This study examined the stability of autobiographical memory dysfunction (i.e., difficulties in retrieving specific memories) during the course of major depressive disorder, its relation to early adverse experiences, and its influence on the course of depressive disorder. Using the Autobiographical Memory Test (AMT), specificity of autobiographical memory was assessed in 25 subjects with a current depressive disorder at baseline, and at 3 and 7 months follow-up. Also, information about self-reported childhood traumatization, and demographic and clinical variables was obtained. Autobiographical memory performance was relatively stable over time despite clinical improvement in the sample. It was not related to depression severity at baseline, while higher levels of childhood traumatization were correlated with more specific memory performance to negative cue words at baseline, but not during follow-up. Specific autobiographical responses to negative cue words predicted a better prognosis, whereas specific responses to positive cue words were not related to prognosis. Autobiographical memory dysfunction in depression appears to be stable over time, is related to short-term prognosis in depression, and may act as a vulnerability factor that influences the long-term course of depressive disorders.

Copyright 2002, Elsevier Science (USA). All rights reserved.

MAJOR DEPRESSION is characterized by a variety of changes in memory performance. In particular, depressed patients' relative inability to retrieve specific autobiographical memories, even when they are explicitly asked to do so, has been the focus of much research in the past decade. For example, depressed patients tend to respond to both negative and positive cue words (e.g., “happy”) with nonspecific, overgeneral memories (e.g., “when I am listening to music”) instead of specific recollections located in time and place (e.g., “the day my child was born”). This phenomenon of overgeneral recall does not pertain to the content of autobiographical memory, but to its lack of spatiotemporal specificity. Overgeneral recall is associated with a diagnosis of major depression and trauma-related disorders such as acute stress disorder and post-traumatic stress disorder. Overgeneral recall is not found in patients with a variety of other psychiatric diagnoses and normal individuals who score high on neuroticism and/or trait anxiety.

METHOD

Subjects

Twenty-five consecutive outpatients (15 women) from the mood-disorders unit of the Community Mental Health Center (CMHC) in Maastricht participated in this study, which was part of a larger research project. Subjects were interviewed using the Structured Clinical Interview for DSM-IV. All subjects met DSM-IV criteria for major depressive disorder (single episode or recurrent) as their primary Axis I diagnosis. Exclusion criteria were a diagnosis of bipolar disorder, organic brain disease, substance abuse, and/or treatment with electroconvulsive therapy within the previous 6 months. Mean age of the subjects was 41.5 years (range, 27 to 58); 44% had an educational level below medium. (Medium education level reflects an intermediate professional qualification; high educational level refers to a college or university degree. Educational level was scored on a 11-point scale, ranging from 1 = no education to 11 = university degree). The mean duration of the current depressive disorder was 14 months (SD = 17; range, 2 to 84). Fifteen (60%) subjects suffered from a first episode. During the study, a majority (72%) of subjects was using antidepressant medication; yet, previous research showed that specificity of autobiographical memory is not influenced by the use of antidepressants. The first measurements were obtained just before or just after the start of treatment, which consisted of a combination of an antidepressant and supportive psychotherapy in the majority
of cases. Psychopharmaceutical treatment in the CMHC was conducted following state-of-the-art guidelines (e.g., daily doses were increased in case of nonresponse after 4 to 6 weeks, and class of antidepressant was changed in case of persisting nonresponse).

**Measures**

*Montgomery-Asberg Depression Rating Scale (MADRS).* The MADRS is an instrument to be used by trained interviewers to assess depression severity. It consists of 10 items, are rated on a scale ranging from 0 to 6. Rating scores are summed to yield a total score, with higher scores indicating higher levels of depression.

*Self-Rating Depression Scale (SDS).* The SDS is a 20-item self-report instrument of depressive symptoms. Items are scored on a 4-point scale; total scores are obtained by summing across all items (range, 20 to 80). Higher scores indicate higher levels of depressive symptoms.

*Autobiographical Memory Test (AMT).* The Dutch translation of the original AMT was used. Five negative and five positive cue words were printed on separate sheets in a booklet. During each test session, a different version (i.e., different cue words) of the AMT was used to exclude practice effects. The numbers of specific autobiographical responses were counted for negative and positive cue words separately. AMT data obtained during the first session were evaluated by two raters. Their specificity ratings reached sufficient to good agreement, with kappa’s ranging between .62 and .86 for different cue words. Follow-up measurements were scored by one of these raters who was blind as to subjects’ actual clinical condition in terms of depression severity.

