False-Memory Creation in Children and Adults

Theory, Research, and Implications

Edited by

David F. Bjorklund
Florida Atlantic University

LEA
LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Mahwah, New Jersey London

2000
In recent years, we have been interested in the conditions under which false memories are likely to be planted in adults and children and the conditions under which adults and children are more likely to be resistant to the implantation of false memories. In the process of accumulating results that address this research question (see, e.g., Pezdek, Finger, & Hodge, 1997; Pezdek & Hodge, 1999; Pezdek & Roe, 1994, 1997), we realized that we had a fairly large corpus of data on a different issue, that is, qualitative and quantitative differences between memories for true and false events. Although some of these data are included in the various articles that we have published, we decided that there was a need to synthesize these findings in one source, along with the findings of other researchers who have compared phenomenological qualities of memories for true and false events. This chapter was written for this purpose.

In the three sections of this chapter, we review three methods that have been proposed to differentiate between accounts of true and false events. Research on the effectiveness of these three methods is included as well. We should say up front that although some measures are more sensitive to differences between accounts of true and false events than others, we agree with Steve Ceci and others that there is no Pinocchio test for
determining whether accounts are true or false (cf. Ceci, Loftus, Leicht-
mant, & Bruck, 1994). We mention this disclaimer here for those who ap-
proach this chapter with the anticipation that we offer some fail-proof

technique for assessing the veracity of reported memories. Such a tech-
nique is not available. Also, as we discuss in this chapter, the problem is
further complicated by the fact that the cognitive characteristics of mem-
ories for true events and false events are overlapping functions; memories
for true and false events share numerous features. Nonetheless, we sug-
gest that a review and critical evaluation of the research on this topic has
utility for practitioners as well as for researchers investigating the nature
of memories for true and false events.

The first method to be discussed for differentiating between accounts
of true and false events is the Statement Validity Assessment (SVA) that
was developed by David Rasin and Max Steller and based on the early
work of Udo Undeutsch in Germany. In the second section, the method
derived from Marcia Johnson’s model of reality monitoring is reviewed.
According to the reality monitoring framework, memories for real and
fictitious events will differ qualitatively. On the basis of this work, the
Memory Characteristics Questionnaire (MCQ) was developed (Johnson,
Foley, Suengas, & Raye, 1988) to assess the qualitative differences between
accounts of real and fictitious events.

In the third section of this chapter, a number of studies are considered
that have examined the conditions under which false memories are more
likely to be suggestively planted. Data have been collected on people’s ac-
counts of events that they believed to be real, although some were true
and some were false. Characteristics of the memory accounts for these
two types of events have been compared. The major difference between
the data generated in the reality monitoring research and data generated
in the false memory research is that the events for which memory was as-
sessed in the reality monitoring research included real events that had oc-
curred, and they were compared with, for example, memories for dreams,
fantasies, and unfulfilled intentions. On the other hand, in the false mem-
ory research, the events for which memory was assessed included similar
types of events that either had occurred (true events) or had not occurred
(false events.)

Several related areas of research are not included in this chapter. Our
goal in this chapter is to assist professionals in discriminating between
memories for true versus false events, both of which individuals believe to
be true. Thus, research on lie detection techniques, credibility assessment,
and malingering are not considered in this chapter. We have also avoided
much of the research comparing accounts of true and consciously fabri-
cated false events because consciously fabricated accounts are considered to be different from events that are erroneously believed to be true.

STATEMENT VALIDITY ASSESSMENT

Much of the early work on what is now known as SVA was conducted in Germany in the 1950s to assist the courts in differentiating between eyewitness accounts that were based on real events and those that were fabricated (see Undeutsch, 1989, for a history of this early work). This work was known as Statement Reality Analysis. The principle hypothesis underlying this work is that accounts of actual experiences differ in content and quality from accounts of experiences that were confabulated, suggested or imagined. Steller (1989) termed this the Undeutsch hypothesis. On the basis of this early work, much of which was conducted by Undeutsch, the Supreme Court of Germany ruled in 1955 that psychological assessments and interviews would be required to determine the credibility of child witnesses in most child sexual abuse cases. As a result of this ruling and a similar one at the time in Sweden, a rash of content criteria was developed to assess the credibility of statements made by alleged sexual abuse victims. However, most of these assessment systems were not systematically developed using sound psychological principles, nor were they scientifically validated.

Current efforts in this area, led primarily by David Raskin (see, e.g., Raskin & Esplin, 1991) and Max Steller (see, e.g., Steller, 1989), represent significant improvements in the application of scientific principles of inquiry and validation. Together, these two researchers began the development of the SVA in 1985. There are three components to the SVA: (a) a structured interview with the witness, (b) a criterion-based content analysis (CBCA) of the interview data, and (c) integration of the CBCA with the information obtained from a set of questions that compose the Validity Checklist. It is important to note that the SVA does not assess the general credibility of a witness but rather assesses the validity of the statements produced by a witness.

It is beyond the scope of this chapter to detail the components of the SVA; however, Table 4.1 summarizes the 18 content criteria for the CBCA. There are no specific rules regarding how many of the criteria must be satisfied to support the conclusion that a particular account is true. As a general rule, however, to support the conclusion that an account is true, it is necessary that the three general characteristics specified at the top of Table 4.1 be satisfied. That is, the account must be logically coherent, have
| TABLE 4.1 |
| Content Criteria for Statement Analysis |

**General Characteristics**

1. Logical structure — Is the statement coherent? Is the content logical? Do the different segments fit together? (Note: Peculiar or unique details or unexpected complications do not diminish logical structure.)
2. Structured production — Are the descriptions unconstrained? Is the report somewhat unorganized? Are there digressions or spontaneous shifts of focus? Are some elements distributed throughout? (Note: This criterion requires that the account be locally consistent.)
3. Quantity of details — Are there specific descriptions of place or time? Are persons, objects, and events specifically described? (Note: Repeating does not count.)

