Journal of Experimental Psychology: General
1994, Vol. 123, No. 2, 201-215

0096-3445/94/$3.00

Mood Dependent Memory for Events of the Personal Past

Eric Eich, Dawn Macaulay, and Lee Ryan

Previous research on mood dependent memory (MDM) suggests that the more one must rely on internal resources, rather than on external aids, to generate both the target events and the cues required for their retrieval, the more likely is one's memory for these events to be mood dependent. To instantiate this "do-it-yourself" principle, three experiments were conducted in which Ss experiencing either a pleasant or an unpleasant mood generated autobiographical events in response to neutral nouns. Subsequently, Ss were tested for event free recall while in the same or the alternative mood state. All three studies showed MDM, such that the likelihood of recalling an event generated 2 or 3 days ago was higher when generation and recall moods matched than when they mismatched. Prospects for future research aimed at elucidating and extending these results are discussed.

During the past decade, many attempts have been made to demonstrate mood dependent memory (MDM) through a diversity of experimental methods and designs (see Blaney, 1986; Bower, 1981; Ucros, 1989). Although some have succeeded in showing that events encoded in a certain state of affect or mood are most retrievable in that state, others have failed. Moreover, efforts to replicate positive results have rarely prevailed, even when undertaken by the same investigator using similar procedures (e.g., Bower & Mayer, 1985, 1989). To date, then, the task of demonstrating mood dependence has proved as maddeningly difficult as trying to catch the wind in one's hands: One might capture it momentarily but cannot contain it for long.

The problem of unreliability that now plagues research on MDM, although serious, may not be insurmountable. Echoing the opinion of others (e.g., Bower, 1987; Kihlstrom, 1989), we suspect that the problem arises not because mood dependence is not real but because the phenomenon is selective in nature. Instead of being spread uniformly across all types of memory materials, tasks, or functions, mood dependent effects may emerge in a clear and consistent manner only under certain circumstances. Just what these circumstances are is not yet known, but past research provides a number of promising leads. The main aim of the present research is to pursue these leads, with a view to engineering an experimental situation in which robust and reliable evidence of MDM is apt to appear.

Arguably the strongest lead suggested by earlier experiments relates to the way in which memory for the to-be-remembered or target events is tested. By several accounts (Bower, 1981; Eich, 1980; Ucros, 1989), MDM is more likely to materialize when retrieval is mediated by "invisible" cues produced by the subject than by "observable" cues provided by the experimenter. Thus, free recall is a more sensitive measure of mood dependence than are either cued recall or old/new recognition memory. Given that we sought to stack the deck in favor of finding MDM, we chose free recall as the test of choice in all three of the studies reported here.

A second clue to detecting mood dependence is, in effect, the complement of the first. Just as the odds of demonstrating MDM may be improved by requiring subjects to generate their own cues for retrieving the target events, so too may these prospects be enhanced by requiring subjects to generate the events themselves. In support of this proposition, Eich and Metcalfe (1989) observed (in three of four studies) a significantly greater mood dependent effect in the free recall of verbal items that subjects had actively generated (e.g., *guitar,* elicited by the cue *musical instruments: banjo—g*), in contrast to items that the subjects had simply read (e.g., *gold,* embedded in the phrase *precious metals: silver—gold*). This was so regardless of whether the overall level of generate item recall was higher than that of the read items—the prototypic "generation effect" (Slamecka & Graf, 1978)—or whether it was lower (as occurred when subjects read some targets three times and generated others only once). Moreover, and in line with remarks made in the preceding paragraph, the results of a test of old/new recognition memory, which was administered soon after recall, revealed no evidence of MDM for either type of target. Thus, it seems that the more one must rely on internal resources, rather than on external aids, to generate both the cues required to effect retrieval of the target events and the events themselves, the more likely is one's memory for these events to be mood dependent.

Eric Eich, Dawn Macaulay, and Lee Ryan, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada.

Preparation of this article was aided by grants from the National Institute of Mental Health (MH55052) and the Natural Sciences and Engineering Research Council (37335). Our thanks to Gordon Bower, Jennifer Campbell, and an anonymous reviewer for their cogent comments and criticisms and to Tanya Anthony, Christine Beck, Andrea Bull, Kirsten Campbell, Dawn Layzell, Colleen Quee, Kirsten Rose, and Reiko West for their expert experimental assistance.

Correspondence concerning this article should be addressed to Eric Eich, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4.

Our research was designed with this "do-it-yourself" principle in mind. As noted earlier, in all three studies we used a retrieval task—free recall—that should prove sensitive to the detection of MDM, given its reliance on *internal* (subject-produced) as opposed to *external* (experimenter-provided) cues. In addition, in all three studies we used an encoding task—autobiographical memory generation—that should prove to be similarly sensitive, given that it places a premium not on the automatic or data-driven perception of *external* events but on *internal* mental processes such as reasoning, imagination, or reflection (Johnson & Multhaup, 1992; Johnson & Raye, 1981; Johnson & Sherman, 1990).

The task at issue involved the recollection, description, and rating of real-life episodes, and it was performed while subjects were experiencing an experimentally engendered mood of either pleasure (P) or displeasure (U). Two different versions of the task—constrained and unconstrained—were used. Participants in Experiments 2 and 3 undertook the latter version, whereby they were asked to recollect or generate specific events from their personal past when prompted with common-word probes, such as *ship* and *street.* The task was unconstrained in that the emotional valence of every generated event—either positive, neutral, or negative—was determined by the subjects themselves. By contrast, participants in Experiment 1 completed the constrained version, whereby they were enjoined by the experimenter to generate positive events in response to certain probes and negative events in response to others.

Whichever version of the task was used, subjects had ample opportunity (up to 2 min) in which to generate a given event. The time they took to do so was recorded, as was their spoken account of what had happened, when and where, and to whom. After describing the event, subjects rated it along several dimensions, including its intensity, importance, and vividness.

There were two reasons for doing the task this way. One was to replicate and extend prior research relating to mood *congruent* memory: the "enhanced encoding and/or retrieval of material the affective valence of which is congruent with ongoing mood" (Blaney, 1986, p. 229). To clarify, consider the results of a study by Teasdale and Fogarty (1979), in which elated and depressed moods were induced by instructing subjects to internalize the emotion implied by a series of self-referential statements, the technique devised by Velten (1968). Using a constrained test of autobiographical memory, Teasdale and Fogarty found that although the mean latency to generate negative events was constant across mood conditions, positive events were retrieved more rapidly on average in the elated than in the depressed state. Our research provides an opportunity both to replicate these results and to pose a number of novel questions concerning mood congruence. For instance, is greater importance ascribed to positive than to negative autobiographical memories generated during a pleasant mood, and do people experiencing an unpleasant affect recollect negative events more vividly than positive ones?

The second, more pressing reason was to promote the occurrence of mood *dependent* memory. Although the subjects did not know it at the time, the autobiographical memories that they generated today (during the encoding session) would serve as the targets of a surprise test of free recall given 2 or more days later (during the retrieval session). Thus, on returning to the laboratory, subjects were reminded that they had previously generated events from their personal past when primed with common-word probes. Following the induction of a mood that either did or did not match the one they had experienced earlier, subjects were asked to recall as many of these events as possible, in any preferred order, *without benefit of any observable reminders or cues.*

This last expression merits emphasis for two reasons. First, as noted before, uncued tests of retention, such as free recall, are especially sensitive to the detection of mood dependent effects. Second, had we cued the subjects with the original word probes and asked for the recall of the same autobiographical events that they had generated earlier, it would have been unclear whether they were actually recalling these events (in an episodic-memory sense) or whether they were simply regenerating episodes or experiences that they had generated before (in a semantic-memory sense). To sidestep this potential problem in interpretation, we deliberately did *not* cue the subjects; rather, we asked them to recall the target events in any order, without the aid of any tangible reminders.

Reasoning that the initial task of generating, describing, and rating autobiographical memories would have drawn heavily on the subjects' internal resources, we surmised that performance on the subsequent task of free recall would manifest MDM. That is, more events would be recalled when encoding and retrieval moods matched (Conditions P/P and U/U) than when they mismatched (Conditions P/U and U/P).

Until now, the focus of discussion has been on ways of enhancing the sensitivity of an encoding or a retrieval task to the detection of mood dependence. It stands to reason, however, that regardless of how sensitive these tasks may be, the chances of demonstrating MDM are slim in the absence of a potent manipulation of mood. In particular, it is hard to imagine how anything less than a substantial shift in mood, between the occasions of event encoding and event retrieval, could lead to a significant loss of memory. In fact, the results of an MDM meta-analysis by Ucros (1989) suggested that the greater the difference in moods—depression versus elation, for example, as opposed to depression versus a neutral affect—the greater the mood dependent effect.

By the same token, it seems implausible to expect MDM to emerge if, for one reason or another, the mood established at the start of a given encoding or retrieval task should cease to exist at its end.[1] Indeed, such a dissipation

[1] These reasons include the possibility that (a) the mood simply decays with the passage of time; (b) the cognitive effort expended in performing the task at hand costs subjects their capacity to maintain the mood; and (c) in the case of an induced state of depression, subjects engage in antidepressive control processes intended to alleviate or repair their mood (see Blaney, 1986; Isen, 1984).

of mood might have undermined Eich and Metcalfe's (1989, p. 454) attempt to demonstrate MDM in a pilot study involving Velten-induced moods, moods that started strong but faded quickly. The point, then, is that mood dependent effects may depend for their expression not only on the sensitivity of the tasks designed to detect them, but also on the strength and stability of the moods in which these tasks are performed.

