Law Hum Behav (2007) 31:463–478
DOI 10.1007/s10979-006-9081-5

ORIGINAL ARTICLE

Interviewing Witnesses: The Effect of Forced Confabulation on Event Memory

Kathy Pezdek · Kathryn Sperry · Shana M. Owens

Published online: 24 January 2007

Abstract After viewing a crime video, participants answered 16 answerable and 6 unanswerable questions. Those in the "voluntary guess" condition had a "don't know" response option; those in the "forced guess" condition did not. One week later the same questions were answered with a "don't know" option. In both experiments, information generated from forced confabulation was less likely remembered than information voluntarily self-generated. Further, when the same answer was given to an unanswerable question both times, the confidence expressed in the answer increased over time in both the forced and the voluntary guess conditions. Pressing eyewitnesses to answer questions, especially questions repeated thrice (Experiment 2), may not be an effective practice because it reliably increases intrusion errors but not correct recall.

Keywords Eyewitness memory · Forced confabulation · False memory · Eyewitness confidence · Event memory

Eyewitness evidence is critical for solving crimes, and it is often the sole source of evidence for determining what actually occurred during a crime. However, psychologists have long been familiar with the shortcomings of eyewitness memory, at least since the early work of Hugo Munsterberg (Munsterberg, 1908). One psychological factor known to affect the accuracy of eyewitness memory is the extent to which suggestive influences intervene between when an eyewitness observes an event and when the eyewitness testifies about the event. Numerous studies have demonstrated that eyewitnesses can be misled by post-event suggestions following an observed event (Loftus, 1975; Loftus, Miller, & Burns, 1978; Pezdek, 1977). This finding is called the misinformation effect. Post-event suggestion can also occur as a result of an intervening photographic lineup that is presented after an eyewitness has viewed an event, but prior to a subsequent lineup (Brigham & Cairns, 1988; Hinz & Pezdek, 2001; Pezdek & Blandon-Gitlin, 2005).

This study examined how eyewitnesses' memories for events are affected by a specific type of post-event suggestion, forced confabulation. This can occur in police interviews when police officers press an eyewitness to answer a question even though the eyewitness has indicated that

K. Pezdek (✉) · K. Sperry · S. M. Owens
Department of Psychology, Claremont Graduate University, Claremont, CA 91711-3955, USA
e-mail: Kathy.Pezdek@cgu.edu

he or she does not know or is unsure of the answer to the question. Although it is clear that providing witnesses with erroneous information during an interview suggestively misleads them to remember the erroneous information as having been observed in the target event (Loftus, 1975; Loftus et al., 1978; Pezdek, 1977), it is less clear whether *self-generated* answers to questions will also produce the misinformation effect. And, given that witnesses to crimes are typically questioned multiple times prior to testifying in court (Cutler & Penrod, 1995; Poole & White, 1995), it is important to examine whether self-generated answers are consistently remembered over time. Eyewitness confidence will also be assessed in light of findings that repeatedly questioning eyewitnesses has separate effects on accuracy and confidence (Shaw, 1996; Shaw & McClure, 1996).

A few studies have examined how forcing individuals to confabulate answers to questions about a target event affects memory for the target event. Hastie, Landsman, and Loftus (1978) had college students view a slide sequence. Some participants then answered a series of questions about the target event and were urged to guess answers to questions about which they were unsure. Control subjects participated in irrelevant distractor tasks. Ten minutes later all subjects completed a forced choice recognition memory test including questions about objects that either were or were not in the target event. Control subjects exhibited higher false alarm rates on the final test than those who answered questions. However, because this finding was stronger for answerable than for unanswerable questions, and because the authors did not examine whether subjects forced to guess tended to give the same wrong answer on the initial and subsequent recognition test, it is not clear whether these findings are a result of changes in memory or simply a criterion shift caused by guessing.

Ackil and Zaragoza (1998) had adults and children view a brief video, followed immediately by a sequence of answerable and unanswerable questions. Half of the subjects were forced to answer every question and were told to guess if they did not know an answer; control subjects were told to respond only to questions for which they knew the answer; they were encouraged not to guess. One week later, all subjects were asked whether they had seen various objects in the video. Subjects were more likely to misattribute an object to the video if they had self generated a response concerning that object in the forced response condition than if the object had been generated by another subject. Similar results were reported with children by Schreiber, Wentura, and Bilsky (2001) using a somewhat different method for inviting speculation. Zaragoza, Payment, Ackil, Drivdahl, and Beck (2001) further reported that confirmatory feedback by the experimenter ("That's right, 'knee' is the correct answer") enhanced these effects. In addition, Shapiro and Purdy (2005) recently reported that answers that resulted from forced confabulations were more likely to be subsequently recalled than information that was included in suggestive questions. Together, these results clearly suggest that forced confabulation increases memory suggestibility.

The interpretation of these effects is that when individuals are forced to guess answers to unanswerable questions, the answers that they generate become incorporated into their memory for the target event. On the subsequent test then, individuals are likely to make a source monitoring error and confuse the self-generated information with information actually observed. Source monitoring errors occur when individuals integrate new information about an event into their memory for the target event, but then do not remember that the new information did not come from the originally experienced target event (Johnson, Hashtroudi, & Lindsay, 1993; Lindsay, 1990; Lindsay & Johnson, 1989).

However, no previous studies were designed to assess the consistency of responses to the same unanswerable questions from time 1 to time 2 for guessed answers that were forced compared to guessed answers that were volunteered because individuals thought they had observed the relevant information. The comparison of answers to unanswerable questions that are forced guessed versus voluntarily guessed is important because it tests whether forced confabulated information that was never actually observed is as likely to persist in memory as information that individuals thought they had observed. In addition, in none of the previous studies on forced confabulation were participants asked the same questions on two occasions so that the effect of forced confabulation at time 1 could be examined on responses to the same questions at time 2. That is, the researchers were not able to examine whether subjects forced to guess gave *the same* wrong answers on the initial and subsequent tests. In terms of the forensic application of these findings, it is important to assess the consistency of responses to the same unanswerable questions at time 1 and time 2 because answers that are consistently given over time are more likely to be judged to be true by law enforcement personnel.