**Specific Contents**

4. Contextual embedding — Are events placed in spatial and temporal context? Is the action connected to other incidental events, such as routine daily occurrences?
5. Interactions — Are there reports of actions and reactions or conversation composed of a minimum of three elements involving at least the accused and the witness?
6. Reproduction of speech — Is speech or conversation during the incident reported in its original form? (Note: Unfamiliar terms or quotes are especially strong indicators, even when attributed to only one participant.)
7. Unexpected complications — Was there an unplanned interruption or an unexpected complication or difficulty during the sexual incident?
8. Unusual details — Are there details of persons, objects, or events that are unusual, yet meaningful in this context? (Note: Unusual details must be realistic.)
9. Superfluous details — Are peripheral details described in connection with the alleged sexual event that are not essential and do not contribute directly to the specific allegations? (Note: If a passage satisfies any of the specific criteria 4-18, it probably is not superfluous.)
10. Accurately reported details misunderstood — Did the child correctly describe an object or event but interpret it incorrectly?
11. Related external associations — Is there reference to a sexually toned event or conversation or a sexual nature that is related in some way to the incident but is not part of the alleged sexual offenses?
12. Subjective experience — Did the child describe feelings or thoughts experienced at the time of the incident? (Note: This criterion is not satisfied when the witness responds to a direct question, unless the answer goes beyond the question.)
13. Attribution of accused’s mental state — Is there reference to the alleged perpetrator’s feelings or thoughts during the incident? (Note: Descriptions of overt behavior do not qualify.)

**Motivation-Related Contents**

14. Spontaneous corrections or additions — Were corrections offered or information added to material previously provided in the statement? (Note: Responses to direct questions do not qualify.)
15. Admitting lack of memory or knowledge — Did the child indicate lack of memory or knowledge of an aspect of the incident? (Note: In response to a direct question, the answer must go beyond “I don’t know” or “I can’t remember.”)
a well-organized structure, and include a convincing quantity of specific
details. Logical coherence refers to how well the different segments fit
together and whether the statement makes logical sense. The structure of the
production refers to whether the statement is well organized or whether
there are digressions or shifts of focus. The quantity of specific details refers
to whether descriptions include, for example, time, place, person, and ob-
jects. In a review of the literature on CBCA by Ruby and Brigham (1997),
the criterion of wealth of details was reported to have the widest utility in
differentiating between accounts of true and false events. It has not been
determined, however, how many of the additional 15 criteria are suf-
cient to draw a conclusion regarding the truth value of an account.

The Validity Checklist was developed by Stellar (1989). The items in the
Validity Checklist probe psychological characteristics of the interviewee,
characteristics of the interview, the motivational context of the report,
and the consistency of the report with other statements and lines of evi-
dence. According to Stellar (1989), although this portion of the SVA is
"more subjective and less formalized" (p. 141) than the CBCA, it provides
additional information that can be used to assess the truth value of a wit-
ness’s statements by determining whether the statements are reasonable
in light of what is known about the victim and the alleged acts.

The validity of the SVA has been tested in a number of studies. Landry
and Brigham (1992) compared judges trained to use the CBCA with un-
trained judges in their ability to discriminate videotaped or transcribed
accounts of adults describing a true or a false personally traumatic event.
The CBCA-trained judges were significantly more accurate (55.3%) than
untrained judges (46.9%). In a similar study, Tye, Henderson, and Honts
(1995) reported that judges trained to use the CBCA could correctly clas-
sify 89% of the accounts of 28 children regarding a mock theft they ob-
erved. The children either reported the theft correctly, or they were
encouraged to lie about it. The level of accuracy for CBCA judges compared favorably with that of lay judges who were significantly better than chance but achieved only 56% (Experiment 1) and 65% (Experiment 2) accuracy. Several experimental studies have demonstrated that judges using the CBCA can discriminate between accounts of true events versus accounts of events that subjects were told to fabricate. Yuille (1988) had children tell a story about a true event or make up a story about a plausible but fabricated event. Using the CBCA, judges correctly classified 96% of the stories. In a similar study with adults, Zaparniuk, Yuille, and Taylor (1995) reported that CBCA-trained judges performed significantly better than chance at discriminating between accounts of subjects who saw a videotape of a crime and subjects who were read a description of the crime and were then asked to fabricate an account of the crime as if they had actually observed it. However, the results were significant for only 2 of the 3 judges who coded the accounts.

Tests of the validity of the SVA have come from field studies as well. For example, Esplin, Boychuk, and Raskin (1988) identified 20 children with documented histories of sexual abuse. They identified a different group of 20 children who had reported sexual abuse, but the cases were not confirmed and were strongly suspect; most of these cases involved custody disputes. Judges using the SVA criteria discriminated between these two groups at a highly significant level. In a similar study, Lamers-Winkelman, Buffing, and van der Zanden (1992) examined the accounts of 103 children who had reported having been sexually abused. Their claims of sexual abuse were independently classified as substantiated, highly probable, or unfounded. Judges using the CBCA criteria successfully discriminated among these three groups with older children but not with younger children. This result is not surprising given that the general characteristics of the CBCA are the most discriminating factors (i.e., the logical structure and amount of detail in the account), and preschoolers' narratives are less well structured than those of older children (Fivush & Slackman, 1986; Fivush, Kuebli, & Clubb, 1992).

The most comprehensive field study to date assessing the validity of the SVA was conducted by Lamb, Sternberg, Esplin, Hershkowitz, and Orbach (1997). Because questions have been raised regarding the representativeness of the cases used in a number of the previous studies assessing the SVA (see, e.g., Wells & Loftus, 1991), Lamb and his colleagues selected 98 active cases for which extensive interviews with youth investigators were available over an extended period of time. Using independent information such as medical reports, physical evidence, statements by other witnesses, and statements by the alleged perpetrator, the researchers clas-
sified each case as plausible or implausible. Significantly higher CBCA scores resulted for the plausible than the implausible cases. However, the differences in CBCA scores between the plausible and the implausible cases did not approach the high level reported by Raskin (see, e.g., Esplin et al., 1988), and the two groups could only be differentiated on 6 of the 14 CBCA criteria used. Lamb and his colleagues concluded that the precision of the CBCA is currently too poor to permit forensic application.