Given that strong and stable moods are keys to demonstrating MDM, how can such moods be induced? In the current research, an answer was sought through the use of the continuous music technique (CMT) devised by Eich and Metcalfe (1989). By this technique, subjects are asked to entertain elating or depressing thoughts while listening to various selections of merry or melancholic music (which, once started, do not stop until the end of the session).

Periodically, subjects rated their current levels of both pleasure/displeasure and arousal/sleepiness, the two bipolar dimensions underlying the circumplex model of mood constructed by Russell and his colleagues (Russell, 1980; Russell, Weiss, & Mendelsohn, 1989). Before they were allowed to begin the task of generating autobiographical events (during the encoding session) or the task of recalling these events (during the retrieval session), subjects were required to reach a critical level of mood—at least moderately pleasant or moderately unpleasant in Experiment 1; at least very pleasant or very unpleasant in Experiments 2 and 3—irrespective of their level of arousal. Understandably, subjects were not told in advance about these criteria, and, as one may expect, some took longer than others to attain the critical level of pleasure or displeasure. Thus, rather than allocate to all subjects the same fixed amount of time for mood induction—the near-universal practice in past research on MDM and one that is apt to produce wide variability in pretask levels of mood—we adopted an ideographic approach to mood manipulation, one that permitted each person to develop a predetermined affective state at his or her own pace.

That the CMT is well suited to studying mood dependence is suggested by several considerations. First, the CMT ensures that subjects assigned to either of the two mismatched conditions (P/U or U/P) will experience a substantial shift in mood between the occasions of event encoding and retrieval (viz., from at least moderately pleasant to moderately unpleasant in Experiment 1 and from at least very pleasant to very unpleasant in Experiments 2 and 3). Second, compared with the moods induced by the Velten technique, those instilled by the CMT are more stable over time and across tasks (see Eich & Metcalfe, 1989, Table 1, for pertinent data). Third, although the CMT is primarily intended to alter one's level of pleasure, it appears to affect one's level of arousal as well. As noted by Eich and Metcalfe (1989, Table 2), subjects asked to think sad thoughts while listening to sad music often, but not always, reported lower levels of arousal than did their pleasant-mood counterparts. The importance of this observation lies in that fact that in all four of their studies using the CMT, Eich and Metcalfe found that subjects who had sustained a substantial shift in both pleasure and arousal tended to recall fewer generate items than did subjects who had experienced an equally extreme change in pleasure but no appreciable alteration in arousal. The implication, then, is that stronger mood dependent effects are associated with two-dimensional (pleasure plus arousal) than with one-dimensional (pleasure only) shifts in mood state. Whether the same applies to the free recall of autobiographical events generated days before is an issue of recurring interest in the rest of this article.

Experiment 1

Method

Design. The experiment consisted of two sessions, here called encoding and retrieval, and conformed to a 2 × 2 × 2 mixed design. Whereas both mood at encoding (P vs. U) and mood at retrieval (again, P vs. U) were varied between subjects, the valence or type of autobiographical event generated during the encoding session (either positive or negative) was varied within subjects. The crossing of the first two variables defined four encoding/retrieval conditions, each of which was represented by 16 randomly assigned subjects. Details on the materials, tasks, and procedures involved in the encoding and retrieval sessions are discussed separately in the following sections.

Encoding session. At the start of this session, subjects were told that the study was part of a research program aimed at understanding how moods influence the performance of various cognitive tasks and how the performance of these tasks in turn influences mood. It was explained that the experiment would be divided into two 60- to 90-min sessions (designated simply as "first" and "second"), spaced 2 or 3 days apart. Subjects were also advised that they would be tested individually throughout the course of the experiment and that each session would entail a different set of tasks.

Next, participants were provided with a page bearing the matrix shown in Figure 1, an adaptation of the "affect grid" invented by Russell et al. (1989). Subjects were told that the matrix would be used to measure two aspects of their prevailing mood: level of pleasure (indicated along the horizontal axis of the matrix) and level of arousal (identified by the vertical axis). It was further explained that, reading from left to right, the columns connote a mood that is extremely unpleasant, very unpleasant, moderately unpleasant, slightly unpleasant, neutral (the shaded center square), slightly pleasant, moderately pleasant, very pleasant, and extremely pleasant. Similarly, the rows signify a state that ranges from extremely high arousal at the top through neutral at the shaded center to extremely low arousal at the bottom (with slightly, moderately, and very high or low levels in between). On receiving these instructions (along with the examples cited by Eich & Metcalfe, 1989, p. 445), subjects marked the one square that best exemplified the levels of pleasure and arousal that they were experiencing at that moment.

After making their mark, subjects were apprised that they soon would listen to a selection of classical music that should help them develop a pleasant (unpleasant) mood. It was stressed that because music alone cannot create the desired state, they should concentrate on ideas or images that make them feel pleasant (unpleasant). Subjects were also advised to develop as intense a state as possible and that as an aid in maintaining their mood, the music would continue to play softly in the background. Finally, subjects were informed that the experimenter would return from time to time to check on their progress and to monitor their mood and that when

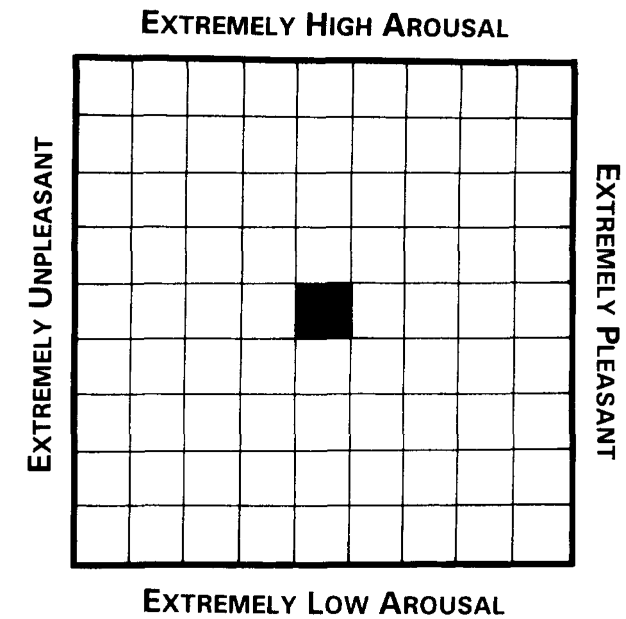

Figure 1. The mood matrix used to measure present levels of pleasure and arousal. From "Affect Grid: A Single-Item Scale of Pleasure and Arousal" by J. A. Russell, A. Weiss, and G. A. Mendelsohn, 1989, *Journal of Personality and Social Psychology, 57,* p. 494. Copyright 1989 by the American Psychological Association. Adapted by permission of the author.

she thought the time was right, they would be asked to perform a series of three tasks: word rating, personality assessment, and autobiographical event generation (the last of which is the focus of discussion in this article).

Participants were then seated in a lounge chair, bordered by stereo speakers. Through these speakers was played, at a comfortable listening volume, one of four cassette tapes, each carrying 60 min of instrumental music. Two of these cassettes contained different selections of "happy" music (excerpts of Beethoven's *Minuet in G,* Boccherini's *Minuet in E,* Mozart's *Divertimento No. 136,* Pachelbel's *Canon in D,* and Vivaldi's *Four Seasons: Spring, Summer, and Fall*), and the other two contained different selections of "sad" pieces (segments of Albinoni's *Adagio in G Minor,* Barber's *Adagio pour Cordes,* Grieg's *Peer Gynt: The Death of Ase,* and Sibelius' *Violin Concerto: Second Movement*). The two tapes representing each type of music were assigned at random, with the proviso that subjects tested under matched mood conditions (P/P or U/U) hear one selection of the appropriate music during the encoding session and the alternate selection during the retrieval session.

Five minutes after music onset, and every 5 min thereafter, subjects marked their current levels of pleasure and arousal on a clean copy of the mood matrix (which, once marked, was promptly collected by the experimenter and never used again). The music continued to play while subjects made these mood ratings and did not stop until the end of the session. On the rare occasions in which subjects took more than 60 min to complete the session, we rewound their tape and replayed it from the beginning.

As noted earlier, subjects understood that they would start the first of three tasks when the experimenter thought the time was right. Depending on whether a pleasant or an unpleasant mood was being induced, the "right time" occurred either (a) when P-mood subjects marked any of the squares included in the two right-most columns of the mood matrix or (b) when U-mood subjects checked any of the squares contained in the two left-most columns. Thus, to advance to the initial task, subjects were required, at a minimum, to rate themselves as feeling either very pleasant or very unpleasant, regardless of their level of arousal. Subjects were not told that the start of the first task was contingent on their achieving a critical level of pleasure or displeasure lest they try to rush matters by rating their mood as being more extreme than it really was.

On reaching the requisite level, subjects were read a list of 16 common unrelated words (such as *banana* and *bride*) and rated each item in terms of its concreteness, imagery, and goodness of meaning on 7-point scales. Subjects were told that the intent of this word rating task was to determine whether people perceive the properties of a word differently depending on their mood. Strictly speaking this was true, but there was more to it than that, as we show in the next section.

After rating the final item, subjects marked a new mood matrix and then began the personality assessment task. The materials for this task consisted of 50 adjectives, displayed one at a time in a random order, representing 25 pairs of trait-descriptive opposites (e.g., *warm, cold; sociable, shy*). Subjects were told that an adjective would appear on a monitor and that they should indicate, as quickly as possible, whether the term is self-descriptive. This task, which was also administered during the retrieval session in the same manner using the same materials, was included in order to test some ideas stemming from Campbell's (1990) work on self-concept clarity. Inasmuch as these ideas are distinct from the issue that chiefly concerns us here—demonstrating mood dependence in the recall of autobiographical events—the results of the personality assessment task, whether administered during the encoding or the retrieval session, are not discussed in this article.