*Childhood Trauma Questionnaire (CTQ).* The CTQ is a retrospective measure of various aspects of abuse and neglect during childhood. Items are scored on 5-point scales (1 = never true to 5 = very often true). In the current study, the short version of the CTQ was used. This version contains five subscales: emotional abuse (EA, 12 items; range, 12 to 60), physical abuse (PA, seven items; range, 7 to 35), sexual abuse (SA, seven items; range, 7 to 35), emotional neglect (EN, 16 items; range, 16 to 80), and physical neglect (PN, eight items; range, 8 to 40). A weighted total score (range, 1 to 25) was used in the analyses.

**Procedure**

Subjects were tested individually by a research assistant. After they had given written informed consent, subjects completed the AMT. They were instructed to write down a specific autobiographical memory for each cue word in the booklet. It was explained to them that a specific memory refers to a personally experienced event that happened at a particular time (within 1 day) and place. The first two items in each booklet were neutral practice items (“car” and “tree”). If necessary, more practice items were given. It was not until the research assistant was confident that the patient understood the instructions and had provided at least one specific memory in response to the practice words that the 10 cue words of the AMT were presented. Subjects wrote under each cue word an autobiographical memory that the word reminded them of and then dated that memory. The test was largely self-paced. In case the subjects did not turn the page of the booklet after 2 minutes, they were instructed to do so even when they had not written down a memory. Following this, the subjects completed the CTQ and SDS. The MADRS was completed by their therapists. Subjects returned for follow-up measurements at 3 and 7 months. During these test occasions, the same procedure was carried out for the AMT and SDS. The MADRS was now completed by a research assistant, while the CTQ was not administered. All raters who completed the MADRS were trained by one of the authors (FP), using video-taped instruction interviews.

**Statistical Analyses**

Paired-sample t tests were used to compare normally distributed variables. For non-normally distributed variables, Friedman’s test was used to compare means. Correlations were computed using Pearson (normal distribution) and Spearman (non-normal distribution) coefficients. Prediction of depression severity was evaluated by linear regression analyses. Significance levels were set at \( \alpha = .05 \). As the sample size was rather small, separate regression analyses for the two follow-ups may produce type I errors. Therefore, to increase the reliability of regression analyses, the average of depression severity scores at 3 and 7 months follow-up was used as the dependent variable. This procedure is legitimate when the same model can be found in the separate regression analyses of the data collected at the two points in time. This can be controlled by computing regression coefficients for independent variables while using the change of the depression scores between 3 and 7 months as dependent variable. If the independent variables do not predict this change, it can be safely assumed that the model for the separate analyses is the same. Linear regression analysis showed that the independent variables could not predict change in depression scores (MADRS, \( F(4,17) = .50, P = .73; \) SDS, \( F(4,18) = .49, P = .74 \)) between 3 and 7 months.

**RESULTS**

**General Statistics and Correlations Between Depression and Memory Specificity**

Overall, the 25 subjects were moderately to severely depressed according to their mean scores on the MADRS and SDS (Table 1). The MADRS baseline score of one patient was missing. At 3 and 7 months follow-up, complete data were available for 24 and 23 subjects, respectively. One patient could not be contacted for both visits, while the other patient refused to participate in the 7-month follow-up visit. Repeated-measures analyses of variance (ANOVAs) were used to test for changes between baseline MADRS and SDS scores and their follow-up scores. The MADRS scores \( F(2,42) = 11.9, P < .001 \) and SDS scores \( F(2,44) = 16.7, P < .001 \) declined significantly over time. This effect comes close to what others studies have reported on the course of symptoms in treated patients. The mean numbers of specific...
memories to the cue words at the three time points are also shown in Table 1.

Next, correlation coefficients were computed between baseline MADRS, SDS, CTQ, and AMT performance (i.e., number of specific responses) on the one hand, and depression severity at follow-up on the other hand. The results are shown in Table 2.

Correlations between AMT performance at baseline and depression severity remained nonsignificant at all points in time, although reaching borderline significance at 3 months ($P = .06$ for MADRS and $P = .07$ for SDS). Baseline depression severity indices were significantly correlated with severity at 3 months follow-up (except for baseline MADRS with SDS), but not with depression severity at 7 months follow-up. Baseline CTQ scores were significantly correlated with depression severity at 7 months.