Similar results followed from a study by Ruby and Brigham (1998) designed to compare the effectiveness of the CBCA specifically for African American speakers. In this study, 6 African American and 6 White university students or staff members presented descriptions of two personal events, one true and one false. The speakers were given 1 or 2 days to prepare their presentations. They were instructed to describe incidents that they considered negative, emotional, and characterized by a loss of control. They were also instructed to be deceptively convincing when relaying the false story. Next, 143 White raters, trained to use the CBCA, then rated transcripts of the presentations. The raters were blind as to the ethnicity of the presenters. The results were that the raters were worse than chance at discriminating true from false accounts by White presenters; for African American presenters, there was no difference in the CBCA ratings for true and false events. In a discriminate analysis performed for the separate CBCA criteria, none of the criteria showed a significant positive correlation with truthfulness for both White and African American presenters. This study raises serious concerns about the utility of the CBCA for assessing the veracity of statements by African American and White individuals.

The results of a study by Steller, Wellershau, and Wolf (1988) raise an interesting question about the validity of the SVA for discriminating between accounts of true versus false events. In this study, children were each asked to tell a story about a true event and one about an event that never really happened to them. The findings were that the SVA was more successful at discriminating accounts of true events from fabricated accounts of unfamiliar events (e.g., medical procedures such as having an operation) than from fabricated accounts of familiar events. If people generate accounts of true and false events on the basis of the schemarelevant information retained in their memory, and if people have more information in their schemata for familiar than unfamiliar events, it is not surprising that accounts of true and false familiar events would be more similar and thus less discriminable than accounts of true and false unfamiliar events. These results relate to the reports of Pezdek et al. (1997) and Pezdek and Hodge (1999) that it is easier to plant false memories for plau-
sible than for implausible events. The interpretation of this result is that it is easier to plant a memory for an episode of an event about which one has a well-developed schema than for an event about which one does not have much schema-relevant knowledge.

The results of Steller et al. (1988) suggest that in real sexual abuse cases, the SVA may be effective at discriminating between true and fabricated accounts only for people without prior knowledge or experience with sexual acts. A similar concern has been expressed by Raskin and Esplin (1991): "A major limitation of CBCA is the difficulty in applying it to situations where the witness has other sources of information from which to invent an accusation that incorporates some or many of the content criteria" (p. 280). Similarly, Lamb et al. (1997) mentioned that the accounts of implausible cases that were most likely to be misclassified as plausible were those in which the central event—for example, an allegation of vaginal penetration—was described in the context of a familiar and more recent interaction with the alleged perpetrator. It may be possible that the SVA does not, in fact, discriminate between true versus false events; rather it discriminates between accounts of familiar versus unfamiliar events. After all, narrative accounts of familiar events would be expected to have a better logical structure and more detailed content than would the narrative accounts of unfamiliar events, and the familiarity of an event can come from direct personal experience or from indirect sources such as conversations with people and various forms of media.

REALITY MONITORING RESEARCH

Reality monitoring refers to the cognitive processes that are used to discriminate between perceived and imagined events (Johnson & Raye, 1981). Using a wide range of tasks, a number of researchers have demonstrated that memories for perceived events are phenomenologically different from memories for imagined events. This includes, for example, memories for perceived and imagined words (Foley, Johnson, & Raye, 1983), pictured objects (Johnson, Raye, Wang, & Taylor, 1979), actions (Foley & Johnson, 1985), and autobiographical events (Johnson et al., 1988).

The essential differences in the phenomenological characteristics of memories for perceived and imagined events were as follows: Whereas memories for perceived events contain more perceptual information (color and sound) and more contextual information (time and place), memories for imagined events contain more information about the cognitive processes that produced the memories. This concept is explained by
the fact that memories for perceived events are retrieved from memory more automatically and thus involve fewer search and decision processes. These search and decision processes that can later serve as reality monitoring cues are thus less abundant for perceived than for imagined events.

In one of the most ecologically relevant studies on reality monitoring, Johnson et al. (1988) had participants think about three autobiographical events (a social occasion, a visit to a library, a trip to the dentist) and three imagined events (a dream, a fantasy, an unfulfilled intention). They then rated their own memories using the Memory Characteristics Questionnaire (MCQ) developed by Johnson et al. (1988). The MCQ, a copy of which is presented in Table 4.2, includes 39 questions that require the participant to rate the quality of a specific memory in terms of various aspects of the dimensions reported to distinguish between perceived and imagined events. Generally, the findings supported the idea that memories of perceived autobiographical events contained more contextual and perceptual information, and memories of imagined autobiographical events contained more information regarding the cognitive processes involved in activating the memory (e.g., "I remember that I did this because..."); "I know I didn't just imagine this because...").

Johnson et al. (1988) also compared MCQ ratings of perceived and imagined events for events that occurred recently or in childhood. Interestingly, although the MCQ ratings for perceived and imagined events were in the same direction for recent and childhood events, most of the differences between perceived and imagined events that were significant for recent events were not significant for childhood events. This finding suggests that perhaps the differences between the phenomenological characteristics of perceived and imagined events dissipate with time.