Following the personality assessment task, subjects rated their current levels of pleasure and arousal in the familiar fashion. Those who rated themselves as feeling at least moderately pleasant (in the P-mood condition) or moderately unpleasant (in the U-mood condition) proceeded directly to the task of autobiographical event generation, detailed shortly. Those who did not were asked to concentrate once again on ideas or images that make them feel pleasant (unpleasant). This "mood boost" period lasted 5–10 min, long enough to ensure that all subjects were experiencing at least a moderate degree of pleasure or displeasure prior to the start of the event generation task.

For purposes of this task, subjects were asked to generate an emotionally positive event, from anytime in their personal past, in response to each of 8 word probes and to generate an emotionally negative episode in response to each of 8 additional probes. These two types of probes were read by the experimenter in a randomly alternating order; to avoid item-selection artifacts, any probe that was associated with a positive event for one subject was linked to a negative event for another. All 16 of the probes, plus 2 others that were used for practice, were common, concrete, semantically unrelated, and affectively neutral nouns culled from Brown and Ure's (1969) word norms.

On presentation of a given probe, subjects were instructed to say OK as soon as they generated an event that was both appropriately valenced (either positive or negative, as stipulated by the experimenter) and suitably specific (one they could date with precision and describe in detail). If the subjects failed to generate such an event within 2 min, then that probe was skipped and the next one was read. If they succeeded, then the experimenter, after logging generation latency (in seconds), asked the subjects to recount aloud the particulars of the event: where and when (month and year) it occurred, what happened, who was involved, and so forth. All of these details were transcribed by the experimenter.

Next, subjects rated (a) how intense was the emotion (perforce either positive or negative) they experienced at the time the event transpired; (b) how personally important did the event seem then; and (c) how vividly can they recollect it now. To simplify matters, subjects were provided with a set of index cards containing the pertinent 5-point scales (1 = *not at all intense/important/vivid,* 5 = *extremely intense/important/vivid*). Subjects spoke their ratings, and the experimenter wrote them down.

After generating, recounting, and rating the final (16th) event, subjects marked a new mood matrix. P-mood subjects were then discharged with a reminder to return, in 2 or 3 days, for a second session of testing, details of which would be divulged at that time. U-mood subjects stayed longer—sharing cookies and conversation with the experimenter while rock or reggae music played in the background—so that they left the lab feeling no worse than neutral.

Retrieval session. Procedures involved in manipulating and measuring mood during the retrieval session duplicated those used in the encoding session.

On reporting the requisite rating of mood (viz., at least very pleasant or very unpleasant, irrespective of degree of arousal), subjects began formal testing, which consisted of three tasks: stem completion, personality assessment, and autobiographical event recall. As was the case during the encoding session, subjects recorded their current levels of pleasure and arousal on a new mood matrix both before and after each task.

The subjects' first task was to speak the first word they thought of that began with a particular three-letter stem, such as *ban* or *bri.* Of the 32 stems that were read to the subjects, 16 corresponded to old words (*banana, bride,* and others they had rated for concreteness, imagery, and goodness of meaning 2 or 3 days earlier), and 16 corresponded to new words (ones they had neither seen nor heard in the course of any encoding session task, including word rating). The question of interest here was whether a larger priming effect—operationalized as the difference in completing the stems with old as opposed to new words—would obtain when encoding and retrieval moods matched than when then mismatched. This question was posed as part of an ongoing, multistudy project investigating mood dependence in implicit memory. A preliminary report on this project, including a précis of the stem completion data, was prepared by Macaulay, Ryan, and Eich (1993).

The subjects' second task, personality assessment, used the methods and materials summarized in the preceding section. The results of this task are not reviewed in this article for the reasons mentioned earlier.

At the outset of the third and last task, autobiographical event recall, all subjects rated themselves as feeling, at a minimum, either moderately pleasant or moderately unpleasant (in the P-mood and U-mood conditions, respectively). If necessary, a brief mood boost (described in the preceding section) was delivered to bring subjects up or down to the requisite level of pleasure or displeasure. The experimenter reminded the subjects that, 2 or 3 days earlier, they had generated both positive and negative autobiographical events in response to common-word probes. She then asked them to spend 5 min trying to recall, in any preferred order, the gist of as many of these events as possible. Subjects were told that they need not give detailed replies; indeed, it was sufficient—even preferable—for them to recall only the probe that was associated with a given event. The experimenter kept a written record of the subjects' spoken responses.

Next, subjects were asked to specify the date (month and year) on which every recalled event had occurred. To make matters easier, the experimenter reminded the subjects of the relevant events by reciting aloud the corresponding probes in the order in which they had been recalled. By comparing these dates with those supplied during the encoding session, we sought to determine whether the subjects were recalling the same events they had generated earlier and whether these events pertained to actual, specific experiences of a personal nature.

After marking yet another mood matrix, subjects were debriefed and then discharged once their level of pleasure was neutral or higher.

Subjects. Sixty-four undergraduates (44 women and 20 men) participated in the study in return for course credit. Criteria for selection included first- or second-year standing, fluency in English, and no prior participation in research involving either autobiographical memory or experimentally induced moods.

Sixteen of the 64 subjects replaced individuals who began but did not complete the experiment either because (a) they were unable to generate specific events in response to at least 75% (12 of 16) of the autobiographical memory probes during the first (encoding) session (1 case); (b) they did not return as scheduled for the second (retrieval) session (4 cases); (c) they became too emotionally distraught during U-mood induction in either session to countenance their continued participation (1 case); or (d) on the basis of periodic ratings of mood taken during the induction of either mood in either session, they appeared to have made little or no progress toward achieving the criterion level of pleasure/displeasure after a considerable amount of time (viz., 20 min or more; 10 cases).

Results and Discussion

Intersession interval. Although most subjects completed the retrieval session 2 days after encoding, some (22 of 64) did so after 3 days in order to accommodate their class schedules. Averaging across all subjects in all encoding/retrieval conditions, which did not differ ($p > .10$), the mean intersession interval was 2.34 days.

Pleasure and arousal ratings. Mean ratings of pleasure recorded on six occasions over the course of the study are shown in Table 1. These ratings correspond to marks made by subjects along the horizontal axis of the mood matrix and range from 4 (extremely pleasant) through 0 (neutral) to −4 (extremely unpleasant).

Several aspects of the data deserve comment. First, as one would expect, baseline (or premood manipulation) ratings of pleasure registered at the start of either the encoding or the retrieval session [encoding session baseline (ESB) and retrieval session baseline (RSB) ratings] were about the same regardless of condition ($ps > .10$).

Second, in keeping with the mood criteria mentioned earlier, subjects rated themselves as feeling, at a minimum, either moderately pleasant (2 or higher) or moderately unpleasant (−2 or lower) just before they began the encoding-session task of autobiographical event generation [before event generation (BEG) rating] and again just before they began the retrieval-session task of autobiographical event recall [before event recall (BER) rating].[2] Although most

[2] Remember that autobiographical event generation and recall were the last tasks performed during the encoding and retrieval sessions, respectively. To advance to the first task in either session

Table 1
Pleasure as a Function of Rating Occasion and Encoding/Retrieval Moods: Experiments 1–3

Encoding/retrieval moods	n	Rating occasion ESB	BEG	AEG	RSB	BER	AER
		Experiment 1					
P/P	16	1.1	2.6	2.0	0.9	2.8	2.3
P/U	16	0.8	2.8	2.3	1.0	−2.6	−2.0
U/P	16	0.7	−2.2	−0.9	1.2	2.8	3.0
U/U	16	0.3	−2.3	−1.3	1.3	−2.2	−1.6
		Experiment 2					
P/P	16	0.6	3.1	2.0	0.8	3.3	2.4
P/U	16	0.9	3.3	2.0	0.7	−3.2	−2.5
U/P	16	0.3	−3.0	−0.8	0.6	3.1	2.3
U/U	16	0.4	−3.3	−0.9	0.9	−3.1	−2.3
		Experiment 3					
P/P	32	1.4	3.2	2.6	1.9	3.3	3.0
P/U	32	1.3	3.2	2.8	1.3	−3.2	−2.6
U/P	32	0.8	−3.1	−0.3	1.3	3.2	2.9
U/U	32	0.8	−3.1	−0.3	1.0	−3.1	−2.3

Note. Ratings reflect a 9-point scale ranging from *extremely pleasant* (4) through *neutral* (0) to *extremely unpleasant* (−4). Data for Experiment 3 represent the average of 2-day and 7-day intersession intervals. P = pleasant mood; U = unpleasant mood; n = number of subjects per mean rating; ESB = encoding session baseline; BEG = before event generation; AEG = after event generation; RSB = retrieval session baseline; BER = before event recall; AER = after event recall.

subjects reached criteria on their own accord, a few required a mood boost to achieve the critical level of pleasure prior to either event generation (14 subjects, 9 U-mood and 5 P-mood, out of a total sample of 64) or event recall (2 subjects, both U-mood, out of the same size sample).