**Autobiographical Memory Performance and Its Stability Over Time**

At baseline, subjects recalled on average 2.32 (SD = 1.4) specific memories to positive cue words and 2.24 (SD = 1.3) specific memories to negative cue words; this difference was not significant (Wilcoxon-test, $z = -1.15$, $P = .25$). The two-tailed Spearman correlation coefficient between positive and negative AMT scores at baseline was $r(25) = .43$, $P = .03$. To test whether AMT scores changed significantly over time, multiple Friedman two-way ANOVAs were performed on the scores obtained at the different points in time. The only significant difference that emerged was that between AMT scores for negative cue words at baseline and those at 3 months ($\chi^2 = 7.04$, $P = .008$). Other differences were not significant (highest $\chi^2 = 3.52$, $P = .06$). Stability of AMT performance over time was also examined by calculating Spearman’s correlations between all scores. All correlations were significant (range of $r$ between .43 and .79) except for the correlation between baseline AMT performance to negative cue-words and both AMT scores after 3 months $(rs < .30)$. Thus, both analyses are indicative of more overgeneral recall at 3 months.

**Relation Between Autobiographical Memory Function and Self-Reported Childhood Traumatization**

To address the relation between overgeneral recall at baseline and childhood traumatization, two linear regression analyses were performed. A higher level of childhood traumatization was found to predict significantly more specific recall in response to negative cue words ($F(1,23) = 8.63$, $P = .007$, $B = .30$, 95% confidence interval [CI] of $B = .09$ to .52). Childhood traumatization had no effect on specific recall in response to positive cue-words ($F(1,23) = 1.71$, $P = .20$). However,

### Table 1. Mean (SD) Scores of Depressed Subjects on the MADRS and SDS, and Mean Number (SD) of Specific Memories at Baseline and Follow-up (3 and 7 months)

<table>
<thead>
<tr>
<th></th>
<th>0 Months (n = 25)</th>
<th>3 Months (n = 24)</th>
<th>7 Months (n = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MADRS</td>
<td>28.6 (6.9)</td>
<td>21.8 (10.7)</td>
<td>16.7 (12.8)</td>
</tr>
<tr>
<td>SDS</td>
<td>59.5 (9.5)</td>
<td>53.5 (9.9)</td>
<td>46.5 (12.9)</td>
</tr>
<tr>
<td>AMT Positive</td>
<td>2.32 (1.4)</td>
<td>1.67 (1.4)</td>
<td>2.22 (1.4)</td>
</tr>
<tr>
<td>AMT Negative</td>
<td>2.24 (1.3)</td>
<td>1.13 (1.3)</td>
<td>1.83 (1.4)</td>
</tr>
</tbody>
</table>

### Table 2. Correlations Between MADRS, SDS, CTQ, and AMT Scores at Baseline and Follow-up (3 and 7 months)

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>3 Months</th>
<th>7 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MADRS</td>
<td>SDS</td>
<td>MADRS</td>
</tr>
<tr>
<td>MADRS</td>
<td>—</td>
<td>—</td>
<td>.53*</td>
</tr>
<tr>
<td>SDS</td>
<td>.75‡</td>
<td>—</td>
<td>.61†</td>
</tr>
<tr>
<td>AMT Positive</td>
<td>.05</td>
<td>.01</td>
<td>-.02</td>
</tr>
<tr>
<td>AMT Negative</td>
<td>.03</td>
<td>-.07</td>
<td>-.35</td>
</tr>
<tr>
<td>CTQ</td>
<td>.03</td>
<td>-.02</td>
<td>-.02</td>
</tr>
</tbody>
</table>

* $P < .5$; † $P < .1$; ‡ $P < .001$ (2-tailed).
we found no significant associations between CTQ score and overgeneral recall at 3 and 7 months.

**Prediction of Depression Severity at Follow-up**

Using the average of MADRS scores at 3 and 7 months as the dependent variable, a multiple hierarchical regression was conducted to determine whether selected variables were able to predict depression severity at follow-up. In general, variables like high baseline depression severity and longer duration of current episode are considered to affect the course of the index episode in a negative way. Accordingly, these variables were entered on the first step in the analysis. However, they were not able to predict depression severity at follow-up ($F(2,19) = 1.3, P = .29$). On the next step, the ability of autobiographical memory to predict course was tested by including baseline negative and positive AMT scores in the model. This resulted in a significant 29% increase in explained variance; $F$ change $(4,17) = 1.76, P = .04$.

In the final equation, more specific memory responses to negative cue words predicted less severe depression at follow-up. The other variables did not contribute significantly to the model. Results of this analysis are summarized in Table 3.

