AUSTRALIAN JOURNAL OF CLINICAL AND EXPERIMENTAL HYPNOSIS

May 1991 **Volume 19 Number 1**

AUSTRALIAN JOURNAL OF CLINICAL AND EXPERIMENTAL HYPNOSIS

Manuscripts and editorial matters should be addressed to the Editor, Dr Wendy-Louise Walker, 14 Hammond Avenue, Croydon, NSW 2132, Australia. All journal business communications and subscriptions should be addressed to the Editor.

Australian Journal of Clinical and Experimental Hypnosis
Vol. 19, No. 1, 1991, pp. 1–18

AMBIGUITY IN SUGGESTION AND THE OCCURRENCE OF PSEUDOMEMORY IN THE HYPNOTIC SETTING

Peter W. Sheehan, Dixie Statham, Graeme A. Jamieson, and Scott R. Ferguson

University of Queensland

Two separate studies were conducted in a program of research designed to analyse the impact of the manipulation of ambiguity of communication on the occurrences of hypnotically induced pseudomemory. Study 1 established base rates of response for conditions associated with previously observed effects. Study 2 varied context by more clearly limiting the setting in which acceptance of false information was appropriate to hypnotic testing. It was predicted that as the communication components of the total context became more distinct, then the incidence of pseudomemory response would reduce. One hundred and forty-two subjects and 116 subjects were tested in Studies 1 and 2, respectively, each sample being tested in a 3 (level of susceptibility: high, medium, low) × 2 (state instruction: waking, hypnosis) design. Past effects were replicated in Study 1 and Study 1 demonstrated stronger behavioural effects than were evident in Study 2. Overall, results highlighted the impact of ambiguity and showed that definition of the context of testing influences the incidence of hypnotically induced pseudomemory.

This paper deals with the phenomenon of pseudomemory and the processes that underlie it. Its theoretical implications and ramifications are necessarily both complicated and diverse, and deal with the essential relationship between the phenomena of hypnosis and memory. Hypnosis has been shown to be highly complex and multifaceted in its own right (Fromm & Shor, 1979; Orne, 1959; Sheehan & McConkey, 1982), as has memory (Jenkins & Dallenbach, 1924; Mayes, 1983; Wessells, 1982). Furthermore, each has shown to be capable of exerting an influence on the other (e.g., Labelle, Laurence, Nadon, & Perry, 1990; Laurence & Perry, 1983; Orne, 1962).

Pseudomemory may be defined as the occurrence of the report of an altered memory reflecting the acceptance of false information that has been suggested

This reearch was supported by a grant to the first author from the Australian Research Council. The authors are grateful to Patricia Truesdale, Vanessa Green, Jo Dower, and Rosemary Robertson for their help in the study.

Requests for reprints should be sent to P. W. Sheehan, Department of Psychology, University of Queensland, St. Lucia, Queensland 4067.

previously. Various parameters affect it, including hypnotic skill (Labelle & Perry, 1986), contextual factors (Lynn, Weekes, & Milano, 1989; McCann & Sheehan, 1988; Spanos & McLean, 1986), and presence of hypnotic instruction (Sheehan, Statham, & Jamieson, in press). What is not clear, however, is the nature of the processes that underlie it.

The study of pseudomemory in the environment of the experimental laboratory essentially began with the work of Laurence and Perry (1983), who suggested to 27 highly hypnotisable subjects, during hypnosis, that they had been awakened during the night in the previous week by loud noises. Seventeen subjects (63%) acknowledged hearing the noises, and in the hypnosis session were prompted to give detailed descriptions. Subjects were then tested for the implanted pseudomemory either immediately following trance, or one week later. Thirteen (48%) accepted the suggestion, while some of the remaining 14 subjects idiosyncratically deduced that the event had never occurred. Six of the 13 subjects who accepted the pseudomemory appeared to adopt it as their own and were convinced of its validity.

Further relevant evidence has flowed from the experimental analysis of pseudomemories since this original work. Lynn et al. (1989), for instance, investigated whether pseudomemories have a persistent nature. This was qualified by the results of McCann and Sheehan (1987), who demonstrated that pseudomemories are reversible, a finding later confirmed by McConkey, Labelle, Bibb, and Bryant (1990). Pseudomemory is clearly a labile phenomenon that is susceptible to alteration, and perhaps especially so when contextual variables act as a vehicle for that alteration.