Third, the impact of the mood manipulation declined over time and across tasks. More specifically, the mean difference in pleasure ratings made by encode-P and encode-U subjects (ns = 32) was 4.9 points at the outset of the event generation task (BEG rating: P-mood = 2.7, U-mood = −2.2), $t(62) = 6.39$, $\eta = .31$, $p < .01$, but only 3.2 points at its end [after event generation (AEG) rating: P-mood = 2.1, U-mood = −1.1], $t(62) = 3.20$, $\eta = .22$, $p < .01$. Similarly, the mean difference in pleasure between retrieve-P and retrieve-U subjects (ns = 32) decreased from 5.2 points before the test of event recall (BER rating: P-mood = 2.8, U-mood = −2.4), $t(62) = 6.11$, $\eta = .30$, $p < .01$, to 4.5 points afterward [after event recall (AER) rating: P-mood = 2.7, U-mood = −1.8], $t(62) = 4.55$, $\eta = .26$, $p < .01$. These diminishing differences may reflect a regression to the mean, or they may be due to the tasks themselves. In any case, it should be noted that although the mood manipulation lost some of its effectiveness, it did not lose it all: As indicated previously, mean differences in rated pleasure between P-mood and U-mood subjects were significant at the conclusion of both event generation and event recall (viz., AEG and AER ratings).

Ratings of arousal are summarized in Table 2. These ratings were assigned in accordance with marks made by subjects along the vertical axis of the mood matrix and varied from 4 (extremely high arousal) through 0 (neutral) to −4 (extremely low arousal).

Comparison of Tables 1 and 2 suggests that the CMT had a marked effect not only on the subjects' ratings of pleasure—which is what the technique was specifically intended to do—but on their ratings of arousal as well. More to the point, P-mood subjects rated themselves as being more aroused than did their U-mood counterparts, and this was true for all ratings taken after the CMT had begun (viz., BEG, AEG, BER, and AER ratings in Table 2), ts(62) > 4.97, ηs > .19, ps < .01.

Additional evidence of covariation between ratings of pleasure and arousal was obtained by computing, for every subject, the product–moment correlation between the six pairs of pleasure and arousal ratings represented in Tables 1 and 2 (viz., ratings ESB through AER). The mean of these correlations, .56, differed reliably from zero, $t(63) = 10.79$, $p < .01$.

(viz., word rating at encoding, stem completion at retrieval), subjects in Experiment 1 had to feel at least very pleasant (3 or higher) or very unpleasant (−3 or lower). On average, these subjects required 15–20 min to reach either criterion (range = 5–60 min). The results were highly similar for participants in Experiment 2 and 3, who had to feel at least very pleasant or very unpleasant before they could begin event generation or event recall, both of which were the first tasks administered in their respective sessions.

Table 2
Arousal as a Function of Rating Occasion and Encoding/Retrieval Moods: Experiments 1–3

Encoding/retrieval moods	n	Rating occasion ESB	BEG	AEG	RSB	BER	AER
		Experiment 1					
P/P	16	−0.1	1.0	0.9	0.7	1.4	1.2
P/U	16	0.3	1.4	1.4	0.2	−1.0	−0.4
U/P	16	−0.1	−1.6	−0.2	0.9	1.9	2.6
U/U	16	−0.2	−1.8	−1.5	0.3	−1.6	−1.2
		Experiment 2					
P/P	16	−0.1	1.0	1.3	0.6	1.3	1.3
P/U	16	0.3	0.4	1.1	0.4	−2.3	−2.1
U/P	16	−0.3	−2.3	−0.6	0.4	1.0	0.9
U/U	16	0.3	−1.7	−0.3	1.1	−1.8	−1.6
		Experiment 3					
P/P	32	0.5	0.4	1.3	0.9	1.4	1.7
P/U	32	0.3	0.3	0.8	0.4	−0.9	−0.8
U/P	32	−0.3	−1.9	−0.5	0.5	1.4	1.4
U/U	32	0.3	−1.6	−0.3	0.2	−1.5	−1.1

Note. Ratings reflect a 9-point scale ranging from *extremely high arousal* (4) through *neutral* (0) to *extremely low arousal* (−4). Data for Experiment 3 represent the average of 2-day and 7-day intersession intervals. P = pleasant mood; U = unpleasant mood; n = number of subjects per mean rating; ESB = encoding session baseline; BEG = before event generation; AEG = after event generation; RSB = retrieval session baseline; BER = before event recall; AER = after event recall.

Autobiographical event generation. To assess performance on this task, we derived six measures for every subject and for each type of event (viz., positive and negative). These measures were (a) the number of events generated (maximum of eight per type) within the allotted time (2 min per event); (b) median generation latency (in seconds); (c) median event age (number of months ago that the events took place); (d) mean event intensity (at the time they occurred; range = 1–5); (e) mean event importance (again at time of occurrence; range = 1–5); and (f) mean event vividness (as currently recollected, range = 1–5). The resulting scores were then averaged over subjects to yield the summary data shown in Table 3.

A 2 × 2 (Encoding Mood × Event Type) mixed-design analysis of variance (ANOVA) was applied to each of the six measures. The results of these analyses revealed three significant main effects for event type and one marginal main effect for encoding mood. Regarding the former, negative events were not only older than positive events (*M*s = 35.6 and 28.1 months, respectively), $F(1, 62) = 5.73$, $p < .01$, but also more intense (*M*s = 3.5 and 3.0, respectively), $F(1, 62) = 24.45$, $p < .01$, and more important (*M*s = 3.0 and 2.7, respectively), $F(1, 62) = 7.16$, $p < .01$. Regarding the latter, U-mood subjects took somewhat longer to generate a given event than did P-mood subjects (*M*s = 14.3 and 10.3 s, respectively), $F(1, 62) = 2.93$, $p < .10$, a result consistent with the claim that depressed mood impedes cognitive processing (see Clark, 1983; Ellis & Ashbrook, 1988).

There were no reliable Encoding Mood × Event Type interactions (*p*s > .10), meaning that there was no sign of any sort of mood congruence. Thus, counter to our earlier conjectures, P-mood subjects did not ascribe greater importance to positive than to negative events, nor did U-mood subjects recollect negative events more vividly than positive ones. Moreover, and in contrast to Teasdale and Fogarty's (1979) findings, there was no evidence that P-mood subjects generated positive events more rapidly than negative ones. Although the cause of this inconsistency is unclear, it may be due to different task instructions (the emphasis on generating specific, datable events was probably greater in Experiment 1 than in Teasdale and Fogarty's study) or different task procedures (e.g., Teasdale & Fogarty's subjects were not required to date a given event, nor did they rate its importance or vividness).[3]

Autobiographical event recall. Figure 2 depicts the mean percentages of previously generated positive and negative events that were recalled in each encoding/retrieval condition. Subjects were credited with recalling a given event if they recalled either its exact corresponding probe (e.g., *ship*) or a specific word or phrase that they had mentioned while describing the event to the experimenter (e.g., *ferry boat*). Instances of the latter were uncommon, accounting for only 13 of the 411 total events that were recalled by all subjects in this study. For each of the remaining 398 events (or 97% of the total), subjects succeeded in recalling the precise corresponding probe.

The recall data were analyzed per a 2 × 2 × 2 mixed design (Encoding Mood × Retrieval Mood × Event Type). Results were negative (*p*s > .10) in all respects but two. First, there was a marginally significant main effect of encoding mood, such that the probability of recall was higher for events (of either type) that had been generated under U-mood as opposed to P-mood conditions (*M*s = 46% and 39%, respectively), $F(1, 60) = 3.71$, $p < .10$. Given that generation latency was somewhat longer for encode-U than for encode-P subjects (as noted in the preceding section), this result may merely reflect a difference in time spent thinking about the events.

Second, and more important, there was a significant interaction between encoding and retrieval moods, $F(1, 60) = 6.79$, $p < .01$. As is apparent in Figure 2, a greater percentage of events (again, of either type) was recalled when encoding and retrieval moods matched than when they mismatched (*M*s = 47% and 38%, respectively), $F(1, 60) = 6.79$, $p = .01$. Because this study involved the induction of distinctive states of pleasure or displeasure, rather than a more commonplace neutral mood, any evidence of mood dependence that one may expect to find should fit a symmetric pattern (see Eich, 1989, for a discussion of this issue). This was indeed the case: the advantage in recall of the P/P condition over the P/U condition (*M*s = 44% and 34%, respectively) was comparable to the difference found between the U/U and U/P conditions (*M*s = 51% and 42%, respectively).

Earlier, we remarked that ratings of pleasure were correlated with those of arousal, so that P-mood subjects were more highly aroused than were U-mood subjects. Nevertheless, although all subjects assigned to either mismatched mood condition (viz., P/U or U/P) showed a substantial shift in rated pleasure between the encoding and retrieval sessions (as was mandated by our methods), only some of these subjects also experienced a marked change in arousal. Does this matter in terms of the magnitude of the MDM effect?

To find out, two scores were found for every subject using Eich and Metcalfe's (1989) formulas. One of these

[3] An alternative possibility relates to the fact that in Teasdale and Fogarty's (1979) study, every subject was given just 15 s in which to generate a given event. Even if the subject failed to do so, a score of 15 s was recorded for each "missing" event, and such scores were counted toward the *mean* latency for that person. In the current experiment, however, the maximum generation latency was 2 min. If no event was forthcoming within that time, the event was omitted from further analysis, and it was not entered into computation of the *median* latency for a given individual. Thus, the discrepancy between Teasdale and Fogarty's results and ours may merely reflect different ways of scoring the latency data. To test this possibility, our data were rescored using Teasdale and Fogarty's (1979) methods. Any event that took longer than 15 s to generate, or was never generated in the first place, received a score of 15 s; such scores were included in figuring the mean (rather than the median) latency for each subject. Although the rescored mean latencies were shorter than those shown in Table 3 (no surprise there), the pattern of results remained the same. On average, positive events were generated in 8.6 s by pleasant-mood subjects and in 9.9 s by unpleasant-mood subjects; the corresponding values for negative events were 9.0 and 10.3. Once again, the results were negative vis-à-vis mood congruence (the encoding Mood × Event Type interaction probability level was greater than .10).

