Routledge
Taylor & Francis Group

Mood and the DRM paradigm: An investigation of the effects of valence and arousal on false memory

Ilse Van Damme

Laboratory of Experimental Psychology, University of Leuven, Leuven, Belgium

Recent studies regarding the effect of mood on the DRM (Deese–Roediger–McDermott) illusion have not been able to clearly establish yet whether valence or arousal is most critical in determining susceptibility to false memories, nor what the underlying processes are. In three experiments, both the valence and the level of arousal of participants' mood were manipulated. Six conditions were used: positive mood with high/low arousal, negative mood with high/low arousal, neutral mood, and a control condition. Memory was tested by means of immediate and delayed recognition and immediate free recall. The mood induction procedure was effective. For recognition memory, there was an effect of arousal on the endorsement of critical lures. Low-arousal moods elicited more false recognition than high-arousal moods, regardless of valence. Based on signal detection analyses, the effect was attributed to more liberal response criteria with low arousal, in combination with a tendency towards improved item-specific memory with high arousal.

Keywords: False memory; Mood; Arousal; Valence; Free recall; Recognition.

By now, it has clearly been established that memory is not a recording device. People are vulnerable to memory distortion and can be tricked into remembering things that never happened (e.g., Brainerd & Reyna, 2005; Loftus, 2005; Schacter, 2001). A simple and straightforward paradigm to illustrate the constructive nature and fallibility of human memory is the "DRM" (Deese–Roediger–McDermott) paradigm (Deese, 1959; Roediger & McDermott, 1995; see also Gallo, 2010). When participants are presented with lists of words that are associatively related to a critical nonpresented lure word, they tend to falsely recall and recognize the lure word afterwards. Moreover, the probability of doing so typically equals or even exceeds the probability of correctly remembering list words.

According to the "activation-monitoring" account (e.g., Roediger & McDermott, 1995, 2000), the lure words are mentally activated due to their associative connection to the list words. When memory is tested, participants are confronted with a source-monitoring problem (cf. Johnson, Hashtroudi, & Lindsay, 1993), as they need to distinguish between items that were

actually studied and items that were merely activated. Source misattribution will induce false memories. According to the "fuzzy-trace" theory (e.g., Brainerd & Reyna, 1998, 2002), verbatim traces and gist traces are stored in parallel during list presentation. Whereas the former preserve specific information about the identity of each item, the latter capture the general meaning and interpretations invoked by the list words. The retrieval of gist traces may generate meaning-consistent false memories. The activation-monitoring account and fuzzy-trace theory are not mutually exclusive and can be integrated (e.g., Van Damme & d'Ydewalle, 2009b). Importantly, both theories attribute false memories (at least in part) to some sort of "relational" encoding (i.e., the encoding of information that is common to various input elements; cf. Hunt & Einstein, 1981; Hunt & McDaniel, 1993).

Recently, researchers have started to look into the effects of mood[1] on susceptibility to the DRM illusion. Storbeck and Clore (2005) induced positive and negative moods, and compared DRM performance in both mood groups with a control group (in which no mood induction took place). Memory was tested by means of free recall immediately after each list. Whereas there were no differences in veridical memory, individuals in a negative mood were significantly less likely to show false memories than those in a positive mood and the control group (see Forgas, Laham, & Vargas, 2005, for comparable findings with the misinformation paradigm). The authors had predicted this based on the "affect-as-information" theory (e.g., Clore & Huntsinger, 2007; Clore & Palmer, 2009), which states that negative affect experienced as task-relevant feedback triggers *item-specific* processing, whereas positive affect triggers *relational* processing (cf. Bless & Fiedler, 2006: stimulus-driven vs. knowledge-driven). Hence, negative moods will reduce the likelihood of false memories due to a stronger focus on item-specific details during the encoding phase. The alternative explanation that negative

moods might induce better source monitoring was ruled out. When participants were asked to recall not only studied words, but all words that came to mind during the experiment (i.e., excluding the need for monitoring), the group difference remained. This strongly suggests that the effect was due to differences at encoding, not retrieval (see also Storbeck & Clore, 2011, but see Wright, Startup, & Mathews, 2005). The fact that there was no difference between positive mood and the control group additionally points to the conclusion that relational processing is the default strategy in any nonnegative state. This has been suggested before, both in the context of the DRM paradigm and in general (e.g., Clore & Huntsinger, 2007; Hunt & McDaniel, 1993; Lampinen, Leding, Reed, & Odegard, 2006; Van Damme & d'Ydewalle, 2009a).

Corson and Verrier (2007) built further on these findings and wanted to examine whether the result pattern obtained was truly due to the valence of participants' mood, or rather to differences in the degree of arousal or activation. Indeed, a widely accepted framework proposes that emotional experiences embody two orthogonal dimensions—valence and arousal—which can either separately or interactively affect cognition (e.g., Revelle & Loftus, 1990, 1992; Russell, 1980, 2003). To evaluate the relative effects of both dimensions, Corson and Verrier induced five specific moods: "serene" (positive valence, low arousal), "happy" (positive valence, high arousal), "sad" (negative valence, low arousal), "angry" (negative valence, high arousal), and "neutral". They used the traditional DRM procedure (Roediger & McDermott, 1995), with free recall tests immediately after each list and a global recognition test at the end of the experiment. The effect of arousal turned out to be the only significant one. Both false recall and false recognition were more frequent in the high-arousal conditions (happy, angry) than in the low-arousal conditions (serene, sad, neutral).

Corson and Verrier (2007) suggested two possible explanations for these findings. On the one

---

[1] "Mood" is typically defined as a diffuse affective state, whereas "emotion" is more short-lived and reflects an underlying judgement of a particular object or situation (e.g., Forgas, 1995).

hand, they pointed out that their results can be fitted within the affect-as-information approach, as the critical factor in this theory is not affect itself, but rather the information conveyed about the validity of one's own point of view. Both anger and happiness induce confidence in one's own cognitions, and it is this inherent positive information that induces relational information processing (Clore & Huntsinger, 2007). Unfortunately, however, this explanation cannot account for the whole result pattern. Similar (low) false memory levels were obtained in sad, serene, and neutral moods, whereas it is clear that the information conveyed about one's own position is quite different in these three conditions. Moreover, results are difficult to reconcile with the finding that relational encoding is the default strategy in nonnegative mood states (e.g., Storbeck & Clore, 2005). An alternative explanation given by Corson and Verrier is the influence of arousal on the encoding strategy used: High arousal elicits elaborative encoding. There seems to be some confusion about the meaning of the concept "elaborative", however. For instance, the authors refer to Schwartz (1975), but in fact this study demonstrates that high arousal induces a focus on physical aspects of verbal material, whereas low arousal induces a focus on semantic, categorical aspects. Applied to the DRM paradigm, this would predict the *opposite* result pattern as the one obtained. Furthermore, there is recent evidence that low arousal induces an overgeneral retrieval style (McBride & Cappeliez, 2004), which would also predict the opposite result pattern. Finally, it is important to note that the effects of extraneous arousal and of emotional arousal elicited by the study object or experience itself are not necessarily identical. The latter has been shown to enhance memory for central or gist information at the cost of (peripheral) details (e.g., Adolphs, Denburg, & Tranel, 2001; Christianson & Loftus, 1991; see also Mather, 2007). With respect to the former, however, the typical finding is that arousal enhances consolidation of studied information and thus improves recollection (e.g., Anderson, Wais, & Gabrieli, 2006; Nielson & Powless, 2007).

To summarize, recent studies regarding the effects of mood on false memory have obtained inconsistent results. It has not been clearly established yet whether valence or arousal is most critical in determining susceptibility to the DRM illusion, nor what the underlying processes are. Three experiments were conducted, with a twofold goal. The *first objective* was to investigate whether Corson and Verrier's (2007) results could be replicated, with some methodological improvements. The type and timing of memory tests were both manipulated: A delayed recognition test was used in Experiment 1, immediate free recall tests in Experiment 2, and immediate recognition tests in Experiment 3. The "classic" procedure in which immediate free recall tests are followed by a global recognition test was not used, since it has been shown that taking a first test enhances susceptibility to false memory when given a later test (e.g., Roediger & McDermott, 1995). In addition to the four specific mood conditions (serene, happy, sad, angry), two control conditions were included: one with a neutral mood induction, and one in which no mood induction took place. This was done to examine whether different operationalizations might have been responsible for the different results obtained for control participants in previous studies. It was hypothesized that a neutral mood induction might not be entirely neutral, but rather boring or sleep inducing (Whissell & Levesque, 1988; as cited in Revelle & Loftus, 1990), eliciting low arousal. The *second objective* was to investigate the processes underlying the effect of mood on false memory. For this purpose, signal detection analyses were carried out.

