FALSE RECOGNITION IN WOMEN REPORTING RECOVERED MEMORIES OF SEXUAL ABUSE

Susan A. Clancy, Daniel L. Schacter, Richard J. McNally, and Roger K. Pitman

Department of Psychology, Harvard University, and Veterans Affairs Medical Center, Manchester, New Hampshire, and Harvard Medical School

Abstract—False recognition—the mistaken belief that one has previously encountered a novel item—was examined in four groups of subjects: women reporting recovered memories of childhood sexual abuse, women who believe that they were sexually abused as children but who cannot recall this abuse (the “repressed” group), women who were sexually abused as children and always remembered the abuse, and women with no history of childhood sexual abuse. Subjects were administered a Deese/Roediger-McDermott paradigm. The results suggest that the recovered-memory group was more prone to false recognition than the other groups. In addition, women reporting recovered and repressed memories showed greater reduction in false recognition across study trials than did other subjects, perhaps reflecting strategic changes in performance.

Reports of childhood sexual abuse (CSA) are plagued by controversy. At the heart of the controversy is the claim that memories of traumatic CSA can be repressed and later recovered in adulthood. Some theorists believe that children exposed to sexual trauma develop dissociative coping skills that enable them to forget memories too upsetting to be consciously accessible (e.g., Terr, 1991). According to this perspective, these repressed (or dissociated) memories of childhood trauma nevertheless influence thought, behavior, and physiological processes (e.g., Brown, Schefflin, & Hammond, 1998) and can be retrieved years later with scant distortion in details (e.g., Terr, 1994).

Other psychologists question these claims. They point to laboratory research demonstrating that human memories are vulnerable to distortion, that illusory memories can be created (see Schacter, 1999, for a review), and that there is little evidence that memories of trauma obey different psychological laws than do memories of nontraumatic events (Shobe & Kihlstrom, 1997). Finally, underscoring the malleability of memory, skeptics have warned that therapies designed to recover memories of repressed abuse may inadvertently foster false memories of trauma (e.g., Loftus, 1993).

Strikingly, in this highly politicized and volatile debate, there are no published experimental data concerning memory functioning in people who report having recovered memories of CSA. Reports of recovered memories, such as Herman and Schatzow’s (1987), are vulnerable to the criticism that the recovered memories were not corroborated and that they may have been inadvertently created in therapy (Ofshe & Watters, 1994). Experimental studies of false memories in nontraumatized volunteers (e.g., Roediger & McDermott, 1995) are criticized because the memories did not involve serious trauma, and therefore may be irrelevant to CSA (Freyd & Gleave, 1996).

False recognition—the mistaken belief that one has previously encountered a novel item—has been well established experimentally (Schacter, Norman, & Koutstaal, 1998). Roediger and McDermott (1995), modifying Deese’s (1959) procedure, demonstrated robust false recognition in the laboratory. In their paradigm, subjects hear a series of word lists; each is composed of associates to a single nonpresented “theme word.” For example, one list comprised words associated with sweet (e.g., sour, candy, sugar, bitter). Following list presentation, subjects received a recognition test composed of studied words, nonpresented theme words (hereafter called false targets), and other nonstudied words. Roediger and McDermott reported high levels of false recognition (e.g., 80%) to false targets in college students. Other researchers have used this paradigm to investigate false recognition in amnesic patients (Schacter, Verfaellie, & Pradere, 1996) and older adults (Norman & Schacter, 1997).

People may differ in their proneness to create false memories. For example, individuals who report frequent episodes of dissociation (disruptions in consciousness) may be especially likely to confuse the products of imagination and the products of perception. Consistent with this hypothesis, people who score high on the Dissociative Experiences Scale (DES; Bernstein & Putnam, 1986) are vulnerable to memory distortions in several laboratory paradigms (e.g., Heaps & Nash, 1999; Hyman & Billings, 1998), including the Deese/Roediger-McDermott (DRM) paradigm (e.g., Winograd, Peluso, & Glover, 1998).