Table 3
Number of Events Generated, Event Generation Latency, Event Age, Event Intensity, Event Importance, and Event Vividness as a Function of Encoding Mood and Event Type: Experiments 1–3

Encoding mood	Event type	*n*	Measure: Gen	Lty	Age	Ity	Imp	Vvd
			Experiment 1					
P	Positive	32	7.9	10.3	22.4	3.0	2.6	3.4
P	Negative	32	7.6	10.3	34.5	3.5	2.9	3.4
U	Positive	32	7.6	13.3	33.7	3.0	2.9	3.5
U	Negative	32	7.5	15.2	36.7	3.5	3.1	3.4
			Experiment 2					
P	Positive	22–32	11.1	9.7	33.7	2.6	2.6	3.4
P	Neutral	22–32	1.2	13.3	50.6		1.8	2.7
P	Negative	22–32	3.3	12.6	39.4	2.2	2.7	3.2
U	Positive	23–32	6.7	10.5	52.6	2.2	2.5	3.2
U	Neutral	23–32	2.0	16.3	57.1		1.9	2.8
U	Negative	23–32	6.8	12.0	48.3	2.3	2.9	3.3
			Experiment 3					
P	Positive	55–64	15.8	6.2	45.2	2.5	2.8	3.4
P	Neutral	55–64	2.8	9.4	48.4		1.7	2.8
P	Negative	55–64	5.2	8.2	43.4	2.1	2.8	3.4
U	Positive	59–64	12.7	7.3	36.7	2.3	2.6	3.3
U	Neutral	59–64	3.2	10.5	37.0		1.6	2.7
U	Negative	59–64	7.9	7.9	40.6	2.2	3.0	3.4

Note. Every subject generated a maximum of 16 events in Experiments 1 and 2 and 24 events in Experiment 3. In Experiment 1, subjects were directed to generate positive events in response to half of the autobiographical memory probes and negative events in response to the other half. In Experiments 2 and 3, the type of event generated (either positive, neutral, or negative) in response to a given probe was determined by the subject rather than by the experimenter. Event generation latency is in seconds; event age is in months. In all cases but two, ratings of event intensity, importance, and vividness reflect a 5-point scale ranging from 1 (*neutral, trivial, or vague*) to 5 (*extremely intense, extremely important, or extremely vivid*). The exceptions occur in Experiments 2 and 3, in which the intensity of positive and negative events was assessed on a 4-point scale ranging from 1 (*slightly intense*) to 4 (*extremely intense*); by definition, neutral events in these experiments had an intensity score of zero. Data for Experiment 3 represent the average of 2-day and 7-day intersession intervals. P = pleasant mood; U = unpleasant mood; *n* = number of subjects per mean score. Gen = number of events generated; Lty = event generation latency; Age = event age; Ity = event intensity; Imp = event importance; Vvd = event vividness.

scores—change in pleasure—was defined as the absolute difference between (a) the mean of the pleasure ratings registered before and after the task of event generation (i.e., BEG and AEG ratings in Table 1) and (b) the mean of the pleasure ratings recorded before and after the task of event recall (i.e., BER and AER ratings). The second score—change in arousal—was defined the same way, except that ratings of arousal were substituted for those of pleasure.

Next, the 32 subjects assigned to the mismatched mood conditions (P/U and U/P) were divided into two groups: large change in pleasure–large change in arousal—the 16 subjects (8 per condition) with the largest arousal change scores, and large change in pleasure–small change in arousal—the 16 subjects (again 8 per condition) with the smallest arousal change scores. For purposes of comparison, a third group—small change in pleasure–small change in arousal—was formed, consisting of the 32 subjects assigned to the matched mood conditions (viz., P/P and U/U), most of whom displayed little, if any, disparity between sessions in their ratings of either pleasure or arousal.

Table 4 shows the mean measures of pleasure and arousal change, plus the mean percentages of total events recalled (i.e., positive and negative types combined), for each of the three groups. Regarding event recall, analysis by planned comparisons disclosed a marginal advantage of the Small/Small group over the Large/Small unit (*M*s = 47% and 39%, respectively), $F(1, 61) = 3.21$, $p < .10$, and a more marked advantage of the former subjects over their Large/Large counterparts ($M = 36\%$), $F(1, 61) = 5.68$, $p < .05$. Although the difference between the Large/Small and Large/Large groups was not significant ($p > .10$), the overall pattern is similar to that found by Eich and Metcalfe (1989), thus providing a small measure of added support for the idea that stronger MDM effects are associated with two-dimensional (pleasure plus arousal) than with one-dimensional (pleasure only) shifts in mood state.

Autobiographical event redating. Immediately following free recall, subjects were asked to specify the month and year that each recalled event took place. How well did these dates correspond to those given originally, during the encoding-session task of autobiographical event generation?

To answer this question, the age of every recalled event (expressed in months) was correlated with the corresponding value obtained from the encoding session. Of the 62 correlations that were computed (one for every subject who recalled at least three events), 61 were significant (*p*s < .05); the overall mean correlation was .97. Additionally, of the 407 total events recalled by these subjects, 358 (88%) were calibrated within 1 year of their original date of occurrence. Such striking accuracy could not have been achieved had subjects disregarded either the directive to generate specific, real-life events during the encoding session, or the request to recall these same events during the retrieval session. By inference, then, the subjects followed their task instructions to a tee.

Experiment 2

Rationale

Although the evidence of mood dependent memory collected in Experiment 1 was clear, it cannot be considered conclusive. As noted at the outset, attempts to replicate MDM effects have seldom succeeded, even when undertaken by the same researcher using similar methods. Moreover, the fact that Experiment 1 showed no sign of mood congruent memory—even in the context of a supposedly

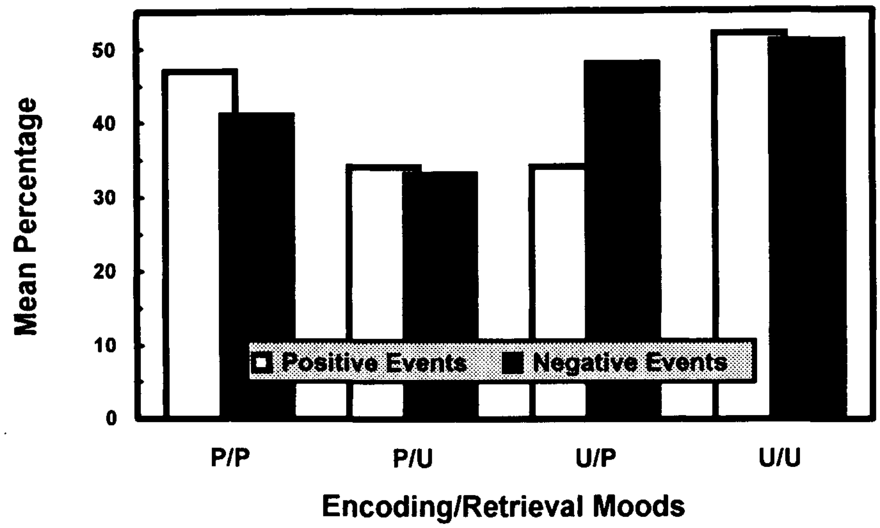

Figure 2. Experiment 1: Recall of positive and negative events as a function of encoding/retrieval moods (P = pleasant; U = unpleasant).

sensitive measure (event generation latency)—is at odds with conventional wisdom, which holds that mood congruence, unlike mood dependence, is a robust and readily reproducible phenomenon. These considerations call for a second study geared toward replicating and extending the results of the first.

Method

Encoding and retrieval sessions. Rather than repeat much of the previous discussion of method, we focus here on the few differences that existed between Experiments 1 and 2. One difference concerned the requirements of the event generation task. Although subjects in the second study, like those in the first, were asked to generate a specific autobiographical incident in response to every probe (16 common, neutral nouns), the type or valence of event generated was determined by the subjects themselves rather than by the experimenter. Thus, after describing and dating a given event, but before appraising its importance and vividness (in the manner mentioned earlier), participants in Experiment 2 rated both the original valence and intensity of the event on a 9-point scale ranging from 4 (extremely positive) through 0 (neutral) to −4 (extremely negative). This measure replaced the 5-point scale used to assess intensity alone in the initial study.

In Experiment 1, autobiographical event generation was the third of three tasks administered during the encoding session, and autobiographical event recall held the same position during the retrieval session. In Experiment 2, however, these tasks were undertaken first in their respective sessions.[4] Accordingly, subjects now had to rate themselves (on the mood matrix) as feeling at least very pleasant or very unpleasant, irrespective of level of arousal, before they could advance to either task, thereby obviating the need for a mood boost.

Table 4
Absolute Difference Between Encoding and Retrieval Ratings of Pleasure and of Arousal, Plus the Percentage of Total Events Recalled as a Function of Pleasure/Arousal Change: Experiments 1–3

Pleasure/arousal change	*n*	|Pls|	|Ars|	EvR (%)
	Experiment 1			
Small/small	32	0.4	0.8	47
Large/small	16	4.6	1.5	39
Large/large	16	4.7	3.9	36
	Experiment 2			
Small/small	32	0.7	0.9	35
Large/small	16	4.8	1.6	26
Large/large	16	5.2	4.1	26
	Experiment 3			
Small/small	32	0.7	0.9	32
Large/small	16	5.1	1.4	27
Large/large	16	5.6	3.8	24

Note. Data for Experiment 3 reflect only the 2-day intersession interval. *n* = Number of subjects per mean score. |Pls| = pleasure; |Ars| = arousal; EvR = events recalled.