One way of usefully studying the processes that underlie pseudomemory is to examine the factors that limit the phenomenon in relation to conditions that establish it initially. This strategy has been shown to be effective, for instance, in isolating the relevance of rapport to the phenomenon of countering preconceptions about hypnosis — which decreases significantly when the rapport between hypnotist and subject is systematically reduced (Sheehan, 1980). Generalising this logic to the study of pseudomemory, it should be particularly useful to isolate a set of conditions that will reduce the phenomenon. Examination of the nature of the effects accompanying the change in procedures should provide relevant cues to the identification of parameters that underlie pseudomemory.

Present research reports two studies that illustrate this logic. Study 1 studied conditions in a 2×3 design involving the manipulation of both the level of susceptibility and state (hypnotic vs. waking) instruction; this provided base procedures of testing. Study 1 also served to test for replication of the effects previously observed by Sheehan et al. (in press), which demonstrated an appreciably higher rate of reported pseudomemory for hypnotic than for waking subjects, and for high than for low or medium susceptible subjects, as well as a significant interaction between level of susceptibility and state instruction. Study 2, using the same comprehensive design on an independent sample of subjects, varied the base procedures of testing in an attempt to reduce the incidence of pseudomemory. Specifically, procedures were altered to reduce the ambiguity in the communicated appropriateness of continued responding in the context of posthypnotic testing (as distinct from the context of the original false suggestions). This involved the elimination of any reference implying a carry-over of effect from the hypnotic to the posthypnotic context, including the elimination of all suggestions about how memory works in hypnosis and any communication implying there was an expectation of particular performance when pseudomemory would be tested later.

It was predicted in the program of work that base procedures of testing would replicate the effects observed previously: main effects should be observed in Study 1 for state instruction and level of hypnotic skill, and the greatest incidence of pseudomemory should occur for high susceptible subjects receiving hypnotic instruction. The incidence of pseudomemory should be lower in Study 2 and the pattern of effects should change to reflect the increased impact of the contextual (ambiguity of communication) manipulation.

Specifically, it is argued that a communication blurring the distinction between responding in a setting which suggests false information and one which tests for its effect (Study 1) should be more effective in establishing pseudomemory than a communication which more sharply separates the two settings and limits expectancy of transfer of effect (Study 2).

STUDY 1

METHOD

Subjects

One hundred and forty-two subjects, ranging in age from 17 to 41 years (mean age 21.0 years) were tested individually in this study. Subjects were preselected for hypnotic responsiveness on the Harvard Group Scale of Hypnotic Susceptibility, Form A (HGSHS:A; Shor & Orne, 1962), and obtained either high (9–12), medium (5–7), or low (0–3) susceptibility scores. All subjects' scores were confirmed on the Stanford Hypnotic Susceptibility Scale, Form C (SHSS:C; Weitzenhoffer & Hilgard, 1962). Mean susceptibility scores of 9.80 for highs, 6.16 for mediums, and 1.58 for lows were obtained on the HGSHS:A. On the SHSS:C, these groups obtained mean scores of 10.23, 6.29, and 1.08. Fifty-six highs, 38 mediums, and 48 lows were allocated randomly to the 3×2 (susceptibility: high, medium, low; state instruction: hypnosis, waking) factorial design. All subjects received course credit and payment for their participation.

Materials

The stimulus material consisted of a videotaped vignette of a bank robbery developed by Yuille (1982). The videotape shows a male bank robber entering a bank from the left of the screen with a pistol in his hand. While gesturing with the pistol he tells the tellers not to press the alarm and instructs them to put the money on the counter. He then places the money into a bag and backs out of the bank. The robber does not have a mask and does not swear. The videotape was approximately 43 seconds long.

Paivio's (1971) Ways of Thinking Questionnaire and the Gordon Test of Visual Imagery Control (Gordon, 1949) were completed by waking subjects in order to equate for time spent by the hypnotic subjects in hypnotic induction and deinduction.