In the present study, we used a variant of the DRM paradigm to investigate false recognition in four groups: women who reported recovering memories of CSA (i.e., recovered-memory group); women who believe that they were sexually abused as children, but who have no memories of the abuse (i.e., repressed-memory group); women who were sexually abused as children and always remembered the abuse (i.e., continuous-memory group); and women with no histories of sexual abuse (i.e., control group). Our use of the terms repressed and recovered reflects the experience of our subjects; it does not imply our belief or disbelief in the veracity of their reports.

We tested two hypotheses. Because subjects who report recovered memories of CSA score higher on the DES than do control subjects who report no abuse history (Clancy, McNally, & Schacter, 1999), and because DES scores predict memory distortion in the DRM paradigm (e.g., Winograd et al., 1998), we tested whether recovered-memory subjects are more prone to exhibit false recognition than other subjects. Inclusion of a group of subjects who have always remembered their abuse and a group of subjects who believe they have been abused (but who have no autobiographical memories of abuse) enabled us to test whether any false-recognition effects are confined to subjects who have reportedly repressed and then recovered their CSA memories.

We also tested whether subjects reporting recovered memories become capable of suppressing their proneness to exhibit false recognition. Despite its robustness, the DRM false-memory effect can be...
counteracted by certain encoding manipulations (Schacter, Israel, & Racine, 1999), retrieval instructions (McDermott & Roediger, 1998), or repeated testing on the same lists (Kensinger & Schacter, 1999; McDermott, 1996). In our experiment on imagination inflation (Clancy et al., 1999), recovered-memory subjects seemed especially vigilant, and they tended to be less vulnerable to memory distortion following guided imagery than were control subjects. Because in the present study we administered a recognition test after each list, we were able to test whether subjects suppressed or reduced their false-recognition rates across trials. To the extent that the motivation for vigilance may be similar for both the recovered- and repressed-memory groups (i.e., both groups might be sensitive to the possibility that the experiment involved attempts to create false memories), we hypothesized that both groups would be especially capable of suppressing their tendency to exhibit false recognition.

**METHOD**

**Subjects**

The recovered, repressed, and control groups were recruited from the community via newspaper notices saying that Harvard researchers were “seeking adult, female volunteers who either recovered memories of having been sexually abused as children, believe they might have been sexually abused as children but have no memories, or have no history of sexual abuse as children, to participate in a study on memory.” Continuous-memory subjects—those who were sexually abused as children and who always remembered the abuse—were recruited from among those who had participated in other studies by our group (e.g., Orr et al., 1998). Subjects provided written informed consent and were paid for their participation.

Fifteen women reporting recovered memories of CSA, 15 women who believed that they were victims of CSA but who had no memories, 12 women with continuous memories of CSA, and 15 nonabused control subjects participated. They were administered a semistructured interview in which they were asked whether they were sexually abused as children, and if so, how old they were when the abuse occurred, what type of abuse was experienced, who the perpetrator was, over what period of time they had been abused, and how old they were when they recovered the memory. The 15 control subjects reported no histories of CSA. Subjects were also asked their age and their highest level of education attained.

To characterize our subjects and to foster their comparison with subjects in previous trauma studies, we asked them to complete the DES (Bernstein & Putnam, 1986), the Beck Depression Inventory (BDI; Beck & Steer, 1987), and the CMISS (Vreven, Gudanowski, King, & King, 1995), the civilian version of the Mississippi Scale for Combat-Related Posttraumatic Stress Disorder (Keane, Caddell, & Taylor, 1988). Designed to assess disruptions in consciousness, the DES contains items related to depersonalization, memory lapses, and absorption. The CMISS assesses symptoms associated with posttraumatic stress disorder (PTSD; e.g., intrusive thoughts, psychological numbing), and the BDI assesses symptoms of depression. Psychiatrically impaired survivors of traumatic events score high on all these self-report measures (Orr et al., 1990).