[4] Subsequent to autobiographical event generation, subjects in Experiment 2 viewed a series of Rorschach-like inkblots, which they described and rated for emotionality. Two days later, subjects were tested first for autobiographical event recall and then for inkblot recognition memory. The results of the latter task are not discussed here because they relate to the continuing, multistudy project (referenced in the *Retrieval session* section of Experiment 1) concerned with implicit measures of mood dependence. Additional data pertinent to this same project were collected in Experiment 3. During the retrieval session of this study, memory for the autobiographical events was assessed initially via an explicit test of free recall (which provided the data of interest in this article) and subsequently via an implicit test of category production (similar to the task described by Macaulay, Ryan, & Eich, 1993). During the encoding session of Experiment 3, the subjects' sole task was that of autobiographical event generation.

Another change in methodology was that all (rather than most) subjects were tested for free recall 2 days after event generation. Finally, the task of event redating was omitted in Experiment 2. This task had served its purpose in Experiment 1 by showing that subjects complied with the request to generate specific autobiographical events during the encoding session and to recall these same events during the retrieval session. Because subjects in the second study received these same instructions, we saw no point in saddling them with the redating task.

Subjects. Sixty-four undergraduates (39 women and 25 men) earned course credit for their participation. Criteria for subject selection were the same as those for Experiment 1. Also like the initial study, 16 students were randomly assigned to each of the four combinations of encoding and retrieval mood, and all testing was conducted individually.

Two students were no-shows for the retrieval session, and 9 others were unable (or unwilling) to attain the requisite level of pleasure or displeasure even after prolonged exposure to the CMT. Substitutes for these 11 individuals were recruited, resulting in a final sample of 64.

Results and Discussion

Pleasure and arousal ratings. Mean ratings of pleasure appear in Table 1. As anticipated, there were no discernible differences among encoding/retrieval conditions in ratings registered at the start of either session ($ps > .10$). Also, reflecting the stricter mood standards set for this study, P-mood subjects rated themselves as feeling very pleasant or better prior to both event generation and event recall, whereas U-mood subjects felt very unpleasant or worse (see footnote 2). Differences between P-mood and U-mood subjects in posttask ratings, although smaller than those seen pretask, were significant, $ts(62) > 8.45$, $\eta s > .35$, $ps < .01$, a finding that further attests to the durability of the moods induced by the CMT.

Mean ratings of arousal are presented in Table 2. Comparison of these data with the pleasure ratings summarized in the preceding table suggests that, as was the case in Experiment 1, ratings of arousal tended to covary with those of pleasure. The mean correlation between the six principal pairs of ratings made by subjects in Experiment 2 was .53, $t(63) = 7.76$, $p < .01$.

Autobiographical event generation. Table 3 shows a summary of several measures of event generation performance, including the mean number of events given a positive (1 to 4), neutral (0), or negative (−1 to −4) rating of original valence and intensity.

A 2 × 3 mixed-design analysis of these scores revealed a reliable effect of event type, $F(2, 124) = 108.08$, $p < .01$. Averaging across encoding moods, subjects rated 8.9 of their events as positive, 5.0 as negative, and 1.6 as neutral; each of these means differed from the other two, $Fs(1, 62) > 36.37$, $ps < .01$. More important, there was also a reliable Event Type × Encoding Mood interaction, $F(2, 124) = 33.06$, $p < .01$. Relative to their U-mood counterparts, P-mood subjects rated more of their events as positive (Ms = 11.1 and 6.7, respectively), $F(1, 62) = 40.86$, $p < .01$, and fewer as negative (Ms = 3.3 and 6.8, respectively), $F(1, 62) = 29.78$, $p < .01$.

Although this pattern signifies the presence of a mood congruent effect, the source of this effect is uncertain. One possibility is that P-mood and U-mood subjects differ in the actual contents of their autobiographical memories, in which case the results could be construed as evidence of mood congruence in event generation. Alternatively, the results may imply mood congruence in event evaluation, such that P-mood subjects interpret personal recollections in a particularly rosy light, even though their recollections are substantively similar or even identical to those generated by the U-mood subjects. Because participants in this study both retrieved and rated autobiographical events in one and the same mood (either P or U), we could not separate mood congruent generation from mood congruent evaluation solely on the basis of the data at hand.

Such separation, however, can be achieved via a different kind of data. We recruited two colleagues to serve as impartial reviewers and asked them to independently read and evaluate all 979 event descriptions that had been supplied by 63 of the 64 subjects in this study. (The experimenter-prepared transcript containing the autobiographical accounts of 1 U-mood subject was lost.) Neither colleague was cognizant of the study's aims or methods, and both were unaware of the subjects' mood at event generation.

Equipped with their own set of transcripts, the reviewers read every event description and coded it as either positive or negative depending on its perceived emotional tone. Given the rarity with which either P-mood or U-mood subjects rated events as neutral, we limited our colleagues' choices to just two, either positive or negative. Doing so also made their task more manageable and less laborious.

Asked to assess the valence of a given event, the two reviewers concurred on 91% of their choices (890 of 979 events), indicating an impressive degree of consensus. On average, they gave a positive evaluation to 72% (356 of 497) of the events that had been generated by P-mood subjects, but they accorded the same judgment to only 52% (253 of 482) of the events that had been generated by U-mood subjects. By way of comparison, the subjects themselves rated 71% and 42% of their events as positive, depending on whether they were in a P or U mood, respectively. The close correspondence between these two sets of numbers implies that moods mainly affect *what* one remembers about his or her past rather than *how* one rates or evaluates it.

The next measure to consider is event intensity. For every subject who had generated at least one positive and one negative event, the mean of all positive (1–4) intensity ratings was found, as was the mean of all negative (−1 to −4) intensity ratings. The latter average was then converted into an absolute value to make it more comparable with the former. No corresponding mean was computed for neutral events because they, by definition, occupied the 0 point on the valence–intensity scale.

Both the main effect of event type and the Event Type × Encoding Mood interaction were reliable, $Fs(1, 61) > 6.46$, $ps < .01$. Whereas P-mood subjects perceived positive events to be more intense than negative events (Ms = 2.6 and 2.2, respectively), $F(1, 61) = 13.79$, $p < .01$, U-mood subjects saw them the same (Ms = 2.2 and 2.3, respectively), $F(1, 61) = p > .10$.

Does this interaction indicate a difference between P-mood and U-mood subjects in the contents of their autobiographical recollections or in the criteria they adopt in assessing the intensity of positive and negative events? This, of course, is the same question that was raised with respect to the probability of generating positive and negative events, and it arises for the same reason: that whichever mood attended the generation of a given event also accompanied its evaluation. Unfortunately, the question does not lend itself to the same solution because, although the event descriptions solicited from subjects in this study were graphic enough to convey a clear sense of emotional valence (either positive or negative) to our nonpartisan reviewers, they were too sparse to allow an accurate, impartial assessment of emotional intensity. Consequently, it is unclear whether the intensity data denote mood congruence operating at the level of event retrieval or at the level of event appraisal. Equally unsettled is why subjects in Experiment 1 (irrespective of their mood state) rated negative events as more intense than positive events, whereas those in Experiment 2 did the opposite (provided they were in a pleasant mood). Although this may have to do with the use of a "constrained" test of event generation in the first study versus an "unconstrained" test in the second, the issue is open.

None of the four remaining measures of event generation performance—latency, importance, vividness, and age—provided evidence of either mood congruence (Encoding Mood × Event Type interaction probability levels greater than .10) or a main effect of encoding mood ($ps > .10$). (To contribute to any of these indexes, a subject had to have generated at least one positive, one negative, and one neutral event.) However, every measure except age did vary reliably as a function of event type, $Fs(2, 98) > 6.17$, $ps < .01$. Concerning latency, negative events were centered between positive events at the short end of the time-to-generate distribution and neutral events at the long end (Ms = 10.1 s, 12.2 s, and 14.9 s, respectively), pairwise comparison $Fs(1, 49) > 2.83$, $ps < .10$. Regarding importance, negative events were rated highest and neutral events lowest (Ms = 2.8 and 1.9, respectively), $F(1, 49) = 28.37$, $p < .01$. The marginal advantage in importance of negative over positive events (M = 2.6), $F(1, 49) = 3.40$, $p < .10$, resembles the pattern found in Experiment 1. Finally, with respect to vividness, neutral events were rated lower than were either positive or negative events (Ms = 2.8, 3.3, and 3.3, respectively), $Fs(1, 49) > 9.89$, $ps < .01$. The lack of a difference in recollective vividness between positive and negative events was not surprising given the analogous null result in Experiment 1.

Autobiographical event recall. To gain a general impression of performance on this task, we determined the percentage of total events recalled (i.e., positive, negative, and neutral events combined) by a given subject and submitted these scores to a 2 × 2 (Encoding Mood × Retrieval Mood) ANOVA.

Although neither main effect was reliable ($ps > .10$), the interaction was robust, $F(1, 60) = 7.87$, $p < .01$. Replicating the mood dependent effect found in the first experiment, proportionately more events were recalled when encoding and retrieval moods matched than when they mismatched (Ms = 35% and 26%, respectively).