The primary issue addressed in the present study is the extent to which memory for a complex event is subsequently suggestively influenced by (a) guessed answers that were forced compared to (b) guessed answers that were volunteered, for answerable versus unanswerable questions. Subjects viewed a five-minute video of a crime, followed by 22 questions: (a) 16 answerable questions, and (b) 6 unanswerable questions (about information not in the video). Subjects in the "voluntary guess" condition (the control condition) answered all 22 questions with the option of responding "don't know" to each; subjects in the "forced guess" condition (the experimental condition) answered all questions without the "don't know" response option. One week later at time 2, all subjects answered the same 22 questions with the "don't know" response option provided. We call the control condition the "voluntary guess" condition recognizing that some degree of response guessing is always present even when individuals are not forced to do so. This is consistent with the Signal Detection Model of memory (cf. Banks, 1970) in which the distribution of memory strength varies along a continuum and the response is determined by where the response criterion is placed along this continuum.

In this paradigm, the memory upon which individuals base their answers is conceptualized as consisting of some information that was observed in the video (i.e., signal) and some related schematic information (i.e., noise). In answering answerable questions, subjects would base their answers on a distribution of information in memory that includes both signal and noise, and thus, would on average be of higher strength than the distribution of information in memory for unanswerable questions, that would consist only of noise. It is thus predicted that the consistency of the answers provided at time 1 and time 2 would be higher for answerable than for unanswerable questions.

Consistent with the previous research, it was also predicted that forced confabulation would contribute additional noise resulting from the suggestive planting of the guessed information in memory. However, because forcing individuals to guess answers to questions would have the effect of lowering their response criterion at time 1, the major prediction in this study is that given that an item was answered at time 1, the mean proportion of items that would be responded to with the same answer at time 2 would actually be lower in the forced guess condition than in the voluntary guess condition. As a corollary, we also predict that given a chance at time 2 to respond "don't know," individuals in the forced guess condition, who generated their forced guesses at time 1 from memories on average of weaker strength, would be more likely than those in the voluntary guess condition to respond "don't know" rather than staying with the answer they gave at time 1. Thus, it is predicted that the consistency of eyewitnesses' responses from time 1 to time 2 (a) would be higher for answerable than for unanswerable questions, and (b) would be greater when individuals are not pressed to answer questions at time 1.

No data exist documenting how frequently officers in real police interviews press an eyewitness to answer questions after the eyewitness indicates that he or she does not know the answer to the question. However, Gudjonsson (1992), Kassin (1997), and Leo (1996) have reported that this is not an unusual practice. They reported that eyewitnesses are often pressed to describe events that the interviewer believed occurred, even when witnesses report that they cannot remember the event or report that they did not even witness the event; false confessions commonly occur under these circumstances. The proposed research investigates whether eyewitness recall of a target event is affected by forcing the eyewitnesses to generate answers to questions probing information about details that they never observed in the target event. This study provides a cognitive processing account of the forced confabulation effect based on the Signal Detection Model of memory.

Experiment 1

Methods

Participants and design

One hundred thirteen participants volunteered to take part in four psychology classes at a State University and 2 psychology classes at a Community College in the Los Angeles metropolitan area. This was a 2 (forced guess versus voluntary guess condition) $\times$ 2 (time 1 versus time 2) mixed factorial design with the first variable manipulated between subjects.

Materials, procedure, and data coding

Subjects participated in two sessions lasting about 15 minutes each. In the first session, participants viewed a 5 min video of a car-jacking. This video was previously used by Clark and Tunnicliff (2001). Participants were told to pay very close attention to the video because afterward they would be asked some questions regarding what they had just watched. The video was followed immediately by 22 open-ended questions including (a) 16 answerable questions, and (b) 6 unanswerable questions. A list of these questions is included in the Appendix. Answerable questions probed information that was presented in the video; unanswerable questions probed information not actually presented in the video. For example, an unanswerable question is, “What was the logo on the perpetrator’s shirt?” and the perpetrator’s shirt did not have a logo. Questions were presented in the same order to all participants. The first three questions were answerable. The next 19 questions followed the chronological order in which the relevant information appeared in the video with no consecutive unanswerable questions.

A pilot study was conducted ($N = 89$) prior to Experiment 1 to determine which questions, from a larger set of answerable and unanswerable questions, were most appropriate for the study. In particular, we wanted to be sure (a) that participants could generate reasonable answers to the unanswerable questions, and (b) that the answerable questions were not so easy that participants could readily recognize the difference between answerable and unanswerable questions. The data coding system was also determined from the pilot data. In this study, each answer was coded as correct or incorrect. Two of the authors together reviewed all of the responses provided to each answerable question in the pilot study and then specified for each question, the responses to be classified as correct answers. For example, regarding question 11 in the Appendix, (“Just outside the building was an archway. What material was this archway made of?”), the acceptable correct answers were “concrete” or “cement” (because most people do not know the difference between concrete and cement). The most common incorrect answer to this question was “brick.” Specifying the correct answers in advance was also necessary because, for example, the colors were slightly different as a function of the LCD projector used (e.g., the victim’s backpack looked purple in some rooms and blue in others).

From the pilot data the experimenters also determined for each question what qualified as "the same answer" at time 1 and time 2 for that question. A lenient coding system was used. Synonyms were considered to be the "the same answer" (e.g., "circular planter" at time 1 and "round planter" at time 2). Responses at time 2 that included the answer from time 1 with an additional descriptor added were also considered to be "the same" response (e.g., "blue shorts" at time 1 and "blue cargo shorts" at time 2).