The same procedure was repeated using the average of SDS scores at 3 and 7 months as dependent variable. The overall regression equation was significant ($R^2 = .41, F(4,18) = 3.10, P = .04$). In this equation, again only more specific responses to negative cue words predicted less severe depression at follow-up ($B = -4.20, 95\% CI = -7.67 to -.73$). Additionally, we tested whether AMT performance at 3 months predicted outcome at 7 months after controlling for severity of symptoms at 3 months. No significant associations were found.

To address the differences between the nonsignificant univariate correlations (baseline AMT scores and depression at follow-up) presented in Table 2 and the significant results from the multiple regression analyses, additional analyses were performed. Instead of using the average of symptom severity at 3 and 7 months, separate regression analyses were conducted with symptom severity at 3 and 7 months as dependent variables (although such analyses are less reliable as argued in the Method section). After controlling for baseline depression severity and episode duration, higher negative AMT baseline scores again predicted lower MADRS scores at 3 months ($B = -4.67, 95\% CI = -8.63 to -.76$), and almost reached significance at 7 months ($B = -5.38, CI = -11.04 to .27, P = .06$). In a similar analysis for SDS scores, a higher negative AMT score reached borderline statistical significance for symptom severity at 3 months ($B = -3.11, 95\% CI = -6.32 to .11, P = .06$), and at 7 months ($B = -5.33, CI = -10.87 to .21, P = .06$). Positive AMT baseline score was not associated with symptom severity at follow-up in any of these analyses. These results suggest that the above mentioned differences between the univariate correlations and initial multiple regression analyses, may be due to both averaging the follow-up data and inclusion of initial depression severity and episode duration as explanatory variables.

**DISCUSSION**

The results of the present study can be summarized as follows. First, AMT performance appears to be relatively stable over time although the clinical condition of the sample improved significantly during follow-up. Furthermore, AMT performance at baseline was not related to depression severity, and...
while correlations with self-reported childhood trauma were predominantly absent or present in an other direction than was anticipated (i.e., higher autobiographical specificity on negative items was related to higher scores on the CTQ). Third, specific autobiographical responses to negative items predicted better prognosis, whereas autobiographical specificity on positive items was not related to prognosis.

Before considering the implications of the current findings, it is worth noting several limitations of the study. These include the small sample size, the reliance on a self-report measure to assess childhood trauma, and the naturalistic, although vigorous, treatment of the subjects. Furthermore, we did not include educational level in our analyses, although it predicts memory specificity to a certain extent. The main reason for doing so was to increase comparability with previous studies, as these did not include educational level in analyses. Educational level was also omitted because it potentially confounds with other variables that may be of influence on course. Inclusion of all these variables would call for unacceptable complex analyses given the small sample size.

Our finding that AMT performance is relatively stable over time even when depression severity decreases is in line with other studies that addressed this issue. The only significant effect in both types of analysis that we used pertained to a decrease in specificity to negative AMT items from baseline to 3 months follow-up. Perhaps, this particular follow-up version contained a more abstract cue word content, as average scores for responses to both positive as negative cues at this point in time were lower in the whole sample, despite clinical improvement. It was shown in the past that abstract cue words may make it more difficult for subjects to generate specific memories. Additionally, subjects did not differ in their response to cue words of different valence, which is in line with most previous research.

As to the connection between autobiographical recall and traumatization, we found a higher level of childhood traumatization to be associated with better baseline memory performance to negative cue words and not related to responses to positive cue words. However, there were no associations between level of traumatization and memory performance at both follow-up measurements. Taken together with the clear lack of such association in a larger mixed sample, one may cautiously conclude that childhood trauma appears not an important antecedent of overgeneral recall. This is difficult to reconcile with the only other study that explicitly addressed this issue. That study did find a relation between early childhood traumas and overgeneral autobiographical memory performance. This discrepancy might be explained by several reasons. First, the level of self-reported traumatization in our sample was rather low (mean CTQ score = 8.99, SD = 2.18; range, 6.09 to 13.3). It remains possible that the level of trauma in our sample varied insufficiently to detect an influence on autobiographical memory performance. Second, given the strong evidence that overgeneral recall is closely tied to intrusions of past experiences rather than to childhood trauma per se, it might well be the case that our sample suffered less from intrusive memories. Third, as the authors already pointed out, it cannot be ruled out that group differences other than history of abuse (e.g., more suicide attempts, more previous episodes, etc. in the abused group) were responsible for their results. Ideally, the association between childhood trauma and overgeneral recall should be readdressed in a large study, testing this association on several occasions while controlling for potential confounders.