One-way analyses of variance (ANOVAs) indicated that the groups differed in age, depressive symptoms, and PTSD symptoms, but not in years of education or dissociative symptoms (see Table 1). According to follow-up tests, the continuous-memory group was older than the control group, and the repressed- and recovered-memory groups reported more symptoms of depression and PTSD than did the continuous-memory and control groups (ps < .05).

**Materials**

Twenty-four lists of semantic associates, each composed of 15 words, were drawn from previous studies (Norman & Schacter, 1997; Roediger & McDermott, 1995). Twelve of these lists were used in the study phase. Four of the 12 lists were modified so they contained only 8 semantic associates, and 4 other lists were modified so they contained only 2 semantic associates. The remaining 4 lists were left unmodified (i.e., they contained all the original 15 semantic associates). We varied the numbers of associates because research has shown a direct relationship between the number of studied associates and false recognition (e.g., Robinson & Roediger, 1997). Because 15-associate lists yield rates of false recognition as high as 80% (e.g.,

<table>
<thead>
<tr>
<th>Table 1. Demographic and psychometric data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>BDI</td>
</tr>
<tr>
<td>CMISS</td>
</tr>
<tr>
<td>DES</td>
</tr>
</tbody>
</table>

*Note.* CMISS = Civilian version of the Mississippi Scale for Combat-Related Posttraumatic Stress Disorder (possible range: 35–175); BDI = Beck Depression Inventory (possible range: 0–64); DES = Dissociative Experiences Scale (possible range: 0–100). Standard deviations are in parentheses.
Roediger & McDermott, 1995), we hoped to counteract a possible ceiling effect by constructing lists that contained fewer semantic associates. Therefore, a 15-associate list comprised 15 words related to a single theme; an 8-associate list comprised 8 words related to a single theme plus 7 words drawn from nonstudied lists and unrelated to any theme; and a 2-associate list comprised 2 words related to a single theme plus 13 words drawn from nonstudied lists and unrelated to any theme. Prior to this experiment, we created another 18 word lists containing 15 semantic associates each. These lists provided the unrelated words incorporated into the 8- and 2-associate lists. On all study lists, associates were presented in decreasing order of associative strength with respect to the nonpresented false targets on which all presented associates converged. For the 8-associate lists, the 8 weakest associates to the false target were presented following 7 unrelated words taken from nonstudied lists. For the 2-associate lists, the 2 weakest associates to the false target were presented following 13 unrelated words taken from nonstudied lists.

Three different sets of study lists were created for counterbalancing purposes. Each set contained 4 lists with 15 semantic associates, 4 lists with 8 semantic associates, and 4 lists with 2 semantic associates. Subjects studied all 12 lists in a set. A 12-item recognition test followed each list presentation. The recognition test for each list included 6 studied items (the 4th, 6th, 8th, 10th, 14th, and 15th items from the studied lists). For all three types of study lists (15, 8, and 2 associates), the 2 associates presented in the 14th and 15th positions on the study lists were tested and labeled true targets. The other 6 items on the recognition test had not been studied. One of these nonstudied items was the lure on which the studied items semantically converged (the false target). The other 5 items were from lists that were not studied and served as control words for the studied words; 2 were semantically related to each other and were labeled true-target controls (the 14th and 15th items from a nonstudied list), 2 were not semantically related to each other, and 1 was the related lure on which all items from one of the nonstudied lists semantically converged (the false-target control).

