<tr>
<td>Hits</td>
<td>2.71 ± 0.08</td>
<td>2.63 ± 0.08</td>
<td>.161</td>
</tr>
<tr>
<td>False alarms</td>
<td>2.39 ± 0.07</td>
<td>2.32 ± 0.06</td>
<td>.538</td>
</tr>
<tr>
<td><strong>Remember</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False memories</td>
<td>0.24 ± 0.07</td>
<td>0.23 ± 0.06</td>
<td>.915</td>
</tr>
<tr>
<td>Hits</td>
<td>0.22 ± 0.06</td>
<td>0.18 ± 0.05</td>
<td>.223</td>
</tr>
<tr>
<td>False alarms</td>
<td>0.17 ± 0.05</td>
<td>0.19 ± 0.07</td>
<td>.688</td>
</tr>
<tr>
<td><strong>Know</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False memories</td>
<td>0.43 ± 0.03</td>
<td>0.53 ± 0.06</td>
<td>.110</td>
</tr>
<tr>
<td>Hits</td>
<td>0.43 ± 0.04</td>
<td>0.42 ± 0.05</td>
<td>.877</td>
</tr>
<tr>
<td>False alarms</td>
<td>0.39 ± 0.05</td>
<td>0.27 ± 0.05</td>
<td>.045</td>
</tr>
<tr>
<td><strong>Guess</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False memories</td>
<td>0.33 ± 0.06</td>
<td>0.23 ± 0.07</td>
<td>.119</td>
</tr>
<tr>
<td>Hits</td>
<td>0.35 ± 0.07</td>
<td>0.39 ± 0.07</td>
<td>.448</td>
</tr>
<tr>
<td>False alarms</td>
<td>0.44 ± 0.06</td>
<td>0.54 ± 0.08</td>
<td>.154</td>
</tr>
</tbody>
</table>

Mean confidence ratings (ranging from 1 = “guess” to 4 = “sure”) and proportions of remember, know, and guess judgments are displayed for items judged as “old.” \(p\) Values are given for the comparison between the effects of cortisol versus placebo. Means ± SEM are shown.

Confidence ratings, as well as remember, know, and guess judgments, did not differ between the cortisol and placebo conditions (see Table 2 for detailed results). Independent of cortisol enhancement, subjects were more confident on hits and false memories than on false alarms, and guess judgments were more frequent for false alarms than for hits and false memories (all \(p \leq .001\)).
This latter result confirms findings from several previous studies in humans and animals of a glucocorticoid-induced impairment of retrieval function (Het, Ramlow, & Wolf, 2005; Kuhlmann et al., 2005; Lupien et al., 2005; de Quervain et al., 2000, 2003; McGaugh & Roozendaal, 2002; Wolf et al., 2001). Both effects occurred in the absence of changes in general response bias. Also, confidence ratings and remember/know/guess judgments were not affected by cortisol. The present study is the first to directly and selectively manipulate cortisol concentrations at retrieval in relation to false memory formation. As cortisol was administered shortly before retrieval and more than 7 hr after initial learning, the observed effect is specifically on retrieval, whereas effects on encoding and consolidation can be excluded.

Whereas the effects of cortisol on false memory generation have not been investigated so far, several foregoing studies investigated the influence of stress on false memories which, among others, is characterized by the profound release of endogenous cortisol. Yet, those studies reported mixed results, with either no changes or enhanced rates of false memories following psychosocial stress (Smeets, Jelicic, & Merckelbach, 2006; Payne et al., 2002). As these studies introduced stress before encoding, effects of stress on retrieval could not be dissociated from those on encoding or consolidation. A recent study separately testing the effects of cold pressor stress on these memory processes found, in line with the present results, slightly decreased false memory rates if stress accompanied retrieval (Smeets, Otgaar, et al., 2008). Nevertheless, the patterns observed in these studies of stress, in principle, remain difficult to interpret in relation to cortisol because stress substantially affects numerous endocrine and cognitive parameters other than cortisol (Ulrich-Lai & Herman, 2009; Lupien et al., 2007; de Kloet et al., 2005).