Procedure

At the beginning of the session, subjects were provided with an explanation of the study in order to establish consistent perceptions of the experiment prior to beginning formal testing. The hypnosis group was told by E1 that the purpose of the study was to investigate the relationship between hypnosis and how people rated their emotional arousal in response to watching particular events. The waking group was informed that the study aimed to investigate the relationship between observation of actual events and ratings of emotional arousal. Subjects were then shown the videotape of the robbery in which the robber was neither swearing nor wearing a mask. They then rated the robber's aggressiveness on a 6-point scale ranging from *low aggressive* (1) to *extremely aggressive* (6). Subjects were then asked to recall the events of the robbery in as much detail as possible and rated their confidence in the accuracy of their recall. Confidence ratings ranged from *not at all confident* (1) to *extremely confident* (6). At this point E1 asked subjects eight questions about the robbery, including whether the robber entered from the left or the right, and obtained confidence ratings of the accuracy of each response.[1]

Following this, subjects in the hypnosis condition received hypnotic induction instructions, while those in the waking condition completed Paivio's Way of Thinking Questionnaire.

All subjects were then administered a range of suggestion tasks that included hand lowering, fractionation (hypnosis group only), arm immobilisation (waking group only), anosmia, pseudo-anaesthesia, dream and age-regression. They were then regressed back to their observation of the video to enable implantation of the false memories and those in the hypnosis group were asked to remember, in hypnosis, the events of the video in as much detail as they could. Subjects were further told routinely that they might remember events and details that their conscious mind had forgotten. It was then suggested that: "Everything is recorded and *in hypnosis* you can remember it if you try."

Suggestions for the three false memory items were then given, beginning with the suggestion that the robber entered from the right and was wearing a mask:

> You're watching the scene unfold before you. First the door opens and the robber comes in from your right . . . he has something like a mask

> over his face. He is wearing a mask. Look carefully now and tell me whether you can see the mask and maybe some of the robber's face through his mask. Look carefully now and tell me when you can see him standing there with the mask over his face and gun in hand.

If the suggestions were rejected, subjects were told:

> That's fine. . . Maybe you don't consciously remember the robber's mask, but as I told you before, you may find there are things your conscious mind has forgotten but your subconscious mind remembers them. Let's review the robbery from the beginning again.

The mask suggestion was repeated and then, regardless of whether it was accepted or rejected, the final false memory suggestion was introduced for all subjects. Subjects were told to:

> Picture the robbery again in hypnosis now . . . let the scene unfold some more. The robber says that it's a stick-up but if everyone does as he says, no one will get hurt. There's a man at the counter — a customer — and the robber tells him to move! Get down on the ground! Spread your arms out! Then he tells the tellers to put the bills up on the counter, the big bills. He is very aggressive and he is swearing, swearing very much, swearing a lot. Look carefully now and tell me when you hear him being aggressive, hear him swearing. Tell me when you hear him swearing a lot, so I know you are right back there, seeing and hearing him being aggressive and swearing at the tellers.

If this suggestion was rejected, subjects were again told that their subconscious mind might be able to remember details their conscious mind had forgotten and the suggestion was repeated.

Deinduction instructions were then administered to subjects in the hypnosis instruction condition, and those in the waking condition completed Gordon's Test of Visual Imagery Control. All subjects then rated the extent to which they felt they had been under hypnosis, ranging from *not at all hypnotised* (1) to *profoundly hypnotised* (6).

At this point, subjects were introduced to E2, who asked them to rate the aggressiveness of the robber on the same scale used previously. Subjects' free recall of the robbery events was then tested and they rated their confidence on the 6-point confidence scale. Structured recall was then tested by E2 asking the same eight questions asked by E1, as well as an additional two questions relating to the mask and swearing. These questions were: "Was the robber wearing a mask?" and "Was the robber swearing a lot?" Subjects rated their confidence for all 10 structured recall questions.

Finally, E2 tested for recall of events by Sheehan and McConkey's (1982) Experiential Analysis Technique (EAT). This technique focused on the subjects' images and cognitive processes associated with their experiences during testing. Specifically, subjects were informed that they were going to watch a videotape of part of the session with E1 where they were reviewing the videotape of the bank robbery. Subjects were told that:

> As you watch the tape now, you'll find that you can remember all sorts of thoughts and feelings you had at that time. I'd like you to stop the playback anytime you recall your reactions to the suggestions, the pictures going on in your mind, or anything else you were experiencing at that time, and just tell me about what it is you are recalling.