Procedure

All subjects were tested individually. Before presentation of the first study list, subjects were told that a series of words would appear on the computer screen and that they should try to remember the words. Subjects were then told that following each list there would be a recognition test during which they would see words on the computer screen, and that some of the words would be those that they studied, and some would be new, nonstudied words. During the recognition test, subjects were asked to press the “R” key, for “remember,” or the “N” key, for “new,” to indicate for each word whether they remembered studying it or it was new. Following these directions, subjects were presented with List 1. Each word remained on the screen for 3 s, so presentation of the list lasted 45 s. The recognition test for List 1 was administered 2 min after List 1 was presented. During the 2 min after the study list was presented, subjects waited while the experimenter accessed the recognition test on the computer. On the recognition test, each word appeared on the computer screen and remained there until the subject responded. When the subject completed the recognition test for List 1, List 2 was presented, followed 2 min later by the recognition test for List 2. This continued until all 12 lists and all 12 recognition tests had been completed.

RESULTS

For each subject, we calculated two main indices of performance: the true-recognition rate (true targets – true-target controls) and the false-recognition rate (false targets – false-target controls). The true-recognition rates and false-recognition rates as a function of group (recovered, repressed, control, and continuous) and number of semantic associates per list (15, 8, and 2) are shown in Table 2. Statistical tests are two-tailed except for contrast analyses that were based on specific directional predictions motivated by previous research.

True Recognition

Because we had no predictions about true recognition, we conducted a mixed-design ANOVA with four levels of a between-subjects factor (subject group) and three levels of a within-subjects factor (number of semantic associates per list: 15, 8, and 2). All effects were nonsignificant: group, F(3, 53) = 0.73, p = .88; number of semantic associates, F(3, 53) = 2.00, p = .14; and Group × Number of Semantic Associates, F(2, 106) = 1.05, p = .40.

False Recognition

First, we conducted a mixed-design ANOVA with four levels of a between-subjects factor (subject group) and three levels of a within-subjects factor (number of semantic associates per list: 15, 8, and 2). There was no significant effect of group, F(3, 53) = 1.74, p = .17, and no significant interaction of group and number of semantic associates, F(6, 106) = 1.68, p = .13. There was a significant effect of number of semantic associates, F(2, 106) = 62.47, p = .01, so we analyzed the 15-, 8-, and 2-associate lists separately.

Table 2. True-recognition and false-recognition data for each group by list type

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of associates</th>
<th>15</th>
<th>8</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>True recognition (true targets – true-target controls)</td>
<td>Recovered memory</td>
<td>.86</td>
<td>.85</td>
<td>.81</td>
</tr>
<tr>
<td></td>
<td>(15)</td>
<td>(.15)</td>
<td>.12</td>
<td>(.16)</td>
</tr>
<tr>
<td></td>
<td>Repressed memory</td>
<td>.84</td>
<td>.85</td>
<td>.83</td>
</tr>
<tr>
<td></td>
<td>(16)</td>
<td>(.16)</td>
<td>.16</td>
<td>(.13)</td>
</tr>
<tr>
<td></td>
<td>Continuous memory</td>
<td>.92</td>
<td>.78</td>
<td>.81</td>
</tr>
<tr>
<td></td>
<td>(11)</td>
<td>(.11)</td>
<td>.31</td>
<td>(.12)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>.88</td>
<td>.86</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td>(12)</td>
<td>(.12)</td>
<td>.18</td>
<td>(.14)</td>
</tr>
<tr>
<td>False recognition (false target – false-target controls)</td>
<td>Recovered memory</td>
<td>.72</td>
<td>.68</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>(26)</td>
<td>(.26)</td>
<td>.27</td>
<td>(.18)</td>
</tr>
<tr>
<td></td>
<td>Repressed memory</td>
<td>.57</td>
<td>.45</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>(24)</td>
<td>(.24)</td>
<td>.22</td>
<td>(.25)</td>
</tr>
<tr>
<td></td>
<td>Continuous memory</td>
<td>.62</td>
<td>.52</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>(38)</td>
<td>(.38)</td>
<td>.25</td>
<td>(.23)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>.52</td>
<td>.38</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>(42)</td>
<td>(.42)</td>
<td>.42</td>
<td>(.18)</td>
</tr>
</tbody>
</table>

Note. Standard deviations are in parentheses.
For the 8- and 15-associate lists, we conducted focused one-tailed contrasts to test our two specific predictions and to compute effect size \( r \) (Rosenthal & Rosnow, 1985). Because the pattern of results was identical in the 8- and 15-associate lists (recovered-memory group showing the highest levels of false recognition, followed by the continuous-memory group, the repressed-memory group, and the control group, respectively), we also present the data for the 8- and 15-associate lists combined, as well as separately. Because the 2-associate lists produced relatively low levels of false recognition in all groups (see Table 2), we omitted them from subsequent analyses.