The MCQ ratings were developed to assess self-reported differences in the nature of memories for perceived and imagined events. Several investigators have tested whether the MCQ ratings can also be useful in discriminating between perceived and imagined events when applied by experimenters to judge the veracity of participants' memory accounts. Schooler, Gerhard, and Loftus (1986, Experiments 4 and 5) tested if judges trained to identify the reality monitoring criteria that discriminate between perceived and imagined events could discriminate between descriptions of perceived and suggested objects. The descriptions were obtained for objects perceived or suggested using the misinformation paradigm of Loftus, Miller, and Burns (1978). In both experiments, trained judges were significantly more accurate than untrained judges in classifying descriptions as originating from real or suggested events, but the difference in classification accuracy was significant only for suggested items and not for
# Table 4.2
Memory Characteristics Questionnaire

<table>
<thead>
<tr>
<th>My memory for this event:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. is 1 = dim; 7 = sharp/clear.</td>
</tr>
<tr>
<td>2. is 1 = black and white; 7 = entirely color.</td>
</tr>
<tr>
<td>3. involves visual detail 1 = little or none; 7 = a lot.</td>
</tr>
<tr>
<td>4. involves sound 1 = little or none; 7 = a lot.</td>
</tr>
<tr>
<td>5. involves smell 1 = little or none; 7 = a lot.</td>
</tr>
<tr>
<td>6. involves touch 1 = little or none; 7 = a lot.</td>
</tr>
<tr>
<td>7. involves taste 1 = little or none; 7 = a lot.</td>
</tr>
<tr>
<td>8. Overall vividness is 1 = vague; 7 = very vivid.</td>
</tr>
<tr>
<td>9. My memory for the event is 1 = sketchy; 7 = very detailed.</td>
</tr>
<tr>
<td>10. Order of events is 1 = confusing; 7 = comprehensible.</td>
</tr>
<tr>
<td>11. Story line is 1 = simple; 7 = complex.</td>
</tr>
<tr>
<td>12. Story line is 1 = bizarre; 7 = realistic.</td>
</tr>
<tr>
<td>13. My memory for the location where the event takes place is 1 = vague; 7 = clear/distinct.</td>
</tr>
<tr>
<td>14. General setting is 1 = unfamiliar; 7 = familiar.</td>
</tr>
<tr>
<td>15. Relative spatial arrangement of objects in my memory for the event is 1 = vague; 7 = clear/distinct.</td>
</tr>
<tr>
<td>16. Relative spatial arrangement of people in my memory for the event is 1 = vague; 7 = clear/distinct.</td>
</tr>
<tr>
<td>17. My memory for the time when the event takes place is 1 = vague; 7 = clear/distinct.</td>
</tr>
<tr>
<td>18. For the year is 1 = vague; 7 = clear/distinct.</td>
</tr>
<tr>
<td>19. For the season is 1 = vague; 7 = clear/distinct.</td>
</tr>
<tr>
<td>20. For the day is 1 = vague; 7 = clear/distinct.</td>
</tr>
<tr>
<td>21. For the hour is 1 = vague; 7 = clear/distinct.</td>
</tr>
<tr>
<td>22. The event seems 1 = short; 7 = long.</td>
</tr>
<tr>
<td>23. The overall tone of the memory is 1 = negative; 7 = positive.</td>
</tr>
<tr>
<td>24. In this event I was 1 = a spectator; 7 = a participant.</td>
</tr>
<tr>
<td>25. At the time, the event seemed like it would have serious implications: 1 = not at all; 7 = definitely.</td>
</tr>
<tr>
<td>26. Looking back, this event did have serious implications: 1 = not at all; 7 = definitely.</td>
</tr>
<tr>
<td>27. I remember how I felt at the time when the event took place: 1 = not at all; 7 = definitely.</td>
</tr>
<tr>
<td>28. Feelings at the time were 1 = negative; 7 = positive.</td>
</tr>
<tr>
<td>29. Were 1 = not intense; 7 = very intense.</td>
</tr>
<tr>
<td>30. As I am remembering now, my feelings are 1 = not intense; 7 = very intense.</td>
</tr>
<tr>
<td>31. I remember what I thought at the time: 1 = not at all; 7 = clearly.</td>
</tr>
<tr>
<td>32. This memory reveals or says about me: 1 = not much; 7 = a lot.</td>
</tr>
<tr>
<td>33. Overall, I remember this event: 1 = hardly; 7 = very well.</td>
</tr>
<tr>
<td>34. I remember events relating to this memory that took place in advance of the event: 1 = not at all; 7 = yes, clearly.</td>
</tr>
<tr>
<td>35. After the event: 1 = not at all; 7 = yes, clearly.</td>
</tr>
<tr>
<td>36. Do you have any doubts about the accuracy of your memory for this event? 1 = a great deal of doubt; 7 = no doubt whatsoever.</td>
</tr>
<tr>
<td>37. Since it happened, I have thought about this event: 1 = not at all; 7 = many times.</td>
</tr>
<tr>
<td>38. Talked about it: 1 = not at all; 7 = many times.</td>
</tr>
<tr>
<td>39. About when did this event happen? Circle one: just today yesterday few days ago last week few weeks ago last month few months ago last year longer (if childhood, indicate age).</td>
</tr>
</tbody>
</table>

perceived items. Also, in both experiments, although the differences between trained and untrained judges on suggested items was significant, the difference was not a big one (20% improvement in Experiment 4 and 13% improvement in Experiment 5).

In a test of the extension of Johnson and Raye’s (1981) model to children, Alonso-Quecuty, Hernandez-Fernaud, and Campos (1994) had 9-year-old children either view a film of a simple staged event (the perceived condition) or listen to a description of the film (the imagined condition). Experimenters then rated their accounts of these events. The accounts of the children in the perceived condition contained more sensory and contextual details than did the accounts of the children in the imagined condition. It is not clear, however, whether listening to a description of a film would qualify as an imagined event according to Johnson and Raye’s framework. Perhaps the children in this condition simply remembered the verbal narration presented and did not actually generate the type of elaborative information that would be expected in Johnson and Raye’s (1981) imagination conditions.

More recently, Roberts, Lamb, Zale, and Randall (1998) modified Johnson et al.’s (1988) MCQ so that the scale could be used by experimenters rating children’s accounts of sexual abuse. A subset of the forensic interviews used by Raskin and Esplin (1991) was rated according to the MCQ criteria. These interviews were with children who had made reports of sexual abuse. The cases were independently classified as confirmed (N = 10) or unconfirmed (N = 16) on the basis of, for example, the availability of confessions, medical evidence, and physical evidence.