A more detailed picture of task performance appears in Figure 3, which depicts the mean percentages of positive, negative, and neutral events recalled by subjects who had generated at least one event of each type (51 of the 64 participants in Experiment 2 met this standard). Despite the reduced sample size, a 2 × 2 × 3 analysis of these percentages revealed a reliable interaction between encoding and retrieval moods, $F(1, 47) = 6.59$, $p = .01$, indicating

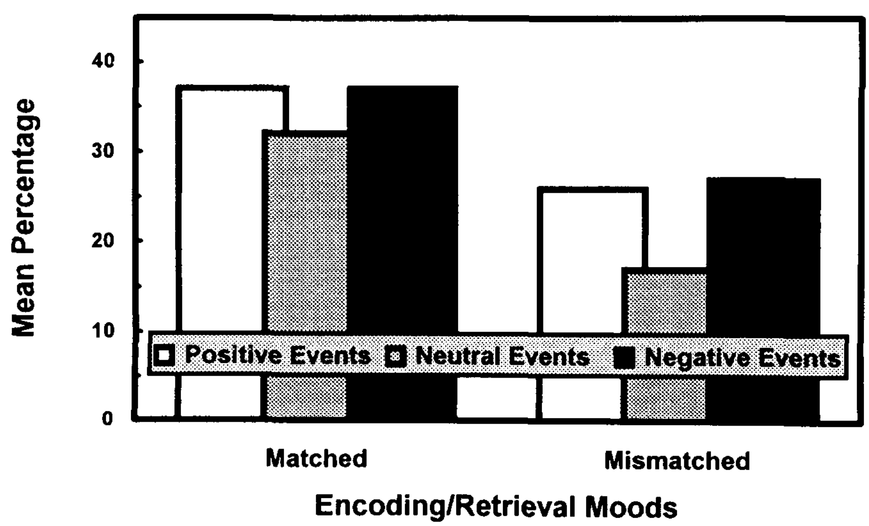

Figure 3. Experiment 2: Recall of positive, neutral, and negative events as a function of encoding/retrieval moods (matched vs. mismatched).

an advantage in recall of matched over mismatched mood conditions (illustrated in Figure 3). Neither the main effect of event type nor any of its affiliated interactions were significant ($ps > .10$), suggesting that positive, negative, and neutral events were equally susceptible to mood dependence.

In Table 4, data on total event recall were reformatted as a function of differing degrees of pleasure/arousal change (defined the same way as in Experiment 1). In comparison with subjects in the Small/Small group, those in the Large/Small and Large/Large groups performed equally poorly (Ms = 35% and 26%, respectively), $Fs(1, 61) > 5.04$, $ps < .05$. Thus, in contrast to the trend observed in earlier experiments (viz., the first study reported here, plus the four described by Eich & Metcalfe, 1989), the results of Experiment 2 did not provide even slim support for the idea that a two-dimensional change in mood is more conducive to demonstrating MDM than is a one-dimensional change. The cause of this apparent conflict is unknown.

Experiment 3

Rationale

Two results of Experiment 2 rise above the rest in importance. First, when asked to generate specific events from their personal past in response to affectively neutral probes, participants in Experiment 2 recollected more positive episodes, and fewer negative ones, if they were in a pleasant as opposed to an unpleasant mood. Second, when asked to freely recall the gist of these events 2 days later, subjects remembered more if they were tested in the same mood they had experienced during event generation than in the contrasting emotional context. Together, these results mark Experiment 2 as the only single study we know of that has succeeded in demonstrating both mood congruent and mood dependent memory. This is reason enough to wonder whether the results are replicable—hence Experiment 3.

As it happens, there was another reason for running a third study, one that is traceable to the results of the first. Recall that in Experiment 1, the interval separating the encoding and retrieval sessions was set at 2 days for some subjects and at 3 days for others. We reanalyzed total event recall (i.e., positive and negative events combined) with a view to this difference, and an intriguing pattern emerged. In particular, whereas the relative advantage in recall of matched over mismatched mood conditions was about 19% at the 2-day interval (Ms = 49% and 41%, respectively; ns = 23 and 19, respectively), the corresponding advantage at the 3-day interval was roughly 38% (Ms = 44% and 32%, respectively; ns = 9 and 13, respectively). These results thus constitute correlational evidence that mood dependent effects in memory become stronger as the gap between event generation and event recall grows longer. By systematically varying the duration of the retention interval in Experiment 3, we sought to secure empirical evidence of this same relation.

Method

Subjects. Selected according to the criteria applied in Experiments 1 and 2, 128 undergraduates (87 women and 41 men) served as subjects in return for course credit. The sample contained substitutes for students who either (a) generated specific events in response to less than 75% (i.e., 18 of 24) of the autobiographical memory probes (2 cases); (b) failed to attend the retrieval session (5 cases); (b) became too upset during U-mood induction to proceed further (2 cases); or (c) seemed unable (or unwilling) to attain the requisite pretask level of pleasure or displeasure (24 cases).

Encoding and retrieval sessions. For a random half of the subjects in this study, the retrieval session met 2 days after the encoding session; for the remaining subjects, the intersession interval was set at 7 days. The factorial combination of these two intervals with two encoding moods (P vs. U) and two retrieval moods (again, P vs. U) defined eight conditions, each of which was represented by 16 individually tested subjects.

In an effort to enhance the sensitivity of memory measurement, participants in Experiment 3 were asked to generate a specific autobiographical event in response to each of 24 emotionally untinged probes. Otherwise, their task was identical to the one performed by subjects in Experiment 2 (who had generated a maximum of 16 events, divided as they saw fit among positive, negative, and neutral types).

As in Experiment 2, autobiographical event generation was the first task undertaken during the encoding session of Experiment 3, and autobiographical event recall was the initial task performed during the retrieval session (see footnote 4). As was also the case in Experiment 2, participants in Experiment 3 were required to reach a critical level of mood (viz., very pleasant or very unpleasant, irrespective of degree of arousal) before they could advance to either task. The methods by which moods were modified and measured paralleled those used in the two preceding studies.

Results and Discussion

Pleasure and arousal ratings. Ratings of pleasure and arousal, averaged across intersession intervals, which did not differ ($ps > .10$), are displayed in Tables 1 and 2, respectively. These averages accord well with those found in the first two studies, and, so rather than discuss the new data in detail, we refer the reader to the earlier relevant accounts (see footnote 2).

Autobiographical event generation. Apart from the fact that subjects in Experiment 3 generated a maximum of 24 rather than 16 events, their task mirrored the one performed by participants in Experiment 2. In principle, then, both studies should show similar patterns of results. As the data appearing in Table 3 make plain, the patterns proved to be virtually identical in terms of event generation latency, event age, and ratings of event intensity, importance, and vividness. In terms of the number of positive, neutral, and negative events generated, Experiment 3 showed a smaller mood congruent effect than did Experiment 2. Nonetheless, the Encoding Mood × Event Type interaction was reliable, $F(2, 252) = 20.77$, $p < .01$, with P-mood subjects generating more positive and fewer negative events than their U-mood peers, $Fs(1, 126) > 21.21$, $ps < .01$.

Autobiographical event recall. A 2 × 2 × 2 analysis of the percentages of total events recalled (i.e., positive, neutral, and negative types combined) yielded two notable results. First, the overall level of recall was higher at the 2-day than at the 7-day intersession interval (*M*s = 29% and 19%, respectively), $F(1, 120) = 30.68$, $p < .01$.

Second, although the triple interaction among encoding mood, retrieval mood, and intersession interval was significant, $F(1, 120) = 6.63$, $p = .01$, its configuration was completely counter to what had been predicted, as is evident in Figure 4. Replicating the results of Experiment 2 (and, for the most part, those of Experiment 1 as well), there was a reliable advantage of matched over mismatched mood conditions at the 2-day intersession interval (*M*s = 32% and 25%, respectively), $F(1, 120) = 7.14$, $p < .01$. As was also the case in the earlier studies, the size of this advantage was invariant with respect to event type. A separate analysis of the recall data supplied by 2-day subjects who had generated at least one positive, one negative, and one neutral event ($n = 58$) showed a significant interaction between encoding and retrieval moods, $F(1, 54) = 4.29$, $p < .05$, but no appreciable effect (whether simple or interactive in nature) of event type ($ps > .10$).

However, instead of seeing a stronger MDM effect at the 7-day interval, we found no effect at all (*M*s = 18% and 20% for matched and mismatched moods, respectively, $ps > .10$). Possible reasons for this negative result are considered momentarily.

Table 4 recasts the recall data for the 2-day interval in relation to pleasure/arousal change. Analysis by planned comparisons showed that subjects who had experienced a large as opposed to a small change in both pleasure and arousal recalled fewer total events (*M*s = 24% and 32%, respectively), $F(1, 61) = 4.12$, $p < .05$. Neither of the other two pairwise contrasts were significant ($ps > .10$). Thus, here, as in Experiment 1, a slightly clearer picture of MDM emerged in conjunction with a two-dimensional than a one-dimensional alteration in affective state.

General Discussion

The research reported here was undertaken in the pursuit of a persistent problem: how to reveal robust and reliable evidence of mood dependent memory. Our approach to this problem was founded on a do-it-yourself principle that maintains that the more one must rely on internal resources, rather than on external aids, to generate both the target themselves and the cues required for their retrieval, the more likely is one's memory for these events to be mood dependent.

To cast this principle, and our approach, in more concrete terms, consider the following scenario. Two individuals—one happy, the other sad—are shown, say, a *rose* and are asked to identify and describe what they see. Both individuals are apt to say much the same thing and to encode the *rose* event in much the same manner. After all, and with all due respect to Gertrude Stein, a rose is a rose is a rose, regardless of whether it is seen through a happy or sad eye. The implication, then, is that the manner in which the perceivers encode the *rose* event will be largely, if not entirely, unrelated to their mood. If true, then when retrieval of the event is later assessed via nominally noncued or "spontaneous" recall, it should make little if any difference whether or not the subjects are in the same mood they had experienced earlier. In short, memory for the *rose* event will probably not appear to be mood dependent under these circumstances.