Subjects' responses were recorded on audiotape.

RESULTS

Validity Check

A 3 × 2 (susceptibility × state instruction) analysis of variance of subjects' ratings of the extent to which they felt they had been hypnotised showed significant main effects for susceptibility (F (2, 136) = 81.27, $p < .001$) and for state instruction (F (1, 136) = 152.52, $p < .001$), and a significant interaction effect (F (2, 136) = 31.77, $p < .001$). High, medium, and low susceptible subjects in the hypnosis group reported a greater depth of hypnosis than did subjects in the waking group. Further, as expected, high and medium susceptible subjects in the hypnosis group indicated a greater depth of hypnosis than lows in the hypnosis group.

Measurement of Accuracy

A pseudomemory score was obtained for each subject by summing the number of false memory items reported for test of memory of the critical items (right entry, mask, swearing). Pseudomemory scores ranged from zero to a maximum of 3 and were analysed for free recall and structured recall.[2]

Accuracy and Free Recall

Analysis of variance of pseudomemory scores for free recall showed significant main effects for susceptibility (F (2, 136) = 9.01, $p < .001$) and for state instruction (F (1, 136) = 15.55, $p < .001$), and a significant interaction effect (F (2, 136) = 3.66, $p < .05$), thus replicating previously observed effects. High susceptible subjects given hypnosis instructions reported a greater number of pseudomemories compared to all other subjects given the same instructions, and compared to all subjects given waking instructions.

Since pseudomemory varies with type of stimulus suggested (McCann & Sheehan, 1988), the incidence of pseudomemory in free recall was analysed for each of the three items individually by conducting logistic regression analyses on the number of subjects in each experimental group who either reported that the robber entered from the right, that he wore a mask, or that he swore. Logistic regression analysis (see Bishop, Fienberg, & Holland, 1975, for details of the technique) showed main effects for susceptibility (χ^2 (2) = 8.67, $p <$ .05) and for state instruction (χ^2 (1) = 8.09, $p < .01$) for the right-entry item. Similar effects were also found for the mask item for susceptibility (χ^2 (2) = 11.89, $p < .01$) and state instruction (χ^2 (1) = 8.91, $p < .01$). The swearing

Table 1 Number of Subjects Reporting Pseudomemory for Right-Entry, Mask, and Swearing Items in Free Recall and Structured Recall in Study 1

Memory test	Stimulus	Level of susceptibility and state instruction					
		High		Medium		Low	
		Hypnosis	Waking	Hypnosis	Waking	Hypnosis	Waking
Free recall	Right	10(35.7)	3(10.7)	2(10.5)	0(0.0)	3(12.5)	0(0.0)
	Mask	15(53.6)	4(14.3)	5(26.3)	2(10.5)	1(4.2)	1(4.2)
	Swear	6(21.4)	0(0.0)	0(0.0)	0(0.0)	1(4.2)	0(0.0)
Structured recall	Right	11(39.3)	4(14.3)	2(10.5)	1(5.3)	4(16.7)	0(0.0)
	Mask	16(57.1)	5(17.9)	4(21.1)	1(5.3)	3(12.5)	3(12.5)
	Swear	10(35.7)	3(10.7)	1(5.3)	1(5.3)	1(4.2)	0(0.0)

Note. Entries in parentheses refer to percentages.

item could not be analysed completely because of the high proportion of empty cells. Table 1 reports the number of subjects in each group for test of free recall (and structured recall) who reported pseudomemories. A significantly greater number of high susceptible subjects given hypnosis instructions reported pseudomemories compared to all other groups for both types of memory test.

Accuracy and Structured Recall

Analysis of subjects' responses to the three critical direct questions asked by E2 showed effects that were very similar to those found for free recall. Data are reported in full in Table 1. A significant susceptibility by state instruction interaction (F (2, 136) = 3.29, $p < .05$) and significant main effects for susceptibility (F (2, 136) = 11.23, $p < .001$) and for state instruction (F (1, 136) = 13.52, $p < .001$) again indicated that incidence of pseudomemory was strongly related to level of hypnotic susceptibility (high vs. medium and low) and to type of instruction (hypnosis vs. waking). Examination of the data for individual items by logistic regression analysis showed significant main effects. For level of susceptibility, effects were significant for right entry (χ^2 (2) = 8.62, $p < .05$), mask (χ^2 (2) = 11.57, $p < .01$), and swearing (χ^2 (1) = 10.08, $p < .01$). For state instruction, effects were significant for right entry (χ^2 (1) = 7.41, $p < .01$), mask (χ^2 (1) = 8.08, $p < .01$), and swearing (χ^2 (1) = 4.56, $p < .05$).