We found more specific responses to negative cue words to be predictive of less depression severity at follow-up using both clinician-rated and self-report instruments in our main analyses. This finding is discrepant from previous studies. Brewin et al. were not able to find an association between overgeneral recall and course using self-reported severity of depression. This may be explained by the fact that in this study, overgeneral recall and depression severity appeared to be less marked than in our sample. Thus, in the Brewin et al. study, floor effects may have made it more difficult to detect a link between overgeneral recall and outcome. As in the Brittlebank et al. study, we found a clear association between memory performance and clinician-rated outcome. However, in our study, more specific AMT performance on negative cue words at baseline predicted better outcome, while Brittlebank et al. found more specific AMT performance on positive cue words predicted a favorable outcome. The fact that the 7 months
follow-up data in the Brittlebank et al. study included only 13 subjects might be a reason for these discrepancies. Statistical analysis of such a small sample may result in spurious findings as these authors themselves admitted. However, a recent study in 21 subjects diagnosed with seasonal affective disorder (SAD) replicated the findings from the Brittlebank et al. and Brewin et al. More specific recall to positive cue words was related to a better prognosis, but only for the clinician-rated symptom severity data. Dalgleish et al. suggested that this difference can be explained by a stronger focus of clinician-rated measures on somatic vegetative symptoms as opposed to the more cognitive focus of the Beck Depression Inventory (BDI). As we used the SDS for self-report, which taps more somatic vegetative symptoms than the BDI, this may also explain why we found an association between overgeneral recall and self-reported course of symptoms. Another explanation for the discrepancies between the current and other studies may be that results depend on characteristics of the population under study. If somatic vegetative symptoms are indeed moderating factors in the association between overgeneral recall and course, it may be that this also explains other differences. Our sample consisted mainly of moderately severe, nonendogenous, depressed outpatients, whereas samples in most other studies were characterized by inpatients with endogenous depression or subjects with SAD. The latter samples suffer by definition more from somatic vegetative symptoms. Perhaps such symptoms are associated differently with overgeneral recall to positive and negative cue words. Furthermore, etiological factors may be relevant. As recent life events are known to be less involved in the onset of recurrent endogenous depression and SAD than in nonendogenous depression, unknown contextual and pathophysiological factors may account for the diverging results. For example, the presence of recent life events perhaps influences both type of overgeneral recall and course, leading to different relations between these variables and accordingly divergent results between studies. However that may be, the current findings along with most previous studies demonstrate that there is a strong link between autobiographical memory specificity and the course of major depressive disorder.

There are several explanations for this link. To begin with, problems with retrieving specific memories beyond the level of intermediate descriptions impairs the use of effective problem-solving strategies. This fits well with our finding that in particular specific responses to negative cue words predict better outcome. Onset of nonendogenous depressive disorders is often related to severe life events. More specific memories of these events might be associated with better access to effective problem-solving strategies and eventually a more favorable prognosis. A second explanation is provided by studies indicating that intrusive cognitions about past events are important determinants of overgeneral recall, because such intrusions consume working memory capacity. It may be that intrusions promote rumination about the past and oneself, and that such ruminative self-focus undermines autobiographical memory specificity. Interestingly, several studies show that rumination both exacerbates and prolongs episodes of depressive disorders and dysphoric mood. Clearly, interrelations between problem-solving strategies, intrusive cognitions, rumination, and autobiographical memory specificity warrant further studies.

Our data support the view that lack of specificity of autobiographical memory is stable, not related to childhood trauma per se or severity of depression, and can be understood as a trait. Therefore, this phenomenon may act as a vulnerability factor that modulates the long-term course of depressive disorders. As to the clinical implications, several remarks are in order. First, patients with poor specific recall of negative events are at risk for chronicity. Thus, it may well be that standard cognitive therapy is likely to fail in these patients. Perhaps other therapeutic strategies, directed at changing their nonspecific cognitive style, are more suited for these patients. Furthermore, continuation and maintenance of a treatment regimen might be indicated in these patients as they may be more susceptible to relapse and recurrence. In patients with more specific recall of negative events, treatment aimed at processing and discussing these experiences in terms of problem-solving strategies (e.g., interpersonal psychotherapy) seems more indicated.

ACKNOWLEDGMENT

The authors would like to thank G. van Breukelen for his statistical advise and L. Finders for her assistance in data collection.
REFERENCES