Two coders, working separately, each coded all of the data for each experiment. The coders for each experiment included one of the authors plus a graduate student who was not an author and was not familiar with the hypotheses tested in the study. Although coders were not blind as to which condition each subject was in while they coded the data, the two experimenters who generated the coding sheet from the pilot study were blind as to which responses were produced in which condition in the pilot study. It was in the pilot study that it was determined which responses would be classified as (a) correct and (b) "the same answer" at time 1 and time 2. In this way, coding biases were avoided. Coding the data simply involved comparing the response to each question for each subject with the correct responses on the coding sheet generated from the pilot study. The only discrepancies between the two coders were a few coding errors, and these were resolved by correcting any erroneous decisions against the coding sheet.

Participants in the forced guess condition ($N = 43$) were instructed to answer all 22 questions. They were told not to leave any questions unanswered, and that they should make their best guess of each answer even if they were unsure. *Participants in the forced guess condition were not given an "I don't know" response option at time 1*. On the other hand, participants in the voluntary guess condition ($N = 52$) were instructed to answer each question, however they were told that if they did not know an answer they should circle the "I don't know" response option that was available on their response sheet. After each of their 22 responses, participants in both conditions also rated their confidence in the accuracy of their response on a 1 (low) to 7 (high) scale. Participants did not give confidence ratings for questions to which they responded, "I don't know."

One week later at time 2, participants were retested on their memory for the target event in the video. At time 2, participants in both groups were provided with the same 22 questions about the target event, but this time *participants in both groups were given the "I don't know" response option* and were instructed to check this option if they did not know the answer to a question. Again, participants provided confidence ratings for each answer provided.

Results and discussion

There were two exclusion criteria. First, 14 participants were excluded from the analyses because they indicated that at least one unanswerable item was "not present" in the video; these included 7 in the voluntary guess and 7 in the forced guess condition. The second exclusion criterion was that participants in the forced guess condition were excluded from all analyses if they failed to answer at least 5 of the 6 unanswerable questions. Four participants were excluded based on this criterion. Analyses were performed on the remaining 95 subjects (M age $= 21.2$ years; 31 males, and 64 females).[1] Results for the two dependent variables—measures of memory and

[1]The data were also analyzed without the exclusion criteria in effect—that is with all 113 participants included in Experiment 1 and 97 participants included in Experiment 2—to see if the exclusion criteria altered the pattern of obtained results. For both experiments (a) the pattern of results was the same when all participants were included in the analyses, and (b) all reported statistical tests that were significant when the exclusion criteria were in effect, were also significant when the exclusion criteria were not in effect. However, there were two statistical tests in Experiment 1 that were not significant when the exclusion criteria were in effect but were significant when the

confidence—were assessed separately. Responses to answerable and unanswerable questions were analyzed separately.

In this study, we examine the pattern of responses at time 2 *given* that participants provided an answer to each question at time 1, in the forced guess versus voluntary guess conditions. There are two ways to achieve this *conditional probability*. In some previous studies (Zaragoza, Payment, Ackil, Drivdahl, & Beck, 2001; Ackil & Zaragoza, 1998) the experimenters made sure that everyone in the forced confabulation condition provided an answer to each question at time 1. Zaragoza et al. (2001), for example reported that, “When participants resisted answering these questions, the experimenter prompted them to provide their best guess (repeatedly, if necessary) until they eventually acquiesced” (p. 474). They could do this because each participant was interviewed separately. However, in our study individuals did not participate individually; they participated with other volunteers from their class. So although participants were instructed to provide a response to every question and not to indicate that the item was “not present” for any question, and the experimenters walked around the room and reminded participants of this when they observed blank spaces for responses, nonetheless, some participants turned in their protocol without answering all questions. Thus, to examine the pattern of responses at time 2 *given* that participants provided an answer to each question at time 1, we had to exclude from the analysis (a) participants who indicated that at least one unanswerable item was “not present” in the video and (b) those in the voluntary guess experimental condition who failed to answer at least 5 of the 6 unanswerable questions.

Analyses of responses to unanswerable questions

The primary analyses addressed the mean proportion of unanswerable questions that were responded to at time 2 with the same answer, a different answer, or a “don’t know” response given that the item was answered at time 1. In these analyses, a “don’t know” response at time 1 was not considered to be “an answer” but rather was classified as a separate response option and was not included in this analysis. The data are presented in Table 1 for participants in the voluntary guess condition and the forced guess condition. In the voluntary guess condition only 29 participants gave an answer to at least 1 unanswerable question (i.e., they did not respond, “don’t know”), compared to all 43 participants in the forced guess condition. Most of the responses to unanswerable questions in the voluntary guess condition were responses of “don’t know” both at time 1 and time 2 (74% of all responses).

If participants were forced to guess answers to unanswerable questions at time 1, were the answers they generated likely to be recalled one-week later at time 2? As can be seen in Table 1, in the forced guess condition, .40 of the responses at time 2 were the same answer that had been generated at time 1, even though the “don’t know” response option was now available. Although .40 is significantly greater than a response rate of 0, $z = 5.35, p < .001$, the difference between the proportion of responses that were the same at time 1 and time 2 ($p = .40$) and the proportion

exclusion criteria were not in effect. For these two comparisons, the direction of the effect matched that reported with the exclusion criteria in effect. First, when an incorrect response to an answerable question was recalled at time 1, the most likely response at time 2 was still that same incorrect response. This was true in both the voluntary guess ($M = .71$) and the forced guess conditions ($M = .61$), and these proportions were significantly different, $t\,(111) = 2.36, p < .05, d = .45$. Similarly, when an incorrect response to an answerable question was recalled at time 1, participants were more likely to change that response to “don’t know” in the forced guess condition ($M = .30$) than in the voluntary guess condition ($M = .20$), $t\,(111) = 2.11, p < .05, d = .40$. Although these latter two tests resulted in significant results without the exclusion criteria, the direction and magnitude of these effects were very similar with and without the exclusion criteria. Together, these results confirm that the exclusion criteria did not altered the pattern of obtained results in general, or the specific pattern of results across conditions.