Confidence and Free Recall

Analysis of variance of subjects' confidence ratings of their free recall accounts of the robbery showed no significant main or interaction effects. A comparison of the confidence ratings of those subjects who displayed pseudomemory (by reporting at least one of the three items in their recall accounts) with those who did not, also failed to show any significant differences in confidence.

Confidence and Structured Recall

ANOVA of subjects' confidence ratings for each of the three pseudomemory items showed a mixture of effects.

The mask and swearing items both demonstrated significant effects. For the mask item, a significant main effect for susceptibility (F (2, 136) = 3.23, $p < .05$) showed that high susceptible subjects were *less* confident than low susceptible subjects, with medium susceptible subjects being the least confident of all groups. For the swearing item, a main effect for susceptibility (F (2, 136) = 9.75, $p < .001$) indicated that low susceptible subjects were significantly more confident of their responses than either highs or mediums, with high susceptible subjects being the least confident of all. A significant main effect for state instruction (F (1, 136) = 13.69, $p < .001$) showed that hypnotic subjects were less confident than waking subjects. No main or interaction effects occurred for the entry from the right item.

Analysis of EAT Data

Analysis of the audiotaped EAT sessions was conducted to explore the nature of subjects' reponses to the pseudomemory suggestions from a phenomenological perspective. Data indicated that many subjects who had responded affirmatively to the false memory suggestions given by E1 did not relinquish or alter their original memory of the robbery events. For example, some subjects accepted E1's version of the events as being correct because of perspectives of her trustworthiness and because "she must know what happened"; others reinterpreted the "graininess" in the videotape as evidence that the robber was wearing a mask; others assumed that E1 was asking them to *imagine* the events as she described them, and then assumed that E2 was asking them to report on these imagined events. For the purposes of analysis, all these responses were regarded as non-pseudomemory responses, and were given a score of 1 on a 3-point pseudomemory-experience scale. If subjects indicated that they were aware of two images of the robbery (the actual robbery and the suggested robbery) but were confused as to which was their actual memory, they were given a score of 2 on the experience scale. A score of 3 was given to those subjects who maintained a delusory response that the false version of the robbery events was their actual memory. Each subject was rated on this 3-point scale for all three items.

Analyses of variance of the pseudomemory experience scores for the right-entry item showed significant main effects for susceptibility (F (2, 136) = 9.71, $p < .001$) and state instruction (F (1, 136) = 4.56, $p < .05$), and a significant interaction effect (F (2, 136) = 3.50, $p < .05$). High susceptible subjects receiving hypnosis instruction showed a greater incidence of distortion in memory experience during suggestion than any other group.

For the mask item significant main effects were found for susceptibility (F (2, 136) = 8.73, $p < .001$), and state instruction (F (1, 136) = 16.45, $p <$ .001), together with a significant interaction effect (F (2, 136) = 4.55, $p <$

.05). Once again, high susceptible subjects receiving hypnosis instruction showed the greatest incidence of subjective distortions in memory for the false suggestion.

For the swearing item significant main effects were found again for susceptibility (F (2, 136) = 8.75, $p < .001$) and state instruction (F (1, 136) = 7.41, $p < .01$). A significant interaction effect failed to be obtained (F (2, 136) = 2.86, p = .06). High susceptibles receiving hypnosis continued to be the most outstanding response group.

Summary

Data for Study 1 importantly reproduced the pattern of findings observed previously (Sheehan et al., in press). Specifically, level of susceptibility and state instruction were firmly related to subjects' pseudomemory reports, and there was a significant interaction between the two factors. Overall, data indicated that the base procedures employed in Study 1 were effective in producing strong pseudomemory response.