Now imagine a different situation. Instead of identifying and describing the rose, the subjects are asked to recall a specific episode, from any time in their personal past, that the object calls to mind. Rather than involving the relatively automatic or data-driven perception of an external stimulus, the task now requires the subjects to engage in internal mental processes such as reasoning, reflection, and cotemporal thought, "the sort of elaborative and associative pro-

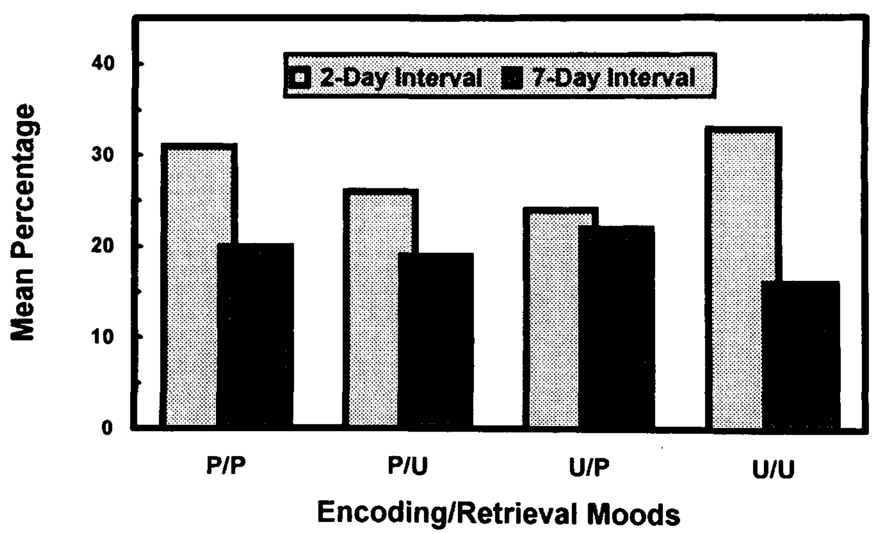

Figure 4. Experiment 3: Total event recall (positive, neutral, and negative types combined) as a function of intersession interval (2 days vs. 7 days) and encoding/retrieval moods (P = pleasant; U = unpleasant).

cesses that augment, bridge, or embellish ongoing perceptual experience but that are not necessarily part of the veridical representation of perceptual experience" (Johnson & Raye, 1981, p. 70). Furthermore, even though the stimulus object is itself affectively neutral, the autobiographical memories it triggers are apt to be strongly influenced by the subjects' mood. Thus, for example, whereas the happy person may recollect receiving a dozen roses from a secret admirer, the sad subject may remember the flowers that adorned his father's coffin. In effect, then, the *rose* event becomes closely associated with or deeply colored by the subject's mood, thereby making mood a potentially potent cue for retrieving the event. Thus, when later asked to spontaneously recall the gist of the episode they had recounted earlier, the subjects should be more likely to remember having related a vignette involving roses if they are in the same mood they had experienced earlier than in a different affective state. In this situation, then, memory for the *rose* event should appear to be mood dependent.

These admittedly sketchy and speculative ideas provided the impetus for our project. The fact that all three studies showed MDM in the free recall of autobiographical events that had been generated 2 or 3 days before suggests that the ideas have some merit and that the paradigm introduced here is a potentially useful one for investigating mood dependence.

To realize this potential, however, many issues will need to be resolved. For instance, if both of the scenarios envisioned earlier were incorporated into an actual experiment, would the results pan out as predicted? That is, would memory for one and the same stimulus object manifest mood dependence if it had been encoded internally (i.e., as a probe for autobiographical event generation), but not if it had been encoded externally (i.e., as a target for perceptual identification and description)? Because the main aim of this project was to find MDM effects—not make them go away—we did not include an "external" encoding task here. It now makes sense to do so, given the positive results found with the "internal" task of autobiographical event generation.

An issue of related interest is whether more potent MDM effects accrue to certain types of internal encoding tasks than to others. Although some internal processing is surely required to generate, say, *rose* in response to the semantic cue *types of flowers: daisy—r* (the internal task used by Eich & Metcalfe, 1989), more such processing would seem to be needed to recollect a specific *rose*-related event from autobiographical memory (the internal task used here). Furthermore, the autobiographically generated *rose* is apt to be more deeply colored by or closely connected to one's mood than is the semantically generated *rose* (see Bower, 1992). For these reasons, it is possible that the former event would show stronger signs of MDM than would the latter, notwithstanding the fact that both refer to the same nominal stimulus and even though both are self-generated or internal in origin. Moreover, it is conceivable that an even more robust MDM effect would emerge if the target event was not merely colored by one's mood (as in the case of the autobiographically generated *rose*) but was construed as actually having caused that mood (see Bower, 1992, for a lucid analysis of the role of "causal belongingness" in mood dependent memory). Just how real or remote these possibilities are remains to be seen.

In addition to the encoding-related issues just raised, a number of other questions deserve deliberation. For instance, are stronger MDM effects associated with two-dimensional (pleasure plus arousal) than with one-dimensional (pleasure only) shifts in mood state? Although most of the data collected to date suggest an affirmative answer (viz., Experiments 1 and 3 in this article and all four of the studies reported by Eich & Metcalfe, 1989), the data are purely correlational in nature. To establish a causal connection, new mood modification techniques need to be developed that permit pleasure to be manipulated independent of arousal.

Turning to a different issue, do MDM effects become stronger as the gap separating event encoding from event retrieval grows longer? Given the results of Experiment 3, which showed evidence of mood dependence at the 2-day but not at the 7-day intersession interval, the obvious answer is no. It should be noted, however, that the low overall level of recall performance at the 7-day interval (M = 19%) left little room for MDM to materialize. (In retrospect, it would have been wiser to have subjects generate no more than 12 or 16 autobiographical events rather than 24.) Also, it is possible that even if MDM was no stronger at 7 days than a 2 (assuming no formidable floor effects), a more robust mood dependent effect may emerge at 2 days than at, say, 2 hr. In view of these considerations, the relation between magnitude of MDM and the duration of the retention interval warrants a second look.

We draw this discussion to a close by calling attention to one other matter. Intuitively, it seems plausible to predict that a shift from a pleasant to unpleasant mood would disproportionately reduce the recall of positive as opposed to negative events, whereas a shift in the opposite direction would lead to a larger loss of memory for negative than for positive events. Indeed, Bower (1992, p. 27) described a study in which subjects were given either success or (false) failure feedback contingent on their performance on each of a series of intellectual tasks. Subsequently, subjects were tested for free recall of the tasks while experiencing either a positive or a negative mood. Whereas positive-mood subjects recalled more of their successful than of their failed tasks, negative-mood subjects recalled more failures than successes. This being so, why is it that in every study reported here, the recall of positive and negative autobiographical events was impaired to the same degree by a shift in mood state, regardless of the direction of the shift (i.e., from P to U or vice versa)? Although unanswerable at present, this question, like those posed previously, may hold the key to a clearer and more complete understanding of mood dependent memory for events of the personal past.

References

Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin, 99,* 229–246.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148.

Bower, G. H. (1987). Commentary on mood and memory. *Behavior Research and Therapy, 25,* 443–455.

Bower, G. H. (1992). How might emotions affect learning? In S.-A. Christianson (Ed.), *Handbook of emotion and memory* (pp. 3–31). Hillsdale, NJ: Erlbaum.

Bower, G. H., & Mayer, J. D. (1985). Failure to replicate mood-dependent retrieval. *Bulletin of the Psychonomic Society, 23,* 39–42.

Bower, G. H., & Mayer, J. D. (1989). In search of mood-dependent retrieval. *Journal of Social Behavior and Personality, 4,* 121–156.

Brown, W. P., & Ure, D. M. J. (1969). Five rated characteristics of 650 word association stimuli. *British Journal of Psychology, 60,* 233–249.

Campbell, J. D. (1990). Self-esteem and clarity of the self-concept. *Journal of Personality and Social Psychology, 59,* 538–549.

Clark, D. M. (1983). On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures. *Advances in Behaviour Research and Therapy, 5,* 27–49.

Eich, E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory & Cognition, 8,* 157–173.

Eich, E. (1989). Theoretical issues in state dependent memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 331–354). Hillsdale, NJ: Erlbaum.

Eich, E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 443–455.

Ellis, H. C., & Ashbrook, P. W. (1988). Resource allocation model of the effects of depressed mood states on memory. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition and social behavior* (pp. 25–43). Göttingen, Germany: Hogrefe.

Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 179–230). Hillsdale, NJ: Erlbaum.

Johnson, M. K., & Multhaup, K. S. (1992). Emotion and MEM. In S.-A. Christianson (Ed.), *Handbook of emotion and memory* (pp. 33–66). Hillsdale, NJ: Erlbaum.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.

Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 482–526). New York: Guilford Press.

Kihlstrom, J. F. (1989). On what does mood-dependent memory depend? *Journal of Social Behavior and Personality, 4,* 23–32.

Macaulay, D., Ryan, L., & Eich, E. (1993). Mood dependence in implicit and explicit memory. In P. Graf & M. E. J. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology* (pp. 75–94). Hillsdale, NJ: Erlbaum.

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

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57,* 493–502.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604.

Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology, 88,* 248–257.

Ucros, C. G. (1989). Mood state-dependent memory: A meta-analysis. *Cognition and Emotion, 3,* 139–167.

Velten, E. (1968). A laboratory task for induction of mood states. *Behavior Research and Therapy, 6,* 473–482.

Received September 17, 1993
Revision received December 20, 1993
Accepted January 6, 1994 ■