Overall, the total presence scores were significantly higher for the confirmed cases than for the unconfirmed cases. Separate t tests were conducted to compare scores for confirmed and unconfirmed cases on each of the 12 criteria of the MCQ. Significantly (p < .05) or marginally significantly (p < .09) higher scores for the confirmed cases occurred on ratings of perceptual features of people, actions, spatial information, temporal information, and self-reference. It appears, however, that the discriminating items on the MCQ were those that specifically predicted higher scores for perceived than for imagined events. Contrary to expectations, the unconfirmed cases did not receive higher self-reference scores nor higher thought and cognition scores than the confirmed cases; in fact, the trend was in the opposite direction. These findings are consistent with pilot results reported by Roberts, Lamb, and Randall (1997) that unconfirmed cases did not produce higher ratings of thoughts, rehearsals, and doubts than confirmed cases.

To date, findings regarding the effectiveness of the MCQ for discriminating between confirmed and unconfirmed cases of sexual abuse have
not been impressive. The results of Roberts et al. (1998) and Roberts et al. (1997) are suggestive rather than conclusive because of the relatively small sample sizes in these studies. However, these results suggest that perhaps the MCQ is only effective at discriminating confirmed from unconfirmed cases of sexual abuse because the total amount of detail elicited by the MCQ differs between these two groups. If this is true, then there should be a significant correlation between the MCQ scores and the CBCA scores. Roberts et al. (1997) confirmed this prediction; the correlation between the two scales was \( r = .76 \) (\( p < .03 \)). Additional research is clearly necessary to assess the utility of the MCQ for discriminating between true and false accounts of childhood sexual abuse.

MEMORIES FOR TRUE EVENTS VERSUS SUGGESTIVELY PLANTED FALSE EVENTS

There is a wealth of information available from which comparisons can be made of phenomenological characteristics of memories for true versus suggestively planted false events. This information has been published in the literature on planting false events in memory, in the suggestibility literature in which accounts of perceived versus suggested events have been compared, as well as the reality monitoring literature in which accounts of perceived versus imagined events have been compared. From the published research literature, we have selected the seven such studies that report recall data from which qualitative comparisons of accounts of true and false events have been made. The purpose of this section of the chapter is to summarize these seven studies and determine if their results suggest a consistent pattern of phenomenological characteristics that differentiate between accounts of true versus false events.

The first four studies are from the research literature on planting false events in memory. These include studies by Pezdek et al. (1997), Pezdek and Hodge (1999), Hyman and Pentland (1996), and Loftus and Pickrell (1995). The next two studies are from the suggestibility research literature. These include studies by Schooler, Clark, and Loftus (1988), and Schooler et al. (1988). The last study, by Johnson et al. (1988), is from the reality monitoring literature.

Pezdek et al. (1997) tested and confirmed the hypothesis that plausible false events are more easily planted in memory than implausible false events. In Experiment 1 of this study, 22 Jewish and 29 Catholic high school students were read descriptions of three true events (from mothers’ reports) and two false events reported to have occurred when they
were 8 years old. One false event described a Jewish ritual, and one described a Catholic ritual. After being read each description, each participant was asked what he or she remembered about the event. The results were that whereas 6 Catholics but 0 Jews remembered only the Catholic false event, 3 Jews but only 1 Catholic remembered only the Jewish false event. One person remembered both false events.

As part of this study, the researchers also compared various characteristics of the memories for the 13 participants who recalled at least one false event. In particular, it was predicted that compared to recalled false events, recalled true events would employ more words and receive higher ratings of clarity and confidence ("How confident are you that you could remember more about this event if given more time to think about it?"). These predictions were derived from a specific process model that describes how false events come to be planted in memory. According to this model, when a person is presented with a description of a specific instance of an activity and asked to verify whether this event happened to him or her, the person compares the instance described with his or her memory for related instances of that activity to determine whether there is a match. The more overlap there is between the information in the description and the information in memory, the more likely it is that a match will be indicated and the described event will be reported as true.

Once a false event is judged to be true, it is predicted that schema-relevant information in memory is activated and linked to the memory for the suggested false event. Memory for the false event thus includes information that was suggestively planted along with activated schema-relevant information. Furthermore, we predicted that when one retrieves a memory to determine whether there is a match between the memory and the described event presented, the schema-relevant information that was activated is less strongly associated with the constructed memory and is thus less clear and less confidently held than the information in memory that directly matched the presented experience. Consequently, because false events are likely to contain a higher proportion of activated schema-relevant information than are true events, memories for false events are predicted to be less clear and less confidently held than memories for true events.

In terms of the verbosity of the memory, memories for true events that are remembered contain the information presented in the described event as well as episodic memory for the event and activated information from schema-relevant knowledge in memory. On the other hand, according to this model, memories for false events contain only the information presented in the description and activated information from schema-
relevant knowledge in the memory. Because the false events did not actually occur, there is no episodic memory for the event. Thus, it was predicted that descriptions of false events would contain fewer words and fewer idea units than would descriptions of true events.

The results confirmed our hypotheses. Summarized in Table 4.3 are the findings from the seven studies to be reviewed in this section of the chapter. The findings are presented separately for the three phenomenological characteristics reported—confidence, verbosity, and clarity. As can be seen in Table 4.3, Pezdek et al. (1997) found that recall for false events \( (M = 15.64 \text{ words, } 3.25 \text{ idea units}) \) contained less information than recall for true events \( (M = 29.67 \text{ words, } 5.57 \text{ idea units}) \), and participants were less confident with false events \( (M = 4.79, \text{ on a } 1 \text{ to } 10 \text{ scale}) \) than with true events \( (M = 6.80) \) that they could recall more information if given more time. In addition, memories for recalled false events \( (M = 4.00, \text{ on a } 1 \text{ to } 10 \text{ scale}) \) were rated as less clear on average than were memories for recalled true events \( (M = 6.90) \).