According to our first hypothesis, recovered-memory subjects should be particularly vulnerable to memory distortion and should show higher rates of false recognition than the other groups. To test this hypothesis, we applied contrast weights of 3, -1, -1, and -1 to the recovered, repressed, continuous, and control groups, respectively. We confirmed this hypothesis for the 8-associate lists, \( t(53) = 2.54, p = .01, r = .32 \), and the results for the 15-associate lists were nearly significant as well, \( t(53) = 1.48, p = .07, r = .20 \). When both kinds of lists were combined, the hypothesis was confirmed, \( t(53) = 2.38, p = .01, r = .31 \).

### Suppression of False Recognition

Because each list type (8 or 15 associates) was presented four times, we examined whether subjects suppressed their false-recognition rates with each subsequent list presentation and test. Findings from our previous study (Clancy et al., 1999) made us curious whether the recovered- and repressed-memory subjects would exhibit heightened vigilance about exhibiting false-memory effects in the laboratory. To test whether these groups were more likely than the other two groups to suppress their false-recognition rates over the course of the experiment, we calculated a mean false-recognition rate for each subject by first multiplying the false-recognition rate for the 2-, 8-, and 15-associate lists by -2, 1, and 1, respectively, and then summing the products. Applying contrast weights of 3, -1, -1, and -1 to the mean \( L \) scores of the recovered-memory, repressed-memory, control, and continuous-memory groups, respectively, we obtained strong support for this hypothesis, \( t(56) = 2.89, p = .01, r = .37 \). In short, whereas other groups’ false-recognition data correspond to a function whereby false alarm rates increase with the number of within-list associates, recovered-memory subjects seem to have a lower threshold for false recognition; their false alarm rate for the 8-associate lists was almost equal to that for the 15-associate lists.

Because the groups differed not only in their reports of CSA, but also on measures of PTSD and depression, we tested whether false recognition (calculated by combining rates from the 8- and 15-associate lists) was related to scores on the BDI and CMISS. It was not: BDI \( r = .01 \), CMISS \( r = -.01 \). Moreover, the correlation between age and false recognition was also nonsignificant, \( r = .10 \). The correlation between DES scores and false recognition was, however, significant, \( r = .32, p = .01 \).

### DISCUSSION

Women who reported recovered memories of CSA were more prone than other subjects to exhibit false recognition of semantic associates. Because recovered-memory subjects showed higher false recognition than either the continuous-memory subjects or the repressed subjects, these results cannot be entirely explained by cognitive impairments related to reported CSA.

Recovered-memory subjects exhibited a lower threshold for false recognition than did the other groups. Whereas other groups’ false-recognition rates conformed to a linear function (highest for 15-associate lists, lowest for 2-associate lists), recovered-memory subjects’ false-recognition rates conformed to a function whereby they reached an asymptote in the 8-associate lists.

Researchers have begun to delineate the mechanisms involved in the creation of false memories. One process implicated is source memory or source monitoring—remembering how, when, and where a memory was acquired. Recollections of perceived events and imagined events can be confused, thereby producing distorted memories.
False Recognition and Recovered Memories

(Johnson, Hashtroudi, & Lindsay, 1993). Schacter et al. (1996) have hypothesized that robust false recognition occurs when subjects retain the common semantic features of presented words—the “gist” of each set of semantic associates—but do not encode or retain distinctive details of individual items (Payne, Elie, Blackwell, & Neuschatz, 1996). Recently, Brainerd and Reyna (1998), using a modified DRM paradigm, showed that events that were not experienced, but were consistent with the gist of the experienced events, could be more memorable than the actual events.