Our finding of reduced false memory rates, in parallel with impaired correct memory retrieval, has implications for the current theorizing on the formation of false memories. Of the two main theoretical frameworks currently discussed, the retrieval monitoring theories assume that subjects during learning consciously or unconsciously generate the prototype of the single exemplars, which are all highly associated with the prototype. This internal generation of the prototype produces a sense of familiarity at subsequent retrieval testing. The cause for the occurrence of a false memory is assumed to be a failure of retrieval monitoring, that is, the subject mistakes this sense of familiarity for having actually encountered the prototype during encoding (Mitchell & Johnson, 2009; Johnson et al., 1993). Effective retrieval monitoring, that is, the ability to discriminate between familiarity due to external presentation or internal generation, has been shown to be essential for avoiding false memories (Curran, Schacter, Johnson, & Spinks, 2001). Retrieval monitoring critically relies on prefrontal cortex (Mitchell & Johnson, 2009; Turner, Simons, Gilbert, Frith, & Burgess, 2008; Dobbins, Simons, & Schacter, 2004), and false memory generation is likewise associated with increased prefrontal cortex activity (Kubota et al., 2006; Schacter & Slotnick, 2004; Schacter, 1996). Retrieval monitoring can be also improved by acute psychosocial stress (Smeets, Sijstermans, et al., 2008; Smeets, Jelicic, Merckelbach, et al., 2006). Because prefrontal cortex, in addition to other brain areas, is a significant target of glucocorticoids (Wang et al., 2005; Radley et al., 2004; Lupien & Lepage, 2001; Sanchez, Young, Plootsky, & Insel, 2000), it could be speculated that cortisol acts on prefrontal cortex to improve retrieval monitoring, thus reducing the occurrence of false memories. However, this view of an improved retrieval monitoring would not integrate the opposite (i.e., impairing) effect of cortisol on retrieval of correct memories.

Gist- or schema-based theories, on the other hand, propose that subjects remember the gist (i.e., the concept or schema) of single events rather than the specific details of the individually learned exemplars (Reyna & Brainerd, 1998; Posner & Keele, 1968). Each of the single exemplars reveals a specific pattern of activation in the associative network during encoding (Gallo & Roediger, 2002; McClelland et al., 1995; Bransford & Franks, 1971; Posner & Keele, 1968). Because all of the exemplars are derived from one prototype, they share common features resulting in the activation of overlapping representations which, most importantly, also overlap with regard to networks that represent the prototype. During encoding, the networks representing the prototype thus become automatically activated due to spreading activation from the individual exemplars and, paradoxically, even receives the greatest activation (because it has the most features in common with all single exemplars), although it was never encountered by the subject as an individual pattern. The more the subject relies on the gist or schema of what he or she experienced during learning, the more he or she falsely remembers the prototype of the learned exemplars (Gallo & Roediger, 2002; Brainerd & Reyna, 2001). It is assumed that, in the course of memory formation, the two features of an episode (i.e., the actually encountered exemplars and the common schema or gist of these exemplars) both become stored as distinct representations, that is, whenever subjects encounter an event, specific details of the individual exemplars are stored as single entities in the associative memory network but, simultaneously, common features of these exemplars, namely, the prototype, become stored as the gist representation. Thus, correct memories of the exemplars and false memories of the prototype refer to discrete entities represented in separate memory traces, though sharing some overlapping features. Although little is known about the neuronal correlates of gist representations, there is some evidence that both representations of the gist and of the specific events depend on hippocampal and medial temporal lobe regions (Garoff-Eaton, Slotnick, & Schacter, 2006; Schacter et al., 1996). These brain regions express a high density of glucocorticoid receptors and are well known to be particularly sensitive to the effects of cortisol (Lupien et al., 2007; Joels, 2001; Lupien & Lepage, 2001; de Kloet, Vreugdenhil, Oitzl, & Joels, 1998).
Assuming that false and correct memories are stored in the same networks but as separate representations, retrieval of false memories would be expected to be reduced by high cortisol levels to the same extent as retrieval of correct memories. Our present findings indeed indicate that both retrieval of correct and false memory was markedly decreased by elevated cortisol, with the magnitude of the decrease being comparable for both types of memory. Finally, we found the reduction of false memories to be strongly correlated with the impairment of correct memory retrieval, a finding in line with the gist-based view on false memories, suggesting that effects of cortisol on both false and correct memories rely on the same mechanism basically impairing retrieval operations.