Table 1 Mean proportion (and standard deviation) of unanswerable questions that were responded to at Time 2 with the same answer, a different answer, or a "Don't Know" response given that the question was answered at Time 1 in experiment 1

	Time 1	
	Voluntary guess	Forced guess
Time 2	($N = 29$)	($N = 43$)
Same answer	.54 (.44)[a]	.40 (.23)[a]
Different answer	.17 (.32)	.15 (.18)
Don't know	.29 (.43)[b]	.45 (.30)[b]

[a] $t\,(70) = 1.81, p < .05$, one-tailed.

[b] $t\,(70) = 1.87, p < .05$, one-tailed.

that changed to "don't know" at time 2 ($p = .45$) was not significant, $z < 1.00$. Thus, although forced confabulation does influence memory, the modal response to unanswerable questions at time 2 was "don't know" rather than providing the answer that was guessed at time 1.

How does the persistence of answers to unanswerable questions from time 1 to time 2 compare in the forced guess versus the voluntary guess conditions? We predicted that forcing individuals to guess answers to questions would have the effect of lowering their response criterion at time 1. Consequently, given that an item was answered at time 1, the mean proportion of items that would be responded to with the same answer at time 2 was actually predicted to be lower in the forced guess condition than in the voluntary guess condition. This prediction was tested with t-tests comparing each of the 3 pairs of responses in Table 1. First, as predicted, the mean proportion of responses that received the same answer at time 1 and time 2 was significantly higher in the voluntary guess condition ($M = .54$) than in the forced guess condition ($M = .40$), $t(70) = 1.81, p < .05, d = .43$. Second, when given a chance at time 2 to respond "don't know," individuals in the forced guess condition ($M = .45$) were more likely than those in the voluntary guess condition ($M = .29$) to respond "don't know" rather than giving the same answer that they gave at time 1, $t(70) = 1.87$, $p < .05$, $d = .45$. However, the proportion of responses that received a different response at time 1 and time 2 did not significantly differ between the forced guess ($M = .15$) and the voluntary guess ($M = .17$) conditions, $t(70) = .29, p = .77, d = .07$.

Analyses of responses to answerable questions

First, for answerable questions at time 1, although more correct responses were generated in the forced guess (.53) than the voluntary guess condition (.44), $t\,(93) = 2.91, p < .01, d = .60$, more incorrect responses were also generated in the forced guess (.47) than the voluntary guess condition (.25), $t\,(93) = 7.88$, $p < .001$, $d = 1.63$. These results suggest that the practice of pressing eyewitnesses to answer questions is likely to increase reports of both correct and incorrect information. In addition, the low overall accuracy rate to the answerable questions suggests that the open-ended test questions used in this study were not easy. Participants were not likely to recognize the difference between answerable and unanswerable questions and thus were not likely to feel tricked by being forced to answer unanswerable questions. In fact, the video was complex with a great deal of information presented.

The results for answerable questions are provided in Table 2. Because a correct answer was possible for answerable questions, Table 2 presents the mean proportion of answerable questions that were responded to at time 2 with a correct answer, an incorrect answer or a

Table 2 Mean proportion (and standard deviation) of answerable questions that were responded to at Time 2 with the correct answer, an incorrect answer, or a "Don't Know" response, given that the question was answered correctly or incorrectly at time 1 in experiment 1

	Time 1			
	Voluntary guess		Forced guess	
Time 2	Correct ($N = 52$)	Incorrect ($N = 51$)	Correct ($N = 43$)	Incorrect ($N = 43$)
Correct	.80 (.16)[a]	.09 (.14)	.72 (.21)[a]	.09 (.08)
Incorrect	.11 (.12)	.71 (.26)	.13 (.15)	.63 (.26)
Don't know	.10 (.13)	.20 (.28)	.15 (.19)	.29 (.25)

[a] $t(93) = 2.03, p = .045$, one-tailed.

"don't know" response given that the question was answered correctly or incorrectly at time 1. Of primary interest was whether these responses differed in the forced guess versus voluntary guess conditions. To compare each pair of responses, *t*-tests were conducted. As predicted, the mean proportion of responses that received the same correct answer at time 1 and time 2 was significantly higher in the voluntary guess condition ($M = .80$) than in the forced guess condition ($M = .72$), $t\,(93) = 2.03, p < .05, d = .42$.

What happened when an incorrect response was provided at time 1 to an answerable question? This accounted for 25% of the total responses in the voluntary guess condition and 47% of the total responses in the forced guess condition. As can be seen in Table 2, if an incorrect response was recalled at time 1, the most likely response at time 2 was that same incorrect response, and this was true in both the voluntary guess ($M = .71$) and the forced guess conditions ($M = .63$), and these proportions did not significantly differ, $t\,(92) = 1.58, p = .12, d = .33$.

Analyses of confidence ratings

The next set of analyses assessed the change in confidence ratings from time 1 to time 2 in the forced guess versus the voluntary guess conditions for responses to answerable and unanswerable questions. A 2 (forced guess versus voluntary guess condition) × 2 (time 1 versus time 2) × 2 (answerable versus unanswerable questions) mixed factorial Analysis of Variance was conducted on the confidence data (range in confidence ratings = 1–7). For this analysis we specifically looked at the answers that were consistently given by each subject at time 1 and time 2 to see if for these answers, the confidence ratings differed between individuals who were forced to guess at time 1 and those who were not. As in previous analyses, a "don't know" response at time 1 was not considered to be "an answer" but rather was classified as a separate response option and was not included in this analysis. This analysis includes responses to (a) answerable questions for which the same correct response was given at time 1 and time 2, and (b) unanswerable questions for which the same guessed response was given at time 1 and time 2. The data are provided in Table 3. Note that the sample sizes across the four conditions vary considerably.