GENERAL OUTLINE

Participants were tested in groups that ranged in size from 10 to 20 people and were randomly assigned to one of six experimental conditions. In five conditions, a specific mood was induced: positive with low arousal (serene), positive with high arousal (happy), negative with low arousal (sad), negative with high arousal (angry), and neutral (control 1). The sixth condition was a control

condition in which no mood induction took place (control 2). The mood induction and assessment methods were taken from Corson and Verrier (2007).

Mood induction

The mood induction procedure was a composite technique based on music and guided imagery (Mayer, Allen, & Beauregard, 1995; see also Jallais & Gilet, 2010). Participants first listened to music, selected for the specific mood being induced. After 1 min, vignettes describing certain situations appeared on the computer screen (while the music continued). Examples are: *"After a long winter's day, you calmly take a gloriously hot bath"* (serene); *"You buy a lottery ticket and you win € 250 instantly"* (happy), *"No one remembers your birthday"* (sad); *"Someone put a big scratch in your car while it was parked in the lot and didn't even bother to leave a note"* (angry); *"The capital is usually one of the biggest cities in the country"* (control).

In all conditions, there were eight vignettes in total, shown for 30 s each. Participants were asked to imagine the situations as vividly as possible. The musical pieces were previously used and validated by Mayer et al. (1995) and/or Corson and Verrier (2007). Several of them had also been validated in various earlier mood induction studies (Gerrards-Hesse, Spies, & Hesse, 1994). The vignettes were taken from Corson and Verrier (2007) and were translated into Dutch.

Mood assessment

Self-report measures were gathered both prior to and after the mood induction phase.[2] First, participants were asked to indicate the status of their current mood by completing a Dutch version of the Brief Mood Introspection Scale (BMIS; Mayer & Gaschke, 1988). This is a 16-item self-report mood questionnaire, in which one is asked to rate adjectives describing mood states on a 7-point Likert scale. Valence scores were calculated via subtractive scoring based on the pleasant–unpleasant subscale (range: –48 to 48); arousal scores were based on the aroused–calm subscale (range: 4 to 68). Second, participants evaluated valence and arousal directly by marking one of the squares of a 9×9 matrix in Experiment 1, and by means of separate rating scales in Experiments 2 and 3.

Procedure

Participants were told that they would take part in two unrelated experiments: one in which they would be asked to imagine several situations, and one in which their memory would be tested. The so-called first experiment consisted of the mood induction phase, preceded and followed by a mood assessment. In the so-called second experiment, participants were told they would be hearing lists of words, of which they would have to try to remember as many as possible. Ten Dutch DRM lists were presented auditorily, in a female voice, at a rate of approximately 1.5 s per word. Memory was tested by means of free recall or old/new recognition. Participants in the negative conditions were given the opportunity to listen to "happy" music at the end of the experiment, in order to improve their mood before leaving the room. The entire procedure lasted about 25 min. E-Prime 1.1 (Psychology Software Tools, Pittsburgh, PA) was used for the presentation of all materials.

Hypotheses

Based on the studies and theories discussed, three alternative hypotheses about the effect of mood on false memory could be put forward. *First*, based on Storbeck and Clore (2005) and the

---

[2] Participants in Control Condition 2 evaluated their mood only once. They skipped the mood induction phase and immediately proceeded with the DRM part. This allowed a comparison between the natural state in which participants came into the lab (i.e., control condition used in Storbeck & Clore, 2005) and a neutral mood induction (i.e., control condition used in Corson & Verrier, 2007).

affect-as-information theory, it could be predicted that false memories would be more likely in conditions conveying positive (or at least nonnegative) information about one's own point of view (serene, happy, angry, control) than in conditions conveying negative information (sad). *Second*, based on Corson and Verrier (2007), it could be predicted that false memories would be more likely in high-arousal conditions than in low-arousal conditions. *Third*, based on evidence that high arousal induces item-specific processing (Schwartz, 1975) and improves recollection (e.g., Nielson & Powless, 2007), whereas low arousal induces an overgeneral retrieval style (McBride & Cappeliez, 2004), it could be predicted that false memories would be *less* likely in high-arousal conditions than in low-arousal conditions.

EXPERIMENT 1

The primary goal of Experiment 1 was to evaluate Corson and Verrier's (2007) recognition results. So far, this was the only study investigating the effect of mood on false recognition in the DRM paradigm, and results turned out to be rather difficult to explain. Based on the finding that taking a first test enhances susceptibility to false memory in a later test (e.g., Roediger & McDermott, 1995), only a delayed recognition test was used in the present experiment (without preceding free recall). In comparison to Corson and Verrier, the composition of the test was improved. Two sets of DRM lists were used as either studied or unstudied materials, in a counterbalanced way, which allowed for specific distractor items for list words and critical lures. Underlying processes were investigated by means of signal detection analyses.

Method

Participants
A total of 301 undergraduate students (24% male, 76% female) participated in the experiment in exchange for course credit or an incentive. Their average age was 20.21 years ($SD = 2.49$; range 17–35).

Materials
Twenty word lists were used. Each list was composed of 15 strong associations to one critical word, which was never presented for study. Words within a list were ordered by association frequency, with the strongest associations occurring first. Lists were taken from Van Damme and d'Ydewalle (2009a, 2009b) and were constructed following the procedure of Roediger and McDermott (1995). Some minor changes were made (i.e., occasionally a word was replaced or switched position), based on recent word association norms collected by De Deyne and Storms (2008).

To allow for counterbalancing of studied and nonstudied materials, the lists were divided into two sets of 10, roughly equated on word frequency of the critical lures ($M = 117$ and 120 occurrences per million, based on the CELEX lexical database of Baayen, Piepenbrock, & Gulikers, 1995). One set was to be used for study, and the other was to remain unstudied and to serve as baseline in the recognition test. Assignment of the sets to the two positions was counterbalanced across participants and conditions. The recognition test was composed of 80 words, which were randomly presented: 30 list targets (1st, 8th, 10th item from each studied list), 30 list distractors (1st, 8th, 10th item from each unstudied list), 10 critical targets (related to the studied lists), and 10 critical distractors (related to the unstudied lists). Test items were visually presented, using black upper-case letters in the centre of the screen.

Design and procedure
Participants were randomly assigned to one of the six mood conditions: serene ($n = 49$), happy ($n = 49$), sad ($n = 49$), angry ($n = 51$), control 1 ($n = 51$), and control 2 ($n = 52$). Before and after the mood induction, they evaluated their mood by means of the BMIS and a 9×9 matrix (cf. Corson & Verrier, 2007). In the latter, valence and arousal were evaluated directly and simultaneously, by marking one of the 81 squares of the matrix. The horizontal dimension indicated valence; the vertical dimension indicated arousal. Next, 10 word lists were presented, with intentional

learning instructions. Participants were asked to try to remember as many of the words as possible, for a memory test that would follow at the end of the experiment. Immediately after each list, five math problems were to be solved (i.e., adding and multiplying, ± 1 min; cf. McDermott, 1996; Roediger & McDermott, 1995). After all lists had been presented, the recognition test was administered. Participants were asked to indicate for each item whether or not it had been part of one of the studied lists.

Results and discussion

Mood manipulation check

At the start of the experiment, there were no significant group differences in either valence or arousal (all Fs ≈ 1). From the first to the second administration of the BMIS, scores changed in accordance with the induced moods. An exception was the sad condition, in which BMIS arousal scores had slightly increased rather than decreased. As predicted, the neutral mood induction significantly lowered participants' arousal scores, $F(1, 241) = 16.68$, $p < .0001$. This supports the hypothesis that such an induction is not entirely neutral, but rather relaxing or boring.

Since the mood difference between groups after the induction was a precondition for testing the effect of mood on false recognition, severely outlying values were removed from the dataset.[3] Eight participants showed valence and/or arousal scores that deviated extremely (>2.5 SDs) from their condition mean. For these (few) individuals, the manipulation had failed completely, and they were therefore left out.[4] Table 1 presents averages for the resulting participants. After the induction, BMIS scores of the six conditions were reliably different in both valence, $F(5, 287) = 25.52$, $MSE = 102.63$, $p < .0001$, $\eta^2 = .31$, and arousal, $F(5, 287) = 11.82$, $MSE = 40.90$, $p < .0001$, $\eta^2 = .17$.

Planned comparisons (Dunn) revealed that valence scores were significantly higher for the two positive groups (serene, happy) than for the two negative groups (sad, angry), $t(287) = 10.12$, $p < .0001$. The two control groups also showed relatively high valence scores, which significantly differed from those of the two negative groups, $t(287) = 8.70$, $p < .0001$. The difference with the two positive groups did not reach statistical significance ($p = .11$). Between-group comparisons of matrix ratings revealed analogous results.