In a similar study with children, Pezdek and Hodge (1999) read children descriptions of two true events and two false events, reported to have occurred when they were four years old. The children, ages 5 to 7 or 9 to 12 at the time of the study, were told what their mother remembered about each event. They were then asked what they remembered about the event. One false event described the child lost in a mall while shopping (the plausible false event); the other false event described the child receiving a rectal enema (the implausible false event). The major results were that whereas 14 children recalled the plausible but not the implausible false event, only 1 child recalled the implausible but not the plausible false event. Three additional children recalled both false events.

As part of this study, the researchers also assessed the verbosity of memories for true and false events. Consistent with the results of Pezdek et al. (1997), Pezdek and Hodge (1999) reported that of the 18 children who remembered at least one false event, significantly more idea units were recalled for true events \( (M = 4.86) \) than for false events \( (M = 3.06) \). Furthermore, of the 17 children who recalled the plausible false event, significantly more idea units were recalled for the true event \( (M = 5.06) \) than for the plausible false event \( (M = 3.24) \). Together these results suggest that the true events could be distinguished from the false events, even the plausible false event, by the number of details recalled.

In a study by Hyman and Pentland (1996), college-age participants were read descriptions of several true events (based on parent reports) and one false event reported to have occurred when the participant was a child. The false event described an event in which the participants attended a
wedding and spilled a punch bowl on the parents of the bride. Participants were asked what they remembered about each event. If they could not recall an event they were asked either to form a mental image of the event and describe the image to the experimenter (the guided imagery condition), or they were simply asked to quietly think about the event (the control condition). This procedure was repeated on each of 3 consecutive days. The major result was that the guided imagery condition produced more confirmed false memories than did the control condition.

Participants in this study were also asked to rate the clarity of their mental image on a 1 (low) to 7 (high) scale, their confidence in the memory (on a 1 to 7 scale), the emotionality of the memory (on a 1 to 7 scale), and the positive–negative valence of the memory on a 1 (negative) to 7 (positive) scale. Although Hyman and Pentland (1996) did not specifically compare phenomenological characteristics of memories for true and false events, the relevant data can be abstracted from their study for this purpose, although significance tests on these comparisons are not available. In each case, comparisons are reported for the participants who always remembered the true events and those who reported that they did remember the false event.

In the control condition, true memories ($M = 5.35$) were rated as more confidently held than false memories ($M = 4.00$). The rated clarity was higher for true memories ($M = 4.87$) than for false memories ($M = 2.75$). Emotionality was rated higher for true memories ($M = 3.79$) than for false memories ($M = 1.75$), and true memories were rated as more positive in emotional tone ($M = 4.02$) than were false memories ($M = 3.25$). A similar pattern was reported in the imagery condition. True memories ($M = 5.57$) were rated as more confidently held than false memories ($M = 3.00$). The rated clarity was higher for true memories ($M = 5.00$) than for false memories ($M = 3.25$). The emotionality of true memories was rated higher for true ($M = 3.90$) than for false memories ($M = 2.17$), and true memories were rated as more positive in emotional tone ($M = 4.37$) than were false memories ($M = 3.58$). These results suggest that there is more information and more elaborated information in memory for true events than exist in memory for confirmed false events. True events are perceptually clearer and more confidently held, and they contain more positive information and more emotional information than do false events.