First, as would be expected, answerable questions were responded to more confidently ($M = 5.09$) than were unanswerable questions ($M = 3.06$), $F(1{,}56) = 120.17, p < .001, \eta^2 = .68$. In addition, there was a significant interaction of time by question type, $F(1{,}56) = 24.02, p < .001, \eta^2 = .30$. As can be seen in Table 3, whereas confidence increased from time 1 ($M = 2.78$) to time 2 ($M = 3.35$) for unanswerable questions, $t\,(57) = 4.02, p < .001, d = 1.06$, confidence decreased from time 1 ($M = 5.24$) to time 2 ($M = 4.93$) for answerable questions, $t\,(92) = 4.55, p < .001, d = .95$. No other effects were significant. Most relevant to understanding forced confabulations, when participants gave the same answer to an unanswerable question at time 1 and one week later at time 2, their confidence in this answer increased over time, and

Table 3 Mean confidence ratings (and standard deviations) for responses to (a) Answerable questions for which the same correct response was given at Time 1 and Time 2, and (b) Unanswerable questions for which the same response was given at Time 1 and Time 2 by participants in the voluntary guess and forced guess conditions in experiment 1

	Answerable questions		Unanswerable questions	
	Voluntary guess ($N = 50$)	Forced guess ($N = 43$)	Voluntary guess ($N = 18$)	Forced guess ($N = 40$)
Time 1	5.02 (.89)	5.34 (.93)	2.92 (1.96)	2.71 (1.41)
Time 2	4.64 (1.19)	5.07 (1.00)	3.42 (1.96)	3.33 (1.28)

this occurred in both the voluntary guess and the forced guess conditions. Guessed answers to unanswerable questions in the forced guess condition were not less confidently held than guessed answers in the voluntary guess condition. These data exemplify another situation in which witness confidence is not a good indication of witness accuracy.

Experiment 2

Experiment 2 assessed the effect of repeated forced guessing on eyewitness memory and confidence. This is a forensically relevant variable because real eyewitnesses are frequently interviewed numerous times (cf., Cutler & Penrod, 1995; Poole & White, 1995), and in each interview, they may be pressed to answer questions for which they indicate that they do not know the answer. It is important to assess whether multiple opportunities to guess the answer to a question increases the likelihood that the self-generated information will be incorporated into memory for the target event and persist over time, especially with questions for which the witness never observed the relevant information in the target event.

In several studies it has been reported that repeated presentation of externally provided misinformation enhances its impact and increases the probability that the suggestive information will be planted in memory. For example, Zaragoza and Mitchell (1996) had subjects view a video followed by questions including misleading questions that were each presented one or three times. With repetition, subjects were more likely to misattribute suggested information to the video, and to do so with higher confidence. The interpretation of this finding is that with repetition, the strength of memory for the suggested information increases and thus, this information becomes more likely to be retrieved when recalling the video. Based on this finding, it is predicted in Experiment 2 that if eyewitnesses are interviewed multiple times about information that they in fact never saw and are encouraged to guess answers to specific questions each time, they are even more likely to remember their forced confabulated answers as information actually presented than if they did so only once.

Methods

Participants and design

Ninety-seven participants volunteered to take part in four psychology classes at a State University in the Los Angeles metropolitan area. This was a 2 (forced guess versus voluntary guess condition) × 2 (time 1 versus time 2) × 2 (once vs. thrice presentation condition) mixed factorial design with only the first variable manipulated between subjects. About half of the participants were assigned to the forced guess condition; the remaining participants were assigned to the voluntary guess condition.

Materials and procedure

The materials and procedure used in Experiment 1 were also used in Experiment 2, with the exception that in Experiment 2, at time 1, three of the six unanswerable questions were randomly selected and presented once each; the other three were presented three times each. Three randomly selected answerable questions were also assigned to the thrice-presented condition to avoid drawing attention to the repeated unanswerable questions. Four versions of test 1 were constructed so that the assignment of unanswerable questions to the once- versus twice-presented conditions and to the forced versus the voluntary guess conditions were counterbalanced across subjects. Each test version included a total of 34 questions. Test questions were arranged so that (a) the first two questions were always answerable, (b) each repeated question was separated by 10 items from its previous presentation, and (c) unanswerable questions were never presented consecutively.

Results and discussion

There were three exclusion criteria.[1] First, 7 participants in the forced guess condition and 3 participants in the voluntary guess condition were excluded from the analyses because they indicated that at least one unanswerable item was "not present" in the video. The second exclusion criterion was that participants in the forced guess condition were excluded from all analyses if they failed to give an answer to at least 5 of the 6 unanswerable questions. Three participants were excluded based on this criterion. Third, 3 participants in the forced guess condition and 2 participants in the voluntary guess condition were excluded for not responding consistently to questions that were repeated three times at time 1; that is, they gave different answers to the same question at time 1. This third exclusion criterion was included in Experiment 2 because we wanted to examine the pattern of time 2 responses *given* that forced confabulated answers were generated once versus thrice at time 1. To include in this analysis, the data from subjects who were doing "something else" at time 1 besides generating the same answer thrice, would render the results uninterpretable. The remaining 79 participants were included in the analyses (M age $= 22.48$ years, $SD = 5.84$; 26 males, 53 females).

Analyses of responses to unanswerable questions

The major results in Experiment 2 pertain to the responses to unanswerable questions. Every participant was presented half of the unanswerable questions once each and the other half thrice each, with the assignment of question to condition counterbalanced across subjects. Although three of the 16 answerable questions were also repeated three times at time 1, the assignment of these three questions to the other experimental conditions was not counterbalanced across subjects; this manipulation was only included to avoid drawing attention to the repeated unanswerable questions.