BMIS arousal scores were significantly higher for the two high-arousal groups (happy, angry) than for the two low-arousal groups (serene, sad), $t(287) = 4.73$, $p < .0001$. As can be seen in Table 1, however, arousal was higher in the sad condition than in the serene condition (Tukey, $p = .002$). Nevertheless, planned comparisons within-valence showed the appropriate differences, both for positive (happy vs. serene), $t(287) = 4.02$, $p < .0001$, and for negative valence (angry vs. sad), $t(287) = 2.65$, $p = .009$. In addition, in the next two experiments it was shown that despite differences in BMIS arousal scores, virtually identical arousal scores were obtained in the serene and sad conditions when participants were asked to evaluate their subjective level of arousal simply and directly on a 0–100 rating scale. When looking more closely at the "aroused-calm" subscale of the BMIS, it turned out that the adjective "sad" is considered to be reflecting arousal rather than calmness in the questionnaire. Hence, indicating that one felt sad (the affective state induced in the negative low-arousal condition) increased a participant's BMIS arousal score rather than decreasing it (albeit not drastically: it is one out of 12 items counted). The specific adjectives in the questionnaire and the composition of the subscale thus unwantedly boosted the arousal score of participants in the sad condition as compared to the serene condition.

Due to the decrease elicited by the neutral mood induction ($p < .0001$; cf. supra), arousal was lower

---

[3] Three in the serene condition, two in the happy condition, one in the angry condition, one in Control Condition 1, and one in Control Condition 2.

[4] Storbeck and Clore (2005, p. 787) reported a similar procedure, by which "participants were retained for analysis only when their self-reports of mood matched the mood intended to be induced".

Table 1. Average valence and arousal scores in Experiment 1, before (1) and after (2) the mood induction

	Valence		Arousal	
Mood	BMIS (1)	BMIS (2)	BMIS (1)	BMIS (2)
Serene	11.09 (1.71)	15.04 (1.55)	25.40 (0.80)	22.63 (0.74)
Happy	12.72 (1.36)	20.21 (1.35)	26.04 (0.93)	27.96 (0.79)
Sad	13.65 (1.64)	2.59 (1.59)	25.59 (0.99)	27.69 (0.90)
Angry	11.24 (1.56)	3.04 (1.50)	26.52 (0.91)	31.10 (1.03)
Control 1	14.27 (1.52)	16.12 (1.36)	26.96 (0.97)	23.52 (0.90)
Control 2	14.45 (1.34)	14.45 (1.34)	25.43 (1.04)	25.43 (1.04)

Note: BMIS = Brief Mood Introspection Scale. In Control Condition 2, the BMIS was administered only once, so the same value was used for group comparisons before and after the mood induction. Standard errors of the mean are given in parentheses.

in Control Condition 1 than in Control Condition 2. Although the difference did not reach statistical significance ($t(287) = 1.50$, $p = .13$; adjusted alpha $= .05/4 = .0125$), Table 1 shows that arousal in Control Condition 1 was comparable to arousal in the serene condition, whereas the score obtained in Control Condition 2 was intermediate between those of the serene and happy conditions. When considering the matrix ratings, a slightly different result pattern emerged. Participants apparently found it hard to disentangle valence and arousal in the matrix. When in a positive mood, they consistently placed a cross in the low-arousal half of the matrix, whereas negative moods tended to be judged as more arousing. This could be attributed to the difficult Dutch translation of the word "arousal" and was taken into consideration in the next two experiments.

Table 2. Proportions of items judged as old in Experiment 1, as a function of Mood

	List words		Critical lures	
Mood	Targets	Distractors	Targets	Distractors
Serene	.71 (.02)	.10 (.01)	.81 (.03)	.11 (.02)
Happy	.72 (.02)	.09 (.01)	.74 (.03)	.06 (.01)
Sad	.70 (.02)	.11 (.01)	.82 (.02)	.10 (.02)
Angry	.67 (.02)	.08 (.01)	.76 (.02)	.08 (.02)
Control 1	.72 (.02)	.10 (.01)	.81 (.03)	.11 (.02)
Control 2	.68 (.02)	.08 (.01)	.77 (.03)	.08 (.02)

Note: Standard errors of the mean are given in parentheses.

Recognition performance

Recognition data for list words and critical lures in all six conditions are presented in Table 2. The average probability of correctly recognizing list targets was .70, and the average probability of falsely recognizing critical targets was .78. False alarms to unrelated distractor words were rather infrequent (.09 for both types of distractors). A one-way analysis of variance (ANOVA) with Mood as a between-subjects variable was carried out to evaluate its general effect on veridical and false recognition. In addition, a Valence × Arousal factorial ANOVA was used to evaluate Hypotheses 2 and 3 more directly. The latter was done in two ways. First, only the critical mood conditions were taken into account (i.e., serene, happy, sad, angry) in a 2 (valence: positive, negative) × 2 (arousal: high, low) ANOVA. Second, a 3 (valence: positive, neutral, negative) × 2 (arousal: high, low) ANOVA was carried out, including both control conditions as well (cf. mood manipulation check: In all three valences, a high- and a low-arousal group could be distinguished).

None of the analyses on the proportions of "old" responses to list targets revealed any significant effects (Fs < 1). Hence, in accordance with previous studies (Corson & Verrier, 2007; Storbeck & Clore, 2005), mood did not significantly affect veridical memory. On average, participants in all six conditions were equally good at recognizing studied words, regardless of their mood during the study phase (see Table 2).

The effect of mood on false recognition of critical targets in a one-way ANOVA did not reach statistical significance, $F(5, 287) = 1.42$, $p = .22$. However, in both types of Valence × Arousal analyses, there was a significant main effect of arousal, $F(1, 188) = 5.37$, $MSE = .03$, $p = .02$, $\eta_p^2 = .03$, and $F(1, 287) = 6.36$, $MSE = .04$, $p = .01$, $\eta_p^2 = .02$. As can be seen in Table 2, participants in the low-arousal conditions (serene, sad, control 1) demonstrated on average more false recognition of critical targets than participants in the high-arousal conditions (happy, angry, control 2). There was no evidence for a main effect of valence, nor for an interaction ($Fs < 1$). The results point to the conclusion that, when valence is kept constant, false memories are more likely with low arousal than with high arousal (cf. Hypothesis 3).

Interestingly, a similar result pattern was observed for false-alarm rates to distractor words (see Table 2). This suggests that the underlying process might be response bias rather than recollection. To clarify this matter, signal detection analyses were carried out.

Signal detection measures

A' was used as a measure of sensitivity or discriminability, and B_D'' was used as a measure of response bias. Both are nonparametric indexes (e.g., Donaldson, 1992; Grier, 1971; Snodgrass & Corwin, 1988) and have been successfully applied to DRM data before (e.g., Schacter, Verfaellie, Anes, & Racine, 1998; Tussing & Greene, 1997). A' is defined as $0.5 + [(H - FA)(1 + H - FA)]/[4H(1 - FA)]$, where H denotes hit rate and FA denotes false-alarm rate, and reflects the ability to discriminate between two types of information (most typically: studied vs. unstudied). Values can vary between 0 and 1, with higher values indicating greater discriminability and 0.5 indicating chance performance. If $FA > H$, the formula is adapted as follows: $0.5 - [(FA - H)(1 + FA - H)]/[4FA(1 - H)]$ (Snodgrass & Corwin, 1988). B_D'' is defined as $[(1 - H)(1 - FA) - HFA]/[(1 - H)(1 - FA) + HFA]$ and reflects the tendency to accept (reject) doubtful information as (in)correct. Values vary between −1 (indicating extremely liberal responding) and 1 (indicating extremely

conservative responding; Donaldson, 1992). As these measures are undefined with hit rates of 1 and false-alarm rates of 0, data were first transformed by recalculating the proportions of "old" responses by adding 0.5 to the frequencies and dividing by $(N + 1)$, as recommended by Snodgrass and Corwin (1988).

Both indexes were calculated in three ways. First, item-specific memory was investigated by treating "old" responses to studied list targets as hits and "old" responses to unstudied list distractors as false alarms (IM-1). Second, item-specific memory was investigated by treating "old" responses to studied list targets as hits and "old" responses to related critical targets as false alarms (IM-2). Third, gist memory (GM) was investigated by treating "old" responses to critical targets as hits and "old" responses to critical distractors as false alarms. This third measure does not necessarily reflect differences in the level of gist memory that was available to participants, but rather indicates differences in the degree to which they tended to rely on gist memory for their recognition judgements (Koutstaal & Schacter, 1997). Estimates are presented in Table 3.

Sensitivity measures revealed that participants were good at distinguishing list targets from unrelated list distractors (IM-1: $M = .88$), but poor at distinguishing list targets from related critical targets (IM-2: $M = .43$). As can be seen in Table 3, all conditions performed below chance making the latter comparison. This confirms that a strong semantic illusion was created. The consistently high level of A' when treating "old" responses to critical targets as hits (third part of Table 3) confirms that participants strongly relied on gist memory ($M = .89$). In addition, when comparing critical targets to list targets, they showed a liberal response bias (i.e., were likely to accept the item in case of doubt), whereas they responded rather conservatively when comparing both critical and list targets to unrelated distractors.