The fourth study from which phenomenological comparisons of true and false events can be compared is that of Loftus and Pickrell (1995). In this study, 24 volunteers suggested to an offspring or younger sibling that they had been lost in a shopping mall when they were about 5 years old. This false event was presented along with three true events. Each partici-
<table>
<thead>
<tr>
<th>Study</th>
<th>Measure</th>
<th>Perceived or Watched</th>
<th>Suggested or Imagined</th>
<th>Statistic</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence</td>
<td>Mean confidence for whether Ss could eventually report more details on 1 (low) to 10 (high) scale</td>
<td>6.80</td>
<td>4.79</td>
<td>t(13) = 3.50</td>
<td>**</td>
</tr>
<tr>
<td>Pezdek et al. (1997)</td>
<td>Mean confidence rating on 1 (low) to 3 (high) scale</td>
<td>2.84</td>
<td>2.57</td>
<td>t(87) = 2.33</td>
<td>*</td>
</tr>
<tr>
<td>Schooler et al. (1986, Exp. 1)</td>
<td>DNR</td>
<td>DNR</td>
<td>DNR</td>
<td>t(12) = .75</td>
<td>ns</td>
</tr>
<tr>
<td>Schooler et al. (1988)</td>
<td>Mean confidence rating on a 1 (low) to 5 (high) scale</td>
<td>DNR</td>
<td>DNR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loftus &amp; Pickrell (1995)</td>
<td>Mean certainty rating for recent memories on a 1 (low) to 7 (high) scale</td>
<td>5.67</td>
<td>5.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson et al. (1988, Exp. 1)</td>
<td>Mean confidence rating on 1 (not) to 7 (very) scale in the control condition</td>
<td>5.35</td>
<td>4.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyman &amp; Pentland (1996)</td>
<td>Mean rating of confidence in the memory from 1 (not) to 7 (very) imagery condition</td>
<td>5.57</td>
<td>3.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbosity</td>
<td>Mean number of words</td>
<td>29.67</td>
<td>15.64</td>
<td>t(13) = 4.65</td>
<td>**</td>
</tr>
<tr>
<td>Pezdek et al. (1997)</td>
<td>Mean number of idea units</td>
<td>6.67</td>
<td>3.25</td>
<td>t(13) = 4.63</td>
<td>**</td>
</tr>
<tr>
<td>Pezdek &amp; Hodge (1999)</td>
<td>Of the 18 children who remembered one false event, mean number of idea units</td>
<td>4.86</td>
<td>3.06</td>
<td>t(17) = 2.36</td>
<td>*</td>
</tr>
<tr>
<td>Schooler et al. (1986, Exp. 1)</td>
<td>Of the 17 children who recalled the plausible false event, mean number of idea units</td>
<td>5.06</td>
<td>3.24</td>
<td>t(16) = 2.26</td>
<td>*</td>
</tr>
<tr>
<td>Schooler et al. (1986, Exp. 2)</td>
<td>Mean number of words</td>
<td>18.34</td>
<td>25.14</td>
<td>t(87) = 2.32</td>
<td>*</td>
</tr>
<tr>
<td>Schooler et al. (1986, Exp. 2)</td>
<td>Mean number of words for real condition and both suggested conditions</td>
<td>7.04</td>
<td>12.69</td>
<td>F(2,113) = 12.85</td>
<td>**</td>
</tr>
<tr>
<td>Study</td>
<td>Measure</td>
<td>Perceived or Watched</td>
<td>Suggested or Imagined</td>
<td>Statistic</td>
<td>Significance</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>-----------------------</td>
<td>-----------</td>
<td>--------------</td>
</tr>
<tr>
<td>Schooler et al. (1988)</td>
<td>Mean number of words</td>
<td>22.12</td>
<td>23.74</td>
<td>t(12) = .50</td>
<td>ns</td>
</tr>
<tr>
<td>Loftus &amp; Pickrell (1989)</td>
<td>Mean word length per description</td>
<td>138</td>
<td>49.9</td>
<td>DNR</td>
<td>*</td>
</tr>
<tr>
<td>Johnson et al. (1988, Exp. 2)</td>
<td>Mean number of idea units for recent events</td>
<td>2.02</td>
<td>1.53</td>
<td>DNR</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Mean number of idea units for childhood events</td>
<td>1.98</td>
<td>1.49</td>
<td>DNR</td>
<td>ns</td>
</tr>
<tr>
<td>Sensory/Clarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pezdek et al. (1997)</td>
<td>Mean clarity rating on 1 (low) to 10 (high) scale</td>
<td>6.90</td>
<td>4.00</td>
<td>t(13) = 4.69</td>
<td>**</td>
</tr>
<tr>
<td>Schooler et al. (1986, Exp. 1)</td>
<td>Percent of reports containing sensory information</td>
<td>41%</td>
<td>19%</td>
<td>z = 1.76</td>
<td>*</td>
</tr>
<tr>
<td>Schooler et al. (1986, Exp. 2)</td>
<td>Percent of reports containing sensory details collapsed</td>
<td>82.5%</td>
<td>55%</td>
<td>z = 2.98</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>across suggested conditions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schooler et al. (1988)</td>
<td>Mean frequency of sensory information included in reports</td>
<td>2.45</td>
<td>1.32</td>
<td>t(12) = 2.96</td>
<td>*</td>
</tr>
<tr>
<td>Loftus &amp; Pickrell (1995)</td>
<td>Mean clarity ratings for the 5 people who recalled the false event. 1 (low) to 10 (high) scale</td>
<td>6.3</td>
<td>3.2</td>
<td>DNR</td>
<td></td>
</tr>
<tr>
<td>Hyman &amp; Pentland (1996)</td>
<td>Mean clarity rating on 1 (not) to 7 (perfectly clear) scale for the control condition</td>
<td>4.87</td>
<td>2.75</td>
<td>DNR</td>
<td>DNR</td>
</tr>
<tr>
<td></td>
<td>Mean clarity rating on 1 (not) to 7 (perfectly clear) scale for the imagery condition</td>
<td>5.00</td>
<td>3.25</td>
<td>DNR</td>
<td>DNR</td>
</tr>
<tr>
<td>Johnson et al. (1988, Exp. 1)</td>
<td>Mean rating on 1 (vague) to 7 (clear) scale for vividness of childhood memories</td>
<td>4.97</td>
<td>4.78</td>
<td>DNR</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Mean rating on 1 (vague) to 7 (clear) scale for recent memories</td>
<td>5.69</td>
<td>5.06</td>
<td>DNR</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Mean rating on 1 (vague) to 7 (clear) scale for visual detail of recent memories</td>
<td>6.00</td>
<td>5.11</td>
<td>DNR</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Mean rating on 1 (vague) to 7 (clear) scale for visual detail of childhood memories</td>
<td>5.22</td>
<td>4.56</td>
<td>DNR</td>
<td>ns</td>
</tr>
</tbody>
</table>

*Note. Exp. = experiment; DNR = Data Not Reported; ns = not significant.  
*P = .05. **P = .01.
pant was asked what he or she remembered about each event. Six of the 24 participants reported either full or partial memory for the false event. The experimenters then examined the verbosity of the recalled descriptions of the true events and the remembered false events. Recalled true events contained more words (M = 138) than did recalled false events (M = 50). Also, mean clarity ratings in the second interview were higher for true events (M = 6.3) than for recalled false events (M = 3.6), and mean confidence ratings in the second interview were higher for true events (M = 2.2) than for recalled false events (M = 1.4). At the end of this study, participants were told that one of the events described had not really occurred. They were asked to pick which one this was, and 19 of the 24 subjects identified the false event correctly.

Schooler has conducted several experiments using the misinformation effect paradigm to compare phenomenological characteristics of memories for perceived and suggested events. Schooler et al. (1986, Experiments 1 and 2) had subjects view a sequence of slides. In the experimental condition, the sequence included a slide with a yield sign in it. In the suggested condition, the sequence did not include this critical target item. In a subsequent phase, a yield sign was suggested in the suggested condition but not in the experimental condition. In the experimental condition, 76% of the subjects reported having seen a yield sign (true), compared with 25% of the subjects in the suggested condition (false). To test the generalizability of the results, Experiment 2 was conducted with a stop sign as the critical item; similar findings were reported.

Phenomenological characteristics of the memories for perceived and suggested events were then examined. Compared with descriptions of perceived events, descriptions of suggested events less frequently included sensory properties of the sign (19% vs. 41%) but more often included functional information about the purpose of the sign (14% vs. 1%), first-person verbal hedges such as “I think” or “I believe” (38% vs. 12%), and descriptions of the cognitive processes associated with remembering or retrieving the event (29% vs. 3%). Similar results were obtained in Experiment 2, as well as in a separate study by Schooler et al. (1988) in which subjects were videotaped as they described the target items. The results of these three experiments are consistent with predictions of the reality monitoring model.