The primary analyses in Experiment 2 addressed the mean proportion of unanswerable questions that were responded to at time 2 with the same answer, a different answer or a "don't know" response given that the item was answered at time 1. These data are presented in Table 4 for participants in the voluntary guess condition and the forced guess condition for questions presented once versus thrice. Separate 2 (voluntary versus forced guess condition) $\times$ 2 (once- versus thrice-presented condition) Analyses of Variance were conducted on the proportion

Table 4 Mean proportion of unanswerable questions (and standard deviation) that were responded to at Time 2 with the same answer, a ifferent answer, or a "Don't Know" response given that the question was answered at Time 1, in the once- versus thrice-presented conditions in experiment 2

Time 2	Time 1 Voluntary guess ($N = 10$)	Forced guess ($N = 39$)
Once		
Same answer	.50 (.53)	.32 (.34)
Different answer	.20 (.42)	.12 (.19)
Don't know	.30 (.48)	.56 (.39)
	Voluntary guess (N = 22)	Forced guess (N = 39)
Thrice		
Same answer	.78 (.40)	.43 (.34)
Different answer	.13 (.31)	.08 (.22)
Don't know	.09 (.29)	.49 (.36)

of responses that received the same answer, a different answer or a "don't know" response at time 2.

The 2 × 2 analysis of the proportion of responses to unanswerable questions that received the same response at time 1 and time 2 will be reported first. Consistent with the results of Experiment 1, more questions received the same response at time 1 and time 2 in the voluntary guess condition ($M = .69$) than in the forced guess condition ($M = .38$), $F(1, 106) = 10.15, p < .01, \eta^2 = .09$. In addition, more questions received the same response at time 1 and time 2 in the thrice ($M = .56$) than the once ($M = .36$) presented condition, $F(1, 106) = 5.61, p < .05, \eta^2 = .05$. The interaction of these variables was not significant $F(1, 106) = 1.10, \eta^2 = .01$.

The 2 × 2 analysis of the proportion of responses to unanswerable questions that received a different response at time 1 and time 2 yielded no significant effects. This is consistent with the results of Experiment 1.

The 2 × 2 analysis of the proportion of responses to unanswerable questions that received an answer at time 1 but then were responded to at time 2 with a "don't know" response yielded only a significant effect of the voluntary guess versus forced guess condition. Consistent with the results of Experiment 1, the proportion of responses for which the answer at time 2 was changed to "don't know" was significantly higher in the forced guess condition ($M = .52$) than the voluntary guess condition ($M = .16$), $F(1, 106) = 15.65, p < .01, \eta^2 = .13$. This factor did not interact with the repetition condition.

In these three analyses, the results involving the effect of the voluntary guess versus the forced guess conditions replicate the findings in Experiment 1. In addition, across these 3 analyses, the effect of repetition significantly affected only the number of questions that received the same response at time 1 and time 2. As predicted, if eyewitnesses are interviewed multiple times about information that they never saw and are encouraged to guess answers to questions each time, they are even more likely to remember their forced confabulated answers as information actually presented than if they did so only once.

Analyses of responses to answerable questions

Responses to answerable questions at time 1 in the once-presented condition were examined to assess the rate of correct and incorrect responses at time 1 in the forced guess versus voluntary guess conditions. Responses to answerable questions in the thrice-presented condition were not

Table 5 Mean confidence ratings (and standard deviations) for responses to unanswerable questions for which the same response was given at Time 1 and Time 2 by participants in the voluntary guess and forced guess conditions in the once- versus thrice-presented conditions in experiment 2

	Presented Once		Presented Thrice	
	Voluntary guess ($N = 10$)	Forced guess ($N = 39$)	Voluntary guess ($N = 22$)	Forced guess ($N = 39$)
Time 1	2.50 (1.85)	2.43 (2.12)	3.84 (1.82)	2.16 (1.25)
Time 2	3.13 (1.73)	2.95 (1.76)	3.84 (1.82)	3.35 (1.86)

assessed because there were only three questions in this condition, and these three questions were not counterbalanced across conditions. Although the number of correct responses generated was similar in the forced guess (.53) and the voluntary guess conditions (.42), $t\,(77) = .97, p > .05$, $d = .22$, significantly more incorrect responses were generated in the forced guess (.48) than the voluntary guess condition (.27), $t\,(77) = 2.56, p < .01, d = .58$.

Analyses of confidence ratings for unanswerable questions

The confidence expressed in the responses to unanswerable questions was assessed next. A 2 (forced versus voluntary guess condition) × 2 (once- versus thrice-presented) × 2 (time 1 versus time 2) mixed factorial Analysis of Variance was conducted on the confidence ratings for responses to unanswerable questions (range in confidence ratings = 1 – 7). For this analysis we specifically looked at the responses that were consistently given by each subject at time 1 and time 2 to see if for these responses, the confidence ratings differed between individuals who were forced to guess at time 1 and those who were not, in the once versus thrice-repeated conditions. These data are provided in Table 5. Note that the sample sizes across the four conditions vary considerably.

Only the main effects of time and condition were significant in the analyses of confidence data. As in Experiment 1, the confidence expressed in answers to unanswerable questions increased from time 1 ($M = 2.72$) to time 2 ($M = 3.32$), $F(1, 269) = 9.06$, $p < .01$, $\eta^2 = .03$. In addition, guessed answers to unanswerable questions in the forced guess condition ($M = 2.72$) were less confidently held than guessed answers in the voluntary guess condition ($M = 3.33$), $F(1, 269) = 4.16, p < .05, \eta^2 = .02$. Although confidence ratings were higher in the thrice- ($M = 3.30$) than the once-presented condition ($M = 2.75$), this difference was not statistically significant, $F\,(1, 269) = 3.41, p < .07, \eta^2 = .012$. No other effects involving confidence were significant.

General discussion

From the results of previous studies on forced confabulation it has been concluded that forced confabulation increases memory suggestibility; that is, participants develop false memories for events about which they previously confabulated answers. The primary issue addressed in the present study is the extent to which eyewitness memory for a complex event is subsequently suggestively influenced by self-generated information in the form of (a) guessed answers that were forced compared to (b) guessed answers that were volunteered, and the persistence over time in these two types of self-generated responses.