A 2 (valence) × 2 (arousal) analysis of variance on A' revealed a marginally significant main effect of arousal for the distinction between list targets and critical targets (IM-2), $F(1, 188) = 3.14$, $MSE = .03$, $p = .078$, $\eta_p^2 = .02$. Sensitivity

Table 3. Estimates of sensitivity (A') and response bias (B_D'') in Experiment 1, as a function of Mood

	Item-specific memory 1		Item-specific memory 2		Gist memory	
Mood	A'	B_D''	A'	B_D''	A'	B_D''
Serene	.88 (.01)	.50 (.06)	.41 (.02)	−.74 (.04)	.88 (.01)	.19 (.07)
Happy	.88 (.01)	.51 (.05)	.47 (.03)	−.64 (.06)	.89 (.01)	.44 (.06)
Sad	.87 (.01)	.50 (.06)	.40 (.02)	−.74 (.04)	.89 (.01)	.19 (.08)
Angry	.87 (.01)	.60 (.05)	.43 (.02)	−.64 (.04)	.89 (.01)	.39 (.08)
Control 1	.88 (.01)	.46 (.06)	.42 (.02)	−.73 (.05)	.89 (.01)	.17 (.09)
Control 2	.88 (.01)	.57 (.06)	.42 (.02)	−.64 (.06)	.89 (.01)	.34 (.07)

Note: Item-specific memory 1 = hits compared to unrelated false alarms; Item-specific memory 2 = hits compared to related false alarms; Gist memory = related false alarms compared to unrelated false alarms. Standard errors of the mean are given in parentheses.

estimates were higher for participants in the high-arousal conditions than for participants in the low-arousal conditions. This suggests that higher levels of arousal enhanced item-specific memory and therefore improved participants' ability to distinguish between actually studied words and related lure words.

2 (valence) × 2 (arousal) analyses of variance on B_D'' revealed significant main effects of arousal for IM-2 and GM, $F(1, 188) = 4.80$, $MSE = .10$, $p = .03$, $\eta_p^2 = .03$, and $F(1, 188) = 9.70$, $MSE = .25$, $p = .002$, $\eta_p^2 = .05$. When comparing critical targets to list targets as well as critical distractors, low-arousal participants responded more liberally than high-arousal participants did. A similar effect of arousal on response bias was observed in a 3 (valence) × 2 (arousal) ANOVA. Again, for both IM-2 and GM, the main effect of arousal was significant, $F(1, 287) = 6.36$, $MSE = .11$, $p = .01$, $\eta_p^2 = .02$, and $F(1, 287) = 11.15$, $MSE = .27$, $p = .001$, $\eta_p^2 = .04$. Hence, when comparing critical targets to other test items, participants in all three low-arousal conditions (serene, sad, control 1) used more liberal response criteria than participants in the other conditions did (see Table 3).

Summary

The mood induction was effective. Valence scores were significantly higher for the positive groups than for the negative groups, with control conditions obtaining intermediate values. Arousal scores were significantly higher for the high-arousal groups than for the low-arousal groups. In addition, support was obtained for the hypothesis that a "neutral" mood induction is not entirely neutral, but rather relaxing or boring. Although the difference between the two control conditions after the induction did not reach statistical significance, BMIS arousal scores were significantly lowered by the neutral mood induction. Assessment of arousal by means of a 9 × 9 matrix turned out to be unreliable, most likely due to the difficult translation of the word "arousal" into Dutch. Hence, a different assessment method was used in the following experiments.

False recognition of critical lures was shown to be more likely in the low-arousal groups than in the high-arousal groups. This is consistent with Hypothesis 3, which was based on evidence that arousal improves item-specific memory (e.g., Nielson & Powless, 2007; Schwartz, 1975) and that low arousal induces an overgeneral retrieval style (McBride & Cappeliez, 2004). Signal detection analyses confirmed that both types of processes played a role in determining the effect of mood on false memory. Sensitivity estimates for IM-2 were higher for happy and angry participants than for serene and sad participants, indicating that high arousal improved participants' ability to discriminate studied list words from related lure words. The effect was only marginally significant, however. Additionally, arousal was shown to influence participants' response criteria. When comparing related lures to both studied list words and

unrelated distractors, moods comprising low arousal induced more liberal responding than moods comprising high arousal. Hence, when feeling calm, participants were more inclined to endorse a semantically related item. Simply put: When in doubt, feeling relaxed made people say "yes" more easily.

EXPERIMENT 2

The goal of Experiment 2 was to investigate whether the effect of arousal on false memory obtained in Experiment 1 could be replicated for free recall. As discussed, differences in both item-specific memory and response bias were shown to be at the basis of the effect, but the former influence was only marginally significant. Accordingly, as free recall is a memory test in which no cues are given and in which intentional recollection is required to retrieve any information, it was expected that the effect of mood would be more modest for free recall than for recognition.

Method

Participants
A total of 302 undergraduate students (19% male, 81% female) participated in the experiment in exchange for course credit or an incentive. Their average age was 19.83 years ($SD = 2.84$; range 17–42).

Materials
The same word lists were used as in Experiment 1. Again, only one set of 10 lists was to be studied, counterbalanced across participants and conditions.

Design and procedure
Participants were randomly assigned to one of the six mood conditions: serene ($n = 50$), happy ($n = 50$), sad ($n = 52$), angry ($n = 52$), control 1 ($n = 50$), and control 2 ($n = 48$). Before and after the

mood induction, they evaluated their mood by means of the BMIS and two separate rating scales ranging from 0 to 100. On the first scale, they were asked to report how positive or negative their mood was at that time (0 = negative, 50 = neutral, 100 = positive). The second scale evaluated subjective arousal. In order to make the concept more clear, it was indicated that: 0 = drowsy (*slaperig*), 25 = calm (*rustig*), 50 = alert (*alert*), 75 = energetic (*energiek*), 100 = aroused (*opgewonden*). After the induction phase, 10 word lists were presented, with intentional learning instructions. Participants were given 1 min to recall each list immediately after its presentation. They were asked to write down all words they could remember, without guessing.

Results and discussion

Mood manipulation check
At the start of the experiment, there were no significant group differences in either valence or arousal (all Fs ≈ 1). From the first to the second administration of the BMIS and the rating scales, scores changed in accordance with the induced moods. As predicted and as in Experiment 1, subjective arousal was significantly lowered by the neutral mood induction, measured by both the BMIS, $F(1, 246) = 11.68$, $p = .0007$, and the rating scale, $F(1, 248) = 17.34$, $p < .0001$. This again supports the notion that such an induction is relaxing or calming.

Since the post-induction mood difference was a precondition for testing the effect of mood on false recall, severely outlying values were removed from the dataset (cf. Experiment 1). Ten individuals showed valence and/or arousal scores that deviated strongly (>2.5 SDs) from their condition mean and were therefore left out.[5] Table 4 presents averages for the resulting participants. After the induction phase, the six conditions were reliably different in both valence and arousal. This was the case when considering BMIS scores, $F(5, 286) = 24.80$, $MSE = 98.95$, $p < .0001$, $\eta^2 = .30$ (valence), $F(5,

---

[5] Two in the happy condition, one in the serene condition, one in the angry condition, one in the sad condition, two in Control Condition 1, and three in Control Condition 2.

Table 4. Average valence and arousal scores in Experiment 2, before (1) and after (2) the mood induction

	Valence				Arousal			
Mood	BMIS (1)	BMIS (2)	Rating (1)	Rating (2)	BMIS (1)	BMIS (2)	Rating (1)	Rating (2)
Serene	17.40 (1.52)	18.47 (1.14)	72.29 (2.05)	75.18 (1.88)	27.77 (0.84)	22.45 (0.71)	49.88 (3.16)	38.59 (2.59)
Happy	14.74 (1.38)	19.50 (1.37)	70.54 (1.94)	75.73 (1.80)	24.87 (1.00)	27.79 (0.76)	47.58 (3.11)	58.35 (2.88)
Sad	14.62 (1.59)	4.10 (1.63)	66.94 (2.30)	49.55 (2.53)	25.24 (0.93)	25.51 (0.67)	45.20 (2.86)	38.88 (2.70)
Angry	12.14 (1.57)	4.16 (1.48)	66.33 (2.25)	55.73 (2.00)	25.96 (0.84)	28.76 (0.92)	44.43 (2.77)	50.18 (2.65)
Control 1	14.42 (1.43)	15.00 (1.43)	68.33 (1.88)	70.92 (1.97)	26.46 (0.89)	23.94 (0.79)	50.02 (2.77)	41.29 (2.80)
Control 2	16.56 (1.43)	16.56 (1.43)	72.13 (2.02)	72.13 (2.02)	26.84 (0.92)	26.84 (0.92)	50.69 (3.19)	50.69 (3.19)

Note: BMIS = Brief Mood Introspection Scale. In Control Condition 2, the BMIS/rating scales were administered only once, so the same value was used for group comparisons before and after the mood induction. Standard errors of the mean are given in parentheses.