It is unknown whether findings concerning false recognition of semantic associates in a laboratory paradigm bear directly on illusory memories of childhood trauma. To the extent that some false memories reflect the gist of past experience, illusory memories of CSA may be accurate representations of some aspect of a person’s past. Memories can be accurate in the sense that they refer abstractly to an experience, yet can contain many details that arise from source-monitoring errors rather than from experience (Schacter et al., 1998).

We previously found that women reporting recovered memories of CSA were not more likely than control subjects to exhibit memory distortion after guided imagery (Clancy et al., 1999). Indeed, some women reporting recovered memories seemed suspicious about the study and vigilant concerning the possible memory-distorting effects of guided imagery. In the present study, the recovered-memory group and the repressed group were more likely than other groups to suppress their levels of false recognition as the experiment progressed; that is, their false-recognition rates were lower at the end of the experiment than at the beginning. However, such suppression did not result in lower overall rates of false recognition, as the recovered group still had significantly higher false recognition (averaged across all lists) than did any other group.

In previous studies that showed reduced false recognition after repeated study and testing of DRM lists, the same lists were presented and tested repeatedly (Kensinger & Schacter, 1999; McDermott, 1996). Thus, false-recognition suppression in these experiments is likely attributable to increasingly detailed episodic memory for words that were actually presented (Kensinger & Schacter, 1999). In our experiment, by contrast, different lists were presented and tested on successive trials, so false-recognition suppression cannot be attributable to increasingly detailed episodic memories of repeated study items. Rather, suppression is more likely attributable to strategic factors, such as vigilance concerning the purpose of the experiment.

DES scores and false recognition were significantly correlated, consistent with the results of other studies. Heaps and Nash (1999) reported a correlation of .34 (p < .001) between DES scores and imagination after guided imagery, and Hyman and Billings (1998) found the correlation between DES scores and false-memory creation to be .48 (p < .001). Winograd et al. (1998) used the DRM paradigm and reported a correlation of .32 (p < .05) between DES scores and false recognition of related lure words.

Because the data are correlational, they do not demonstrate any causal connection, or even temporal sequencing, among false recognition, proneness to dissociation, and recovered memories of CSA. One possible interpretation is that CSA reported by the recovered-memory group actually occurred, was subsequently forgotten, and then was recovered, and that these experiences also induced high dissociation and a proneness for false recognition as detected by the DRM paradigm. However, such an interpretation cannot explain why the continuous-memory group showed less false recognition than the recovered-memory subjects did. Such an interpretation would require individual differences in either the nature of the abuse, the person who experienced it, or both. Another possible interpretation is that individuals who are prone to false recognition by virtue of high dissociation are more likely to falsely “remember” CSA experiences that were only suggested or imagined.

Research on memory functioning in people who report recovered memories of CSA is needed. Prior research on memory distortion has been chiefly confined to laboratory studies on college students, amnestic, the elderly, and, more recently, trauma victims with PTSD (e.g., Bremner, Shobe, & Kihlstrom, 1999). Although the advantages of laboratory research include good internal validity and experimental control, the disadvantages include potential artificiality. Our experiment concerns false recognition for nontraumatic events; therefore, care must be taken when extrapolating our findings to clinical settings.

However, the results are consistent with the hypothesis that women who report recovered memories of sexual abuse are more prone than others to develop certain types of illusory memories. They appear to have been more drawn by gist than were the other groups.

Acknowledgments—We thank Eric Loken and Christopher F. Chabris for their statistical advice.

REFERENCES

Bremner, C.J., & Reyna, V.F. (1998). When things that were not experienced are easier to “remember” than things that were. Psychological Science, 9, 484-489.


(RECEIVED 3/9/99; REVISION ACCEPTED 6/15/99)