Consistent with previous research, we reported evidence that forced confabulation suggestively influences recall of the target event. Responses to unanswerable questions are most revealing in this study because we know that individuals did not observe in the target video, the information relevant to answering these questions. When participants were forced to guess answers to unanswerable questions at time 1, .40 of the time 2 responses in Experiment 1 and .38 of the time 2 responses in Experiment 2 were the same specific answer to the question that had been generated at time 1, even though the "don't know" response option was available at time 2. These are the rates of forced confabulation. Further, when the same specific answer was given to an unanswerable question at time 1 and time 2, the confidence expressed in the answer increased over time, and this occurred at similar rates in the voluntary guess and the forced guess conditions. However, although forced confabulation does occur as a result of being forced to answer questions, the rate of these false memories is not overwhelming; in no condition in the present study was the modal response to unanswerable questions at time 2 the same answer that had been forcibly guessed at time 1.

In both Experiments 1 and 2, the mean proportion of responses to unanswerable questions that received the same response at time 1 and time 2 was significantly higher in the voluntary guess than the forced guess condition. Further, in both experiments, when given a chance at time 2 to respond "don't know," individuals in the forced guess condition were more likely than those in the voluntary guess condition to shift and respond "don't know" rather than giving the same answer that they gave at time 1. These results suggest that although forced confabulation does occur, information that resulted from forced confabulation is less likely to persist in memory than information that individuals voluntarily responded because they thought they had observed it. To understand these results, the distribution of the strength of information in memory for answerable and unanswerable questions can be conceptualized in terms of the Signal Detection Model of memory (cf. Banks, 1970). In this paradigm, the memory upon which individuals base their answers consists of a network of mutually supporting memories including some information that was observed in the video (i.e., signal) and some related schematic information (i.e., noise). In responding to answerable questions, individuals would base their answers on a distribution of the strength of information in memory that includes both signal and noise, and thus, would on average be of higher strength than the distribution of information in memory for unanswerable questions, that would consist only of noise. Although our theoretical conceptualization of the cognitive processes underlying forced confabulation is based on the Theory of Signal Detection, a signal detection analysis of the results is not possible because we utilized a recall rather than a recognition memory test. A recall test rather than a recognition test was used in this study so that the procedures would be more relevant to interview procedures utilized by police officers.

The manipulation of whether at time 1 responses are forced or voluntarily guessed is considered to affect the placement of the response criterion, with forced guessing lowering the criterion to include more information of lower memory strength. This would account for the two major findings in this study: (a) given that an unanswerable question was answered at time 1, the mean proportion of items that were responded to with the same answer at time 2 was lower in the forced guess condition than in the voluntary guess condition, and (b) given a chance at time 2 to respond "don't know" to an unanswerable question, individuals in the forced guess condition, who generated their forced guesses at time 1 from memories on average of weaker strength, were more likely than those in the voluntary guess condition to respond "don't know" rather than staying with the answer they gave at time 1. Again, these results suggest that forced confabulated information that was never actually observed is less likely to persist in memory than information that individuals volunteered because they thought they had observed it.

Experiment 2 tested the effect of repeated questioning on these outcomes. Unanswerable questions were even more likely to receive the same answer at time 1 and time 2 if the questions had been answered thrice than once at time 1. Our interpretation of this effect is that repetition increases the strength of memory for the suggested information. However, the fact that this occurred at similar rates in the forced guess and voluntary guess conditions suggests that similar processes underlie memory for voluntary and forced confabulated information. These results suggest that if eyewitnesses are interviewed multiple times about information that they in fact never saw and guess answers to specific questions each time, they are more likely to remember their confabulated answers as information actually presented than if they did so only once, and this occurs whether eyewitnesses' confabulated responses were forced or voluntarily provided.

In light of the dearth of research on forced confabulation, we are cautious suggesting conclusions from this study regarding police practices. However, some relevant findings are noteworthy. Responses to answerable questions reveal the effect of forced guessing on questions about which individuals are likely to have observed the relevant information for answering the question. Although more correct responses were generated to answerable questions at time 1 in the forced guess than the voluntary guess condition, this difference was significant only in Experiment 1. Of greater concern, and consistent with the predictions of the Signal Detection Model of memory, in both experiments significantly more incorrect responses were generated to answerable questions in the forced guess than the voluntary guess condition. Together, these results suggest that the practice of pressing eyewitnesses to answer questions may not be an effective one because it reliably increases intrusion errors without reliably increasing recall of correct information. However, this conclusion should be qualified by the practical value of correct, incorrect, and missing information at any specific stage in an investigation. From the standpoint of police investigators, there may be situations in which having more information (including both correct and incorrect information) is more important for generating leads than having less information. Under these conditions, it may be productive to press an eyewitness to answer a question even though the eyewitness has indicated that he or she does not know or is unsure of the answer to the question.

Responses to unanswerable questions reveal the effect of forced guessing on questions for which we know that individuals did not observe in the target video, the information relevant to answering these questions. The results of both experiments suggest that pressing eyewitnesses to answer unanswerable questions is likely to produce a pattern of responses that incorrectly suggests that the responses are credible. That is, 40% of the forced responses to unanswerable questions in Experiment 1 received the same response at time 1 and time 2, even when the "don't know" response was available at time 2, and the confidence expressed in responses to unanswerable questions increased from time 1 to time 2. A similar pattern of results occurred in Experiment 2.

It should be mentioned that because the 6 answerable questions in this study were always different from the 16 unanswerable questions, strictly speaking, question type was confounded by the specific questions themselves. These possible item effects may be a problem to the extent that different questions require more specific or less salient information to be answered. Although there is a potential problem resulting from this confounding, this potential problem is not restricted to this study. In most previous studies in which answerable and unanswerable questions have been included, the question type has been confounded by the specific questions themselves (however, see Shaw, 1996 for an exception). In future studies, question type could be unconfounded from the questions themselves by making up two versions of the video such that the across the two videos, the questions themselves could be counterbalanced across the two question types.