286) = 9.06, $MSE = 31.05$, $p < .0001$, $\eta^2 = .14$ (arousal), as well as rating scores, $F(5, 286) = 29.81$, $MSE = 205.97$, $p < .0001$, $\eta^2 = .34$ (valence), $F(5, 286) = 8.17$, $MSE = 380.33$, $p < .0001$, $\eta^2 = .12$ (arousal).

Planned comparisons (Dunn) on the BMIS scores revealed that valence was significantly higher for the two positive groups (serene, happy) than for the two negative groups (sad, angry), $t(286) = 10.53$, $p < .0001$. The two control groups showed intermediate values, which differed significantly from valence in the negative groups, $t(286) = 8.17$, $p < .0001$, and marginally significantly from valence in the positive groups, $t(286) = 2.22$, $p = .0272$ (adjusted alpha $= .05/3 = .0167$). The same analyses on the rating scores revealed comparable results (see Table 4) and are therefore not reported.

BMIS arousal scores were significantly higher for the two high-arousal groups (happy, angry) than for the two low-arousal groups (serene, sad), $t(286) = 5.44$, $p < .0001$. As in Experiment 1, arousal was somewhat higher in the sad condition than in the serene condition (Tukey, $p = .07$), most likely due to the composition of the questionnaire. Comparisons within-valence again showed the appropriate differences for both positive, $t(286) = 4.72$, $p < .0001$, and negative groups, $t(286) = 2.95$, $p = .003$. Arousal was significantly lower in Control Condition 1 than in Control Condition 2, $t(286) = 2.51$, $p = .01$. Control Condition 1 differed significantly from the high-arousal groups, $t(286) = 4.43$, $p < .0001$, but not from the low-arousal groups. Control Condition

2 differed significantly from the low-arousal groups, $t(286) = 2.86$, $p = .004$, but not from the high-arousal groups. The same analyses on the rating scores revealed analogous results, with the important exception that for this measure the serene and sad conditions did show comparable arousal scores (see Table 4).

Free recall performance

Table 5 presents recall data for studied list words and critical lures in all six conditions. Overall, studied words were correctly recalled with a probability of .65, and critical lures were falsely recalled with a probability of .23. Noncritical intrusions were extremely infrequent ($M = .2$ words/list, so on average only two words throughout the whole experiment). Again, data were analysed in two ways. A one-way ANOVA with Mood as a between-subjects variable was carried out to evaluate its general effect on veridical and false recall. In addition, both 2×2 and 3×2 factorial analyses were used to evaluate the effects of valence and arousal more directly.

As in Experiment 1 and in previous studies, the mood groups did not differ with respect to veridical memory. None of the analyses on the proportions of correct recall revealed any significant effects ($Fs \approx 1$). Hence, participants in all six conditions were equally good at recalling studied list words, regardless of their mood during the study phase (see Table 5). In addition, none of the analyses on the proportions of false recall (both one-way and two-way) revealed any significant effects ($Fs < 1$). Thus, neither valence nor arousal affected participants' tendency to falsely recall critical lures. Taking into account that the mood induction might have worn out by the end of the DRM procedure, analyses were repeated considering only the first half of the word lists. False recall still did not differ between groups.

Table 5. Proportions of items recalled in Experiment 2, as a function of Mood

Mood	List words	Critical lures
Serene	.64 (.01)	.22 (.02)
Happy	.66 (.01)	.23 (.02)
Sad	.65 (.01)	.25 (.02)
Angry	.65 (.01)	.23 (.02)
Control 1	.66 (.01)	.21 (.02)
Control 2	.63 (.01)	.23 (.02)

Note: Standard errors of the mean are given in parentheses.

Summary

The mood induction was effective. Valence scores were significantly higher for the positive groups than for the negative groups, with control conditions obtaining intermediate values. Arousal scores were significantly higher for the high-

arousal groups than for the low-arousal groups. In addition, strong support was obtained for the hypothesis that a neutral mood induction is relaxing, as subjective arousal was significantly lowered by this type of induction, and there was a significant post-induction difference between Control Condition 1 and Control Condition 2. In accordance, the former differed from the high-arousal mood groups but not from the low-arousal mood groups, whereas the latter differed from the low-arousal mood groups but not from the high-arousal mood groups.

False recall of critical lures did not show any effect of mood. Based on the fact that the effect of arousal in Experiment 1 was elicited mainly by differences in response bias and free recall is a test in which no cues are given, it was predicted that the effect would be smaller for free recall than for old/new recognition. However, an alternative explanation might be found in the different timing of the tests. Evidence exists that the effect of arousal on memory performance is different for immediate and delayed tests (see Revelle & Loftus, 1990, 1992, for a review), in that it might only facilitate long-term recollection.

EXPERIMENT 3

The goal of Experiment 3 was to investigate whether the difference in results between Experiments 1 and 2 was due to the *type* of test (cues vs. no cues) or the *timing* of the tests (delayed vs. immediately after each list). This was done by using immediate recognition tests. They were composed and administered in the same way as the delayed recognition test in Experiment 1, but shared the timing and the amount of information tested with the immediate free recall tests in Experiment 2.

Method

Participants
A total of 297 undergraduate students (14% male, 86% female) participated in the experiment in exchange for course credit. Their average age was 18.55 years ($SD = 1.18$; range 17–29).

Materials
The same word lists were used as in Experiments 1 and 2. One set of 10 lists was to be studied and the other served as baseline, counterbalanced across participants and conditions. Each recognition test was composed of eight words, which were randomly presented: three list targets (1st, 8th, 10th item from preceding study list), three list distractors (1st, 8th, 10th item from nonstudied list), one critical target (related to preceding study list), and one critical distractor (related to nonstudied list). Test items were visually presented, using black uppercase letters in the centre of the screen.

Design and procedure
Participants were randomly assigned to one of the six mood conditions: serene ($n = 50$), happy ($n = 50$), sad ($n = 50$), angry ($n = 50$), control 1 ($n = 50$), and control 2 ($n = 47$). The procedure was the same as in Experiment 2, with the exception that recognition tests were administered immediately after each list instead of free recall. Participants were asked to indicate for each item whether or not it had been part of the preceding list.

Results and discussion

Data of two participants (both in the happy condition) were omitted prior to the analyses due to a failure to comply with the instructions.

Mood manipulation check
At the start of the experiment, there were no significant group differences in either valence or arousal. From the first to the second administration of the BMIS and rating scales, scores changed in accordance with the induced moods. The only exception were BMIS arousal scores in the sad condition, which stayed constant. Nevertheless, the rating scores did change in the expected direction. As in the previous two experiments, subjective arousal was lowered by the neutral mood induction, measured by both the BMIS, $F(1, 242) = 3.01$, $p = .08$, and the rating scale, $F(1, 240) = 6.38$, $p = .01$.

Since the post-induction mood difference was a precondition for testing the effect of mood on

false recognition, severely outlying values were again removed from the dataset. Seven individuals showed valence and/or arousal scores that deviated strongly (>2.5 SDs) from their condition mean and were left out.[6] Table 6 presents averages for the resulting participants. After the induction phase, the six conditions were reliably different in both valence and arousal. This was the case when considering BMIS scores, $F(5, 282) = 24.20$, $MSE = 121.22$, $p < .0001$, $\eta^2 = .30$ (valence), $F(5, 282) = 7.59$, $MSE = 50.09$, $p < .0001$, $\eta^2 = .12$ (arousal), as well as rating scores, $F(5, 280)^7 = 27.48$, $MSE = 257.36$, $p < .0001$, $\eta^2 = .33$ (valence), $F(5, 280) = 5.09$, $MSE = 494.55$, $p = .0002$, $\eta^2 = .08$ (arousal).

Planned comparisons (Dunn) on the BMIS scores revealed that valence was significantly higher for the two positive groups (serene and happy) than for the two negative groups (sad and angry), $t(282) = 10.49$, $p < .0001$. Control groups showed intermediate values, which differed significantly from those of the negative groups, $t(282) = 7.01$, $p < .0001$, and from those of the positive groups, $t(282) = 3.45$, $p = .0007$. The same analyses on the rating scores revealed comparable results (see Table 6) and are therefore not reported.

BMIS arousal scores were significantly higher for the two high-arousal groups (happy and angry) than for the two low-arousal groups (serene and sad), $t(282) = 5.56$, $p < .0001$. Planned comparisons on the rating scores revealed comparable results. Arousal scores in the two control conditions were virtually identical (see Table 6). Thus, although the neutral mood induction did lower arousal ratings in Control Condition 1 (cf. supra), this did not create a difference with Control Condition 2, as subjective arousal in the latter was quite low in itself.