Thus, our results essentially support gist-based theories of false memory formation rather than retrieval monitoring theories. To be noted, we applied a nonverbal version of the DRM paradigm using abstract shapes instead of word lists, which have been more commonly employed in previous studies. It has been argued that the use of abstract shapes prevents subjects from internally generating the gist representation at encoding, which might also prevent subsequent retrieval monitoring (Koutstaal & Schacter, 1997). Although this possibility cannot be completely ruled out without further testing, in our view, it is not likely that the occurrence of retrieval monitoring essentially depends on the stimulus material because the basic process promoting the encoding of a gist representation is the activation of overlapping representations of the single exemplars during the learning phase. This activation of overlapping representations, resulting in an internal “generation” of the gist representation, presumably occurs similarly with abstract shapes and words (although only in the case of words it may occasionally occur that activation of a specific gist representation enters consciousness). Importantly, in this view, gist representations of both words and abstract shapes similarly produce a sense of familiarity at subsequent retrieval testing, provoking failures in retrieval monitoring. Hence, if false memories were due to erroneous retrieval monitoring, this indeed should have been detected in the present study using abstract shapes. Rather than pointing toward an impaired retrieval monitoring as cause of false memories, our data speak for the notion that such false memories are generated as part of the process leading to the formation of correct memories, with common underlying neurophysiological (i.e., cortisol-dependent) mechanisms. It will be an intriguing issue of future studies to further specify the particular brain circuitry by which cortisol impacts retrieval of false and correct memories. Because it is well known that glucocorticoid receptors are expressed throughout prefrontal cortex and the hippocampus, one outstanding question centers around understanding the relative contributions of these two structures to correct memory retrieval and false memory suppression within the context of elevated cortisol.

In sum, our finding of a parallel modulation of false memories and correct memories by elevated cortisol suggests that both types of memory share common general mechanisms, with false memories possibly being the cost of an otherwise adaptive memory system that is able to extract general knowledge from single encountered events (Schacter, 1999). This view is eventually supported also by studies in amnesic patients who do not only display impaired memory for true events but likewise exhibit, in parallel, a distinctly reduced production of false memories in comparison with healthy controls (Koutstaal, Verfaellie, & Schacter, 2001; Schacter, Verfaellie, & Anes, 1997), possibly due to a diminished capability of these patients to extract a gist representation from the learned exemplars (Verfaellie, Schacter, & Cook, 2002). Also, the consolidating influence of sleep on memory has been revealed to concurrently enhance veridical memories and the occurrence of false memories (Diekelmann et al., in press; Payne et al., 2009). Thus, false memories appear to be tightly linked to the formation of correct memories, with both types of memory relying on basically similar neurophysiological mechanisms.

Acknowledgments

We thank Johannes Schneeweiß for help with data collection and Heidi Ruf and Ingrid von Lützau for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 654 “Plasticity and Sleep”).

Reprint requests should be sent to Jan Born, Department of Neuroendocrinology, University of Lübeck, Ratzeburger Allee 160, Haus 25a, 23538 Lübeck, Germany, or via e-mail: born@kfg.uni-luebeck.de.

REFERENCES


Cognitive Psychology, 2, 41–100.

Human Behavior, 28, 41–100.

Psychoneuroendocrinology, 26, 307–317.

Journal of Cognitive Neuroscience, 13, 201–216.

Nature Reviews Neuroscience, 6, 463–475.