This research is forensically relevant because virtually 100% of all eyewitnesses to crimes, who eventually testify in court, are interviewed by police officers at least once, and typically multiple times (cf. Cutler & Penrod, 1995; Poole & White, 1995). Encouraging eyewitnesses to guess answers to questions about which they report having no memory leads to forced confabulation and an increase in response confidence. With additional research on this topic, the results from this body of research would be important to provide to attorneys and law enforcement personnel (Brigham & Wolfskeil, 1983), as well as to jurors (Brigham & Bothwell, 1983; Noon & Hollins, 1987) and the courts (*Manson v. Braithwaite*, 1977; *Neil v. Biggers*, 1972/1979), who often assume that confident eyewitnesses are accurate eyewitnesses.

Appendix

List of the Answerable and Unanswerable Questions Used in Both Experiments

Answerable Questions:

1. Describe the victim's shirt.
2. Describe the victim's tattoo.
3. What building material were the walls in the computer lab made of?
4. What color was the mouse pad in the computer lab?
5. What time was indicated on the clock in the hallway outside the computer lab?
6. In the lab, what color was the shirt worn by the victim's co-worker?
7. Describe the handrail in the interior stairway.
8. What color backpack was the victim carrying?
9. Inside the building, how many flights of stairs did the victim walk down?
10. How many floor mats were in front of the exit to the building?
11. Just outside the building was an archway. What material was this archway made of?
12. What shape were the planters in the courtyard just outside the building?
13. What traffic sign was posted at the entrance to the parking lot?
14. What type of car was stolen?
15. Describe the shorts the perpetrator was wearing.
16. Which direction did the perpetrator drive as he exited the parking lot?

Unanswerable Questions:

1. What type of soda was the victim's co-worker drinking in the computer lab?
2. What was depicted in the framed picture on the wall in the computer lab?
3. Describe the victim's watch.
4. How many panels did the window in the stairway have?
5. What was in the bed of the white pick-up truck parked in the parking lot?
6. What was the logo on the perpetrator's shirt?

Acknowledgements We are very grateful to Steve Clark for providing us with the car-jacking video that we used in this study. We also thank Shirley Lam for her assistance with data coding. We continue to be grateful to Victor Gambos and Iris Blandon-Gitlin for permitting us to recruit participants from their classes. This project was supported in part by a grant from the Fletcher Jones Foundation to the first author. Kathy Pezdek, in the Department of Psychology at Claremont Graduate University, can be contacted by email at Kathy.Pezdek@cgu.edu.

References

Ackil, J. K., & Zaragoza, M. S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology*, *34,* 1358–1372.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, *74,* 81–99.

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior*, *7,* 19–30.

Brigham, J. C., & Cairns, D. L. (1988). The effect of mugshot inspection on eyewitness identification accuracy. *Journal of Applied Social Psychology*, *18,* 1394–1410.

Brigham, J. C., & Wolfskeil, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identification. *Law and Human Behavior*, *7,* 337–349.

Clark, S. E., & Tunnicliff, J. L. (2001). Selecting lineup foils in eyewitness identifications: Experimental control and real-world simulation. *Law and Human Behavior*, *25,* 199–216.

Cutler, B. L., & Penrod, S. D. (1995). Mistaken identification: The eyewitness, psychology, and the law. Cambridge: Cambridge University Press.

Gudjonsson, G. (1992). The psychology of interrogations, confessions and testimony. New York: John Wiley & Sons.

Hastie, R., Landsman, R., & Loftus, E. F. (1978). Eyewitness testimony: The dangers of guessing. *Jurimetrics Journal*, *19,* 1–8.

Hinz, T., & Pezdek, K. (2001). The effect of exposure to multiple lineups on face identification accuracy. *Law and Human Behavior*, *25,* 185–198.

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

Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist*, *52,* 221–233.

Leo, R. A. (1996). Inside the interrogation room. *The Journal of Criminal Law and Criminology*, *86,* 266–303.

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16,* 1077–1083.

Lindsay, D. S., & Johnson, M. K. (1989). The reversed eyewitness suggestibility effect. *Bulletin of the Psychonomic Society*, *27,* 111–113.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7,* 560–572.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4,* 19–31.

Manson v. Braithwaite, 432 U.S. 98, 112, 97 S.Ct. 2243, 53 L.Ed.2d 140 (1977).

Munsterberg, H. (1908). On the witness stand: Essays on psychology and crime . Garden City, NY: Doubleday.

Neil v. Biggers, 409 U.S. 188, 34 L.Ed.2d 401 (1972), *cert. denied*, 444 U.S. 909, 100 S.Ct. 221, 62 L.Ed.2d 144 (1979).

Noon, E., & Hollins, C. R. (1987). Lay knowledge of eyewitness behaviour: A British survey. *Applied Cognitive Psychology*, *1,* 143–153.

Pezdek, K. (1977). Cross-modality semantic integration of sentence and picture memory. *Journal of Experimental Psychology: Human Learning and Memory*, *3,* 515–524.

Pezdek, K., & Blandon-Gitlin, I. (2005). When is an intervening lineup most likely to affect eyewitness identification accuracy? *Legal and Criminological Psychology*, *10,* 247–263.

Poole, D. A., & White, L. A. (1995). Two years later: Effects on the eyewitness testimony of children and adults. *Developmental Psychology*, *29,* 844–853.

Schreiber, N., Wentura, D., & Bilsky, W. (2001). What else could he have done? Creating false answers in child witnesses by inviting speculation. *Journal of Applied Psychology*, *86,* 525–532.

Shapiro, L. R., & Purdy, T. L. (2005). Suggestibility and source monitoring errors: Blame the interview style, interviewer consistency, and the child's personality. *Applied Cognitive Psychology*, *19,* 489–506.

Shaw, J. S., III. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied*, *2,* 126–146.

Shaw, J. S., III, & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior*, *22,* 629–653.

Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science*, *7,* 294–300.

Zaragoza, M. S., Payment, K. E., Ackil, J. K., Drivdahl, S. B., & Beck, M. (2001). Interviewing witnesses: Forced confabulation and confirmatory feedback increase false memories. *Psychological Science*, *12,* 473–477.