Recognition performance

Recognition data for list words and critical lures in all six conditions are presented in Table 7. The average probability of correctly recognizing list targets was .86, and the average probability of falsely recognizing critical targets was .69. False alarms to unrelated distractor words were rare, as could be expected in an immediate test (.01 for both types of distractors). As in the previous experiments, both a one-way ANOVA in function of mood and a Valence × Arousal factorial analysis were carried out. Since arousal ratings in the two control conditions did not differ, only the four critical mood conditions (i.e., serene, happy, sad, angry) could be taken into account in the latter analysis.

As before, there was no effect of mood, nor of valence or arousal in particular, on veridical recognition ($Fs < 1$). The result pattern for false recognition was similar to that in Experiment 1. The one-way effect of Mood on the proportions of "old" responses to critical targets was not significant, but the main effect of arousal in a Valence × Arousal factorial analysis was marginally significant, $F(1, 189) = 2.43$, $MSE = .07$, $p = .12$, $\eta_p^2 = .01$. In line with Experiment 1, there was a tendency towards increased false recognition with low arousal as compared to high arousal ($M = .71$ vs. .65). As can be seen in Tables 6 and 7, the group with the highest arousal score (angry) showed the lowest false-recognition rate. There was no evidence for a main effect of valence, nor for an interaction between valence and arousal ($Fs < 1$).

Signal detection measures

As in Experiment 1, A' was used as a measure of discriminability, and B_D'' was used as a measure of response bias. Both indexes were again calculated in three ways. Estimates are presented in Table 8.

Sensitivity measures showed that participants were extremely good at distinguishing list targets from unrelated list distractors (IM-1: $M = .95$), but less good at distinguishing list targets from related critical targets (IM-2: $M = .63$). Nevertheless, unlike in Experiment 1, all conditions still performed above chance when making the latter

---

[6] Two in the serene condition, one in the happy condition, two in the sad condition, one in Control Condition 1, and one in Control Condition 2.

[7] Two participants forgot to fill in the rating scales after the mood induction, but did complete the BMIS.

Table 6. Average valence and arousal scores in Experiment 3, before (1) and after (2) the mood induction

Mood	Valence				Arousal			
	BMIS (1)	*BMIS (2)*	*Rating (1)*	*Rating (2)*	*BMIS (1)*	*BMIS (2)*	*Rating (1)*	*Rating (2)*
Serene	16.27 (1.78)	19.58 (1.44)	70.08 (2.16)	73.40 (1.87)	25.23 (0.96)	21.25 (0.90)	51.04 (3.10)	38.23 (3.23)
Happy	14.40 (1.39)	20.70 (1.32)	70.66 (2.18)	77.15 (2.00)	25.72 (0.90)	27.89 (1.07)	39.09 (3.11)	50.70 (3.93)
Sad	12.94 (2.09)	2.50 (1.88)	66.04 (2.79)	46.00 (2.75)	24.29 (0.86)	24.56 (0.87)	39.73 (3.15)	33.63 (2.58)
Angry	14.62 (1.61)	4.52 (1.94)	66.38 (2.32)	55.22 (2.80)	24.30 (0.88)	29.26 (1.18)	47.96 (3.09)	52.73 (3.57)
Control 1	13.98 (1.69)	17.27 (1.50)	68.69 (2.53)	71.94 (2.44)	26.06 (1.05)	24.92 (1.04)	49.44 (3.03)	43.80 (3.03)
Control 2	12.00 (1.25)	12.00 (1.25)	69.09 (1.75)	69.09 (1.75)	25.54 (1.02)	25.54 (1.02)	44.00 (2.77)	44.00 (2.77)

Note: BMIS = Brief Mood Introspection Scale. In Control Condition 2, the BMIS/rating scales were administered only once, so the same value was used for group comparisons before and after the mood induction. Standard errors of the mean are given in parentheses.

Table 7. Proportions of items judged as old in Experiment 3, as a function of Mood

	List words		Critical lures	
Mood	Targets	Distractors	Targets	Distractors
Serene	.85 (.01)	.01 (.00)	.71 (.03)	.00 (.00)
Happy	.85 (.02)	.01 (.00)	.68 (.04)	.01 (.00)
Sad	.87 (.02)	.01 (.00)	.71 (.04)	.00 (.00)
Angry	.86 (.01)	.01 (.00)	.62 (.04)	.00 (.00)
Control 1	.85 (.01)	.01 (.00)	.70 (.04)	.01 (.01)
Control 2	.86 (.01)	.01 (.00)	.71 (.04)	.01 (.00)

Note: Standard errors of the mean are given in parentheses.

distinction (see Table 8). This difference between experiments could be expected based on the difference in the delay of testing. The consistently high level of A' when treating "old" responses to critical targets as hits ($M = .89$; i.e., strong reliance on gist memory) and the liberal response bias used for the comparison between critical targets and list targets (as opposed to the conservative bias for the distinction between critical/list targets and unrelated distractors) both indicate that the semantic illusion elicited was still quite strong.

As in Experiment 1, a 2 (valence) × 2 (arousal) analysis of variance on A' for the distinction between list targets and critical targets (IM-2) revealed a marginally significant main effect of arousal, $F(1, 189) = 2.72$, $MSE = .04$, $p = .10$, $\eta_p^2 = .01$. Sensitivity estimates were again higher for participants in the high-arousal groups than for participants in the low-arousal groups,

suggesting that higher arousal enhanced item-specific memory and improved participants' ability to distinguish between studied words and related lure words. The same analysis on A' for the distinction between critical targets and distractors (GM) also revealed a marginally significant main effect of arousal, $F(1, 189) = 3.39$, $MSE = .01$, $p = .067$, $\eta_p^2 = .02$, but in the opposite direction. Low-arousal participants showed higher sensitivity for critical targets, and therefore stronger reliance on gist memory, than high-arousal participants did.

As in Experiment 1, response bias was more liberal in the low-arousal groups than in the high-arousal groups, both for the distinction between list targets and critical targets (IM-2) and for the distinction between critical targets and distractors (GM; see Table 8). The effect was not as strong as that in Experiment 1, however, as 2 × 2 analyses of variance revealed a marginally significant main effect of arousal only for the former, $F(1, 189) = 3.50$, $MSE = .09$, $p = .06$, $\eta_p^2 = .01$. The nonsignificance of the latter was most likely due to the extremely low proportions of false alarms to critical distractors in the present experiment.

Summary

As before, the mood induction was effective. In addition, subjective arousal scores were lowered by the neutral mood induction (but there was no post-induction difference between the two control conditions). In line with Experiment 1, there was a tendency towards increased false recognition of

Table 8. Estimates of sensitivity (A') and response bias (B_D'') in Experiment 3, as a function of Mood

	Item-specific memory 1		Item-specific memory 2		Gist memory	
Mood	A'	B_D''	A'	B_D''	A'	B_D''
Serene	.95 (.00)	.68 (.03)	.61 (.03)	−.82 (.02)	.90 (.01)	.65 (.05)
Happy	.95 (.00)	.59 (.06)	.64 (.03)	−.73 (.05)	.88 (.01)	.65 (.05)
Sad	.96 (.00)	.64 (.04)	.62 (.03)	−.79 (.04)	.90 (.01)	.60 (.05)
Angry	.96 (.00)	.69 (.03)	.68 (.02)	−.71 (.05)	.87 (.01)	.72 (.04)
Control 1	.95 (.00)	.66 (.04)	.62 (.03)	−.73 (.05)	.89 (.01)	.57 (.06)
Control 2	.95 (.00)	.68 (.04)	.61 (.03)	−.80 (.04)	.89 (.01)	.62 (.05)

Note: Item-specific memory 1 = hits compared to unrelated false alarms; Item-specific memory 2 = hits compared to related false alarms; Gist memory = related false alarms compared to unrelated false alarms. Standard errors of the mean are given in parentheses.

critical lures with low arousal as compared to high arousal. This is consistent with Hypothesis 3, although it is important to note that the effect was only marginally significant. Sensitivity estimates revealed two complementary trends (albeit also only marginally significant). High arousal was shown to enhance item-specific memory and to decrease reliance on gist memory. In addition, there was an effect of arousal on participants' response criteria. When comparing related lures to other words, low arousal induced more liberal responding than high arousal did. The effect only approached statistical significance for the comparison between critical lures and studied list words. The nonsignificance of the comparison between critical lures and distractors can most likely be attributed to the fact that it is extremely easy to reject unrelated items in a short, immediate test.

Together, these findings suggest that the difference in results between Experiments 1 and 2 was mainly due to the different test type. Similar effects of arousal on memory were obtained for immediate and delayed recognition, but not for free recall. Nevertheless, the timing of the tests might still play a role as well, since effects were less robust for immediate than for delayed recognition.