On the other hand, inconsistent with the results of the studies described above, in the Schooler et al. (1986) study, subjects who saw the yield sign used an average of 18.34 words to describe it, compared with 25.14 words for subjects for whom the yield sign was suggested. This finding was replicated in Experiment 2 of this study; the mean number of words used by
4. Accounts of True and False Events

subjects to describe the suggested stop sign was 12.69 (11.23 words in another suggested condition), compared with 7.04 words used by participants who saw the stop sign. However, in the study by Schooler et al. (1988), a range of objects less stereotypic than traffic signs (a robe, a shovel, a picture frame, a mailbox, etc.) were included in target slides, and no difference resulted between the verbosity of descriptions of perceived ($M = 22.12$ words) and suggested ($M = 23.74$ words) items. These findings are discrepant with the results of the other studies addressed in this section of the chapter, as well as with what might be predicted from the CBCA regarding the verbosity of true and false events.

One interpretation of the different results obtained by Schooler as compared with other researchers regarding the verbosity of descriptions of perceived and suggested items is that unlike the researchers in the other studies reported here, Schooler and his colleagues included in their word count all words provided. This included the words in statements such as, “I think it must have been there because . . . ,” and “When I think about it, it seems to me that . . .” Because participants made more self-referencing statements in describing suggested items than perceived items, this would have inflated the word count for suggested items.

Another interpretation of why Schooler obtained higher verbosity ratings for suggested than perceived items is that perhaps the difference relates to whether participants were describing a simple object or a complex event. With a simple object, the verbal label captures a great deal of the information about the object (i.e., “stop sign,” “yield sign”). If a person has seen a simple object, elaborate descriptive information is redundant with the simple verbal label. Thus, people who have seen a simple object are not likely to provide a great deal of descriptive information about the object. On the other hand, complex events such as getting lost in a mall and spilling punch at a wedding are not easily verbally labeled, and there are very few constraints on the range of elaborative details that might describe such events. It does appear, however, that with complex events, memories of perceived events will be described with significantly more words than will memories of suggestively planted events.

The seventh study to be reported in this section is the reality monitoring study by Johnson et al. (1988). In this study, people’s reality monitoring ability for perceived and imagined autobiographical events was compared. Although this study was discussed previously in this chapter, some of the findings will also be reported here for comparison with the above studies on memory for perceived and suggested events. The findings were that perceived events were described using more idea units ($M = 2.02$ for recent events and $1.98$ for childhood events) than were imagined events ($M = 1.53$
for recent events and 1.49 for childhood events). Perceived events ($M = 6.00$ for recent events and $5.22$ for childhood events) were also rated higher in clarity of visual detail on a scale from 1 (vague) to 7 (clear) than imagined events ($M = 5.11$ for recent events and $4.56$ for childhood events), although this difference was significant only for recent events. Differences in ratings of vividness of memories and confidence in the veracity of the memory did not significantly differ between the perceived and the imagined events in the Johnson et al. (1988) study (these means are presented in Table 4.2).

In summary, with the exception of the reports of verbosity in descriptions of perceived and suggested events by Schooler et al. (1986, 1988), these seven studies present a consistent pattern of results regarding phenomenological characteristics of memories for perceived and suggested or imagined events. Although people may erroneously report some false events as true, their descriptions of true events contain a greater clarity of perceptual details, they are more confidently held, and they are described with more words. These findings are consistent with the prediction that when one retrieves a memory to determine whether there is a match between the memory and the described event presented, schema-relevant information that is associated with the memory is less clear and less confidently held than information in the episodic memory that directly matches the presented experience. Furthermore, whereas memories for remembered true events contain the information presented in the described event as well as episodic memory for the event and schema-relevant information that is associated in memory, memories for false events contain only the information presented in the description and associated information from schema-relevant knowledge in memory.

**CONCLUSION**

Although accounts of memories for true and false events tend to differ in predictable and statistically significant ways, the tools available for determining the veracity of memory accounts are far less than perfect. Findings from the research on memory for true events versus suggestively planted false events dovetail nicely with findings from the reality monitoring research in suggesting that accounts of true events contain a greater clarity of perceptual details, they are more confidently held, and they are described with more words. This pattern is consistent for memory accounts of adults as well as children.

To the extent that the CBCA component of the SVA tool also assesses these qualities, it is not surprising that the CBCA facilitates the discrimina-
tion of accounts of true and false events. However, there is a need for additional research on the value added of the additional factors in the CBCA beyond the three general criteria. The relative importance of the 18 CBCA content criteria needs to be determined to provide a more precise interpretation of the score obtained in any case. Also, as we discussed earlier, it is not clear whether the organization and logical structure of a memory account are indicative of the veracity of the account, as is predicted by the CBCA, or whether these features simply correlate with the familiarity of the domain described. If accounts of familiar events received higher CBCA scores than accounts of unfamiliar events, this would be a serious limitation of the CBCA because the familiarity of an event can come from indirect sources such as conversations with people and various forms of media as well as from direct personal experience.

In the work reviewed in this chapter, the veracity of memory accounts has been determined by the relative quality of specific characteristics; the qualities of some event known to be false have been compared with the qualities of some event known to be true. This procedure has important applied implications. Social service agencies seek validity assessment tools to determine, for example, the veracity of reports of abuse in which the veracity of the alleged event is not known. The work reviewed in this chapter suggests that the veracity of an alleged event will be determined more accurately if blind raters can compare the qualities of the account of the reported event with the qualities of the account of some comparable event that is known to be true (a within-subject control condition). Because the qualities known to differentiate between accounts of true and false events also vary widely among individuals, we suggest that the predictive validity of any memory assessment tool would be enhanced if interviewers made within-subject comparisons of the account of the alleged event with the account of a comparable event known to be true.

REFERENCES


Ruby, C. L., & Brigham, J. C. (1997). The usefulness of the criteria-based content analysis
4. Accounts of True and False Events