GENERAL DISCUSSION

The aim of the present experiments was to clarify the effect of mood on false memory in the DRM paradigm. The few studies carried out so far were not yet able to clearly establish whether valence or arousal is most critical in determining susceptibility to the illusion, nor to identify the underlying processes. Based on previous findings and theories, three alternative hypotheses could be proposed. *First*, based on Storbeck and Clore (2005) and the affect-as-information theory, it could be predicted that false memories would be more likely in conditions conveying positive information about oneself than in conditions conveying negative information. *Second*, based on Corson and Verrier (2007), it could be predicted that false memories would be more likely in high-arousal conditions

than in low-arousal conditions. *Third*, based on evidence that high arousal induces item-specific encoding (Schwartz, 1975) and improves recollection (e.g., Nielson & Powless, 2007), whereas low arousal induces an overgeneral retrieval style (McBride & Cappeliez, 2004), it could be predicted that false memories would be more likely in low-arousal conditions than in high-arousal conditions. To disentangle these hypotheses, both the valence and the level of arousal of participants' mood were manipulated, and memory for DRM lists was evaluated by means of delayed recognition, immediate free recall, and immediate recognition.

The mood induction procedure used (cf. Mayer et al., 1995) was effective in eliciting both positive and negative moods with low and high arousal (see also Jallais & Gilet, 2010). This implies that the translated vignettes for the guided imagery were adequate, and we now have a set of Dutch materials to experimentally induce serene, happy, sad, and angry moods. Two control conditions (no vs. neutral mood induction) were included to investigate whether a neutral induction would have an effect on participants' feeling of arousal. This was indeed the case. In all three experiments, subjective arousal was significantly lowered by the neutral mood induction. This implies that the induction was not entirely neutral, but rather boring or relaxing (cf. Whissell & Levesque, 1988; as cited in Revelle & Loftus, 1990). It can be explained by the fact that the vignettes in the neutral condition contained general descriptions rather than specific situations, making it rather difficult for participants to truly imagine them. Without anything more to do than reading a short sentence, 30 s of exposure to eight of these vignettes was most likely quite boring. It should be acknowledged that mood in the condition *without* mood induction cannot be said to have been entirely neutral either, as participants in this condition evaluated their mood only once, at the start of the experiment. Mood in Control Condition 2 thus rather reflected the natural state in which participants came into the lab. Nevertheless, the finding of differing arousal in Control Conditions 1 and 2 still has two important implications. First, it implies that different

operationalizations might have been responsible for the different result pattern obtained for control participants in Storbeck and Clore (2005; no induction) and Corson and Verrier (2007; neutral induction). Second, in Experiments 1 and 2, it allowed for an investigation of the effect of arousal on false memory not only for positive and negative valence, but also for neutral valence. In Experiment 3, the latter was not possible due to relatively low arousal in the condition without mood induction as well.

Whereas there was no effect of either valence or arousal on false recall, arousal did influence false recognition. The effect was significant when a delayed test was used and marginally significant when short tests were provided immediately after each list. Consistent with Hypothesis 3, participants in low-arousal conditions were more inclined to falsely endorse critical lures than participants in high-arousal conditions, regardless of valence. Based on signal detection analyses, the effect could be attributed to several underlying processes. First and most importantly, there was a difference between low- and high-arousal groups in response bias. When comparing critical targets to other test items, participants in low-arousal conditions (serene, sad, control 1) used more liberal response criteria than participants in the other conditions (happy, angry, control 2) did. This was the case for both immediate and delayed recognition tests, although the effect was strongest with delayed testing. The finding is in line with previous evidence that low arousal induces an overgeneral retrieval style (McBride & Cappeliez, 2004). Second, for both types of recognition test, there was suggestive evidence for an effect of arousal on item-specific memory. Happy and angry participants were better at distinguishing studied words from related lure words than serene and sad participants were. Third, with immediate testing, there was a tendency towards reduced reliance on gist memory by happy and angry participants as compared to serene and sad participants. Although both latter effects were only marginally significant, the trend of increased item-specific memory and decreased gist reliance with higher arousal is also consistent with Hypothesis 3. It is in line with

earlier work showing that high arousal induces verbatim encoding (Schwartz, 1975) and with studies showing a positive effect of arousal on memory (e.g., Nielson & Powless, 2007) and on conscious recollection in particular (Anderson et al., 2006).

Whereas the effect of arousal on recollection was weak, the effect of arousal on response bias was more robust. When feeling calm, participants were more inclined to endorse a semantically related item on a recognition test. Importantly, this effect was also obtained when the two control conditions were included in the analyses. This suggests that a relaxed state, even when elicited by boredom, induces liberal responding. Feeling relaxed apparently leads people to accept doubtful items more easily, especially when taking a long, delayed test. In that sense, the relatively stronger results for delayed than for immediate recognition might possibly be due to the difference in duration and therefore dullness of the study and test phases. In Experiment 1, participants listened to 10 lists of 15 words in a row, separated by each other only by some relatively easy-to-solve math problems. Next, the memory test comprised a long series of 80 words, presented one by one on the computer screen. In both Experiments 2 and 3, study and test phases were alternated, providing more variety. The long duration and hence dullness of the two phases in Experiment 1 might have intensified participants' arousal levels. In the low-arousal conditions, participants who already felt relaxed might have gotten more relaxed out of boredom. In the high-arousal conditions, participants who already felt aroused might have gotten annoyed, further increasing their arousal level. In follow-up studies, subjective arousal should be assessed both prior to and after the DRM part in order to evaluate this hypothesis.

As opposed to immediate recognition, immediate free recall of critical lures did not show any effect of mood. Given the only marginally significant effect of arousal on signal detection measures of item-specific and gist memory in Experiments 1 and 3, the influence of arousal on recollection was most likely too weak to affect performance in a test in which intentional recollection is crucial. Arousal-induced group differences in response

bias were more robust, but are trivial in a test in which no cues are given. The weak effect of arousal on recollection in the present study may be at least partly due to the relatively short retention interval in all three experiments. A longer delay between study and test might increase the effect size, as effects of arousal on memory typically increase with longer delays (e.g., LaBar & Phelps, 1998; McGaugh, 2000; Revelle & Loftus, 1992; Sharot & Yonelinas, 2008). An additional explanation might be the fact that in a DRM experiment, participants tend to focus on both item-specific and gist information. They are asked to remember the individual words, but as relational processing is the default encoding strategy (e.g., Hunt & McDaniel, 1993; Van Damme & d'Ydewalle, 2009a), they will also attend to the gist of the lists. Mather and Sutherland (2011) recently suggested that arousal specifically strengthens memory for high-priority information. Priority may be determined by both bottom-up sensory influences and top-down cognitive influences, such as goal relevance (see also Levine & Edelstein, 2009). Thus, in a regular memory experiment, arousal will strengthen attention to and recollection of the information that is most relevant to the goal of memorizing. In the DRM paradigm, it is difficult to establish what is most goal relevant: specific list words or the gist of the lists. In fact, both are. This confusion might have tempered the effect of arousal on recollection. In a follow-up study, we plan to investigate this by using more detailed study instructions, focusing on either item-specific or relational encoding.

The present findings are consistent with Hypothesis 3, but completely opposite to Hypothesis 2, which was based on Corson and Verrier's (2007) results. It is not entirely clear what caused the failure to replicate their findings. Several methodological aspects might have contributed to it: the composition of the recognition test (separate distractor items for list words and critical lures vs. one set of unrelated words), the way in which test items were presented (one by one vs. in two columns, allowing participants to scan back and forth), subtle differences in the mood induction procedure (translation of the vignettes),

and differences in the mood evaluation procedure (different use of the BMIS; matrix vs. rating scales). It seems rather unlikely that any of these differences was responsible for a huge difference in results. It might be the case, however, that the methodological differences contributed to differences in the induced moods and the subjective arousal levels obtained. It is possible that higher levels of arousal than the ones obtained in the present study would lead to a drop in performance and hence more false memories (cf. Yerkes-Dodson law, Yerkes & Dodson, 1908; see also Baldi & Bucherelli, 2005; Deffenbacher, Bornstein, Penrod, & McGorty, 2004). Further research is needed to determine whether different degrees of "high" arousal have different effects on the DRM illusion. It is noteworthy, however, that recent studies investigating the effect of psychosocial stress on false memory also obtained inconsistent findings (i.e., Smeets et al., 2006: *better source monitoring*; Smeets, Jelicic, & Merckelbach, 2006, and Smeets, Otgaar, Candel, & Wolf, 2008: *no effect*; Payne, Nadel, Allen, Thomas, & Jacobs, 2002: *increased false recognition*). Second, because valence and arousal were not evaluated at the end of the study, it might be that the moods induced by Corson and Verrier had worn out by the time of the recognition test. This would have excluded any effects of arousal on response bias, which were in fact the strongest effects found in the present study (however, see Storbeck & Clore, 2011). Finally, it is possible that Corson and Verrier's participants had a stronger relational focus during encoding, turning the gist of the lists into high-priority information. Future studies will need to distinguish between these alternative explanations.

The present results also did not support Hypothesis 1, which was based on the affect-as-information theory. This can be explained by the theory itself, however. According to Clore and Palmer (2009; see also Clore & Huntsinger, 2007), participants' positive/negative feelings will exert an influence on performance only when they consider these feelings as task-relevant feedback (i.e., informative to what they are doing: I feel bad, so this probably means that I'm doing badly

and therefore need to concentrate better). The effect disappears if feelings can be attributed to an alternative source (see also Schwarz & Clore, 1983). Storbeck and Clore (2005), whose results supported the affect-as-information theory, induced positive and negative moods by means of music, used a cover story, and evaluated moods with just one sentence at the end of the experiment. Clearly, the current procedure was more transparent. Participants were asked to imagine emotional situations as vividly as possible, and a whole mood questionnaire was to be filled out both before and after. Hence, it was much more likely that participants attributed their changed mood to the induction phase and therefore did not experience it as task-relevant feedback. This implies that the present findings should be nuanced as follows: Participants who were aware of how they felt and of what caused it, showed more false recognition in low-arousal moods than in high-arousal moods.

To conclude, the present results are consistent with studies showing that low arousal induces an overgeneral retrieval style (McBride & Cappeliez, 2004), and that high arousal induces verbatim processing (Schwartz, 1975) and improves memory (e.g., Nielson & Powless, 2007). The effect on response bias was most robust, with more liberal responding in a relaxed state than in an aroused state. Low subjective arousal seems to increase the tendency to accept doubtful information, especially when taking a long, boring test. The relatively weak effect of arousal on recollection might be attributed to the relatively short study–test delay and to uncertainty about whether specific list items or rather the gist of the lists should have highest priority (cf. Mather & Sutherland, 2011). Due to the fact that mood was manipulated prior to the encoding phase (cf. Corson & Verrier, 2007; Storbeck & Clore, 2005), it is difficult to determine where the most important effects took place. Effects on information processing are confounded with effects on consolidation and retrieval. In accordance with the present results, English and Nielson (2010) recently found reduced susceptibility to the misinformation effect (i.e., memory distortion due to misleading postevent information; Loftus, 2005)

with high arousal. After a 1-week delay, false memories were less likely when a highly arousing stimulus had followed the provision of misleading information. Thus, although arousal was only induced after the encoding phase, it reduced source confusion in a similar way as it did in the present experiments. Future studies are needed to fine-tune the importance of various underlying processes and to separate effects at the encoding, consolidation, and retrieval stages. Important aspects to consider will be the study instructions used, regular arousal assessments, and a fine-grained discrimination between various arousal levels.

REFERENCES

Adolphs, R., Denburg, N. L., & Tranel, D. (2001). The amygdala's role in long-term declarative memory for gist and detail. *Behavioral Neuroscience*, 115, 983–992.

Anderson, A. K., Wais, P. E., & Gabrieli, J. D. E. (2006). Emotion enhances remembrance of neutral events past. *Proceedings of the National Academy of Sciences, USA*, 103, 1599–1604.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database [CD-ROM]. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

Baldi, E., & Bucherelli, C. (2005). The inverted "U-shaped" dose–effect relationships in learning and memory: Modulation of arousal and consolidation. *Nonlinearity in Biology, Toxicology, and Medicine*, 3, 9–21.

Bless, H., & Fiedler, K. (2006). Mood and the regulation of information processing and behavior. In J.P. Forgas (Ed.), *Affect in social thinking and behavior* (pp. 65–84). New York, NY: Psychology Press.

Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to "remember" than things that were. *Psychological Science*, 9, 484–489.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, 5, 164–169.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York, NY: Oxford University Press.

Christianson, S., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition and Emotion*, *5*, 81–108.

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, *11*, 393–399.

Clore, G. L., & Palmer, J. (2009). Affective guidance of intelligent agents: How emotion controls cognition. *Cognitive Systems Research*, *10*, 21–30.

Corson, Y., & Verrier, N. (2007). Emotions and false memories: Valence or arousal? *Psychological Science*, *18*, 208–211.

De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*, 198–205.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, *28*, 687–706.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*, 275–277.

English, S. M., & Nielson, K. A. (2010). Reduction of the misinformation effect by arousal induced after learning. *Cognition*, *117*, 237–242.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model. *Psychological Bulletin*, *117*, 39–66.

Forgas, J. P., Laham, S. M., & Vargas, P. T. (2005). Mood effects on eyewitness memory: Affective influences on susceptibility to misinformation. *Journal of Experimental Social Psychology*, *41*, 574–588.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833–848.

Gerrards-Hesse, A., Spies, K., & Hesse, F. W. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, *85*, 55–78.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*, 424–429.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*, 497–514.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*, 421–445.

Jallais, C., & Gilet, A. (2010). Inducing changes in arousal and valence: Comparison of two mood induction procedures. *Behavior Research Methods*, *42*, 318–325.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*, 555–583.

LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science*, *9*, 490–493.

Lampinen, J. M., Leding, J. K., Reed, K. B., & Odegard, T. N. (2006). Global gist extraction in children and adults. *Memory*, *14*, 952–964.

Levine, L. J., & Edelstein, R. S. (2009). Emotion and memory narrowing: A review and goal-relevance approach. *Cognition and Emotion*, *23*, 833–875.

Loftus, E. F. (2005). Planting misleading information in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361–366.

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*, 33–52.

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science*, *6*, 114–133.

Mayer, J. D., Allen, J. P., & Beauregard, K. (1995). Mood inductions for four specific moods: A procedure employing guided imagery vignettes with music. *Journal of Mental Imagery*, *19*, 133–150.

Mayer, J. D., & Gaschke, Y. N. (1988). The experience and meta-experience of mood. *Journal of Personality and Social Psychology*, *55*, 102–111.

McBride, C., & Cappeliez, P. (2004). Effects of manipulating valence and arousal components of mood on specificity of autobiographical memory. *Psychological Reports*, *95*, 615–630.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language*, *35*, 212–230.

McGaugh, J. L. (2000). Memory—A century of consolidation. *Science*, *287*, 248–251.

Nielson, K. A., & Powless, M. (2007). Positive and negative sources of emotional arousal enhance long-term word-list retention when induced as long as 30 min after learning. *Neurobiology of Learning and Memory*, *88*, 40–47.

Payne, J. D., Nadel, L., Allen, J. J. B., Thomas, K. G. F., & Jacobs, W. J. (2002). The effects of experimentally induced stress on false recognition. *Memory*, *10*, 1–6.

Revelle, W., & Loftus, D. A. (1990). Individual differences and arousal: Implications for the study of mood and memory. *Cognition and Emotion*, *4*, 209–237.

Revelle, W., & Loftus, D. A. (1992). The implications of arousal effects for the study of affect and memory. In S.-A. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 113–149). Hillsdale, NJ: Lawrence Erlbaum Associates.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.

Roediger, H. L., III, & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science*, *9*, 123–127.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. New York, NY: Houghton Mifflin.

Schacter, D. L., Verfaellie, M., Anes, M. D., & Racine, C. (1998). When true recognition suppresses false recognition: Evidence from amnesic patients. *Journal of Cognitive Neuroscience*, *10*, 668–679.

Schwartz, S. (1975). Individual differences in cognition: Some relationships between personality and memory. *Journal of Research in Personality*, *9*, 217–225.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*, 513–523.

Sharot, T., & Yonelinas, A. P. (2008). Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition*, *106*, 538–547.

Smeets, T., Jelicic, M., & Merckelbach, H. (2006). Stress-induced cortisol-responses, sex differences, and false recollections in a DRM paradigm. *Biological Psychology*, *72*, 164–172.

Smeets, T., Jelicic, M., Merckelbach, H., Peters, M., Fett, A., Taverniers, J., et al. (2006). Enhanced memory performance on an internal–internal source monitoring test following acute psychosocial stress. *Behavioral Neuroscience*, *120*, 1204–1210.

Smeets, T., Otgaar, H., Candel, I., & Wolf, O. T. (2008). True or false? Memory is differentially affected by stress-induced cortisol elevations and sympathetic activity at consolidation and retrieval. *Psychoneuroendocrinology*, *33*, 1378–1386.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.

Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory. *Psychological Science*, *16*, 785–791.

Storbeck, J., & Clore, G. L. (2011). Affect influences false memories at encoding: Evidence from recognition data. *Emotion*, *11*, 981–989.

Tussing, A. A., & Greene, R. L. (1997). False recognition of associates: How robust is the effect? *Psychonomic Bulletin & Review*, *4*, 572–576.

Van Damme, I., & d'Ydewalle, G. (2009a). Implicit false memory in the DRM paradigm: Effects of amnesia, encoding instructions, and encoding duration. *Neuropsychology*, *23*, 635–648.

Van Damme, I., & d'Ydewalle, G. (2009b). Memory loss versus memory distortion: The role of encoding and retrieval deficits in Korsakoff patients' false memories. *Memory*, *17*, 349–366.

Wright, D. B., Startup, H. M., & Mathews, S. A. (2005). Mood, dissociation and false memories using the Deese–Roediger–McDermott procedure. *British Journal of Psychology*, *96*, 283–293.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482.