STUDY 2

In Study 2, an attempt was made to lower the incidence of pseudomemory reports by reducing the ambiguity in communication about the appropriateness of the subject's continuing to respond outside the hypnotic context. This was done by eliminating any cues that implied or suggested specifically that there would be a carry-over effect from suggestion with E1 to testing with E2. Replication of effects found in Study 1 allows one to infer that, if the incidence of pseudomemory report is altered in Study 2, then the changes in procedures introduced in the previous experiment are responsible. In this way the design focuses on the impact of the contextual variable (ambiguity of communication) on the production of subjects' pseudomemory reports. Samples tested in Studies 1 and 2 were entirely independent, no subject participating across the two experiments. As in Study 1, two experimenters were involved, E1 being common to the two studies.

METHOD

Subjects

One hundred and sixteen subjects with ages ranging from 17 to 48 years (mean age 21.0) participated in the study. All subjects were pretested for hypnotic susceptibility on the HGSHS:A and had their susceptibility scores confirmed on the SHSS:C. Subjects obtained either high (9–12), medium (5–7), or low (0–3) susceptibility scores. Fifty highs, 34 mediums, and 32 lows obtained mean susceptibility scores of 10.02, 5.94, and 2.09, respectively, on the HGSHS:A. On the SHSS:C, these groups obtained mean scores of 10.06, 6.12, and 1.64.

Subjects were assigned to a 3 (susceptibility: high, medium, low) × 2 (state instruction: hypnosis, waking) independent groups design and were tested individually. All subjects were paid and given course credit for their participation in the study.

Materials

The videotape stimulus was identical to that used in Study 1. The Tellegen Absorption Scale (Tellegen & Atkinson, 1974), the shortened form of the Betts Questionnaire of Mental Imagery (Sheehan, 1967) and the Gordon Test of Visual Imagery Control were used as filler activities for the waking group to equate for time spent by the hypnotic group in induction and deinduction.

Procedure

Prior to the introduction of hypnotic instruction for the experimental group, procedures were identical to those used before. All subjects were shown the videotape and then rated the robber's aggressiveness on a 6-point scale ranging from low aggressive to extremely aggressive. Subjects were then asked to recall the events of the robbery and, following this, they were asked eight questions about the robbery (including the direction of entry question). Confidence ratings of responses (which ranged from (1) *not at all confident* to (6) *extremely confident*) were also obtained. Individuals in the hypnosis group were then given hypnotic induction instructions, while those in the waking group completed the Tellegen Absorption Scale and the Betts Questionnaire of Mental Imagery. Finally, the same suggestion tasks were then administered to all subjects.

Regression instructions, which took subjects back to their earlier viewing of the video, implanted the false information and incorporated the changes in procedure that defined the main context manipulation.

Verbatim instructions were as follows:

> Earlier in this session, before you were hypnotised, we watched a video together, a video of a bank robbery. And as I talk to you now, while you are here with me in hypnosis, you will remember those events very clearly. It is important, here with me now, that you remember the events of the video in as much detail as you can. Everything that happened, even if you do not immediately recall it, you can remember it here with me now, in hypnosis, if you really try. As I recount the events of the bank robbery to you in hypnosis you will be able to see and hear the events in your mind as they are happening, in hypnosis, here with me now. You can remember if you try.

At no time were subjects (as in Study 1) told that their memories would always be available subconsciously. Such was seen to imply access to particular memories at a later point in time and could reasonably be interpreted as implying or suggesting transfer or carry-over of effect.

Actual suggestions for the three false memory items were given in exactly the same way as Study 1. Subjects were first told that the robber entered from the right and then that he was wearing a mask. If the suggestion about the mask was rejected, subjects were told:

> That's fine . . . Maybe you don't think you remember the robber's mask. But as I told you before, you can remember it for me now in hypnosis. These things are happening just as I say they are. Let's review the robbery from the beginning again.

The suggestion was repeated, and then followed by the false memory suggestion for the robber swearing. If the swearing suggestion was not accepted, subjects were told: "Maybe you don't think you see him being aggressive and swearing. But as I said before, you will find these things happening." The suggestion was then repeated.

Deinduction instructions, which incorporated further separation of the hypnotic–nonhypnotic contexts, were then given to the hypnosis group:

> After you wake up, you will remember everything . . . all that you saw on the video and experienced here with me now. You will remember everything that happened.

Waking subjects were also told that they would remember everything they saw on the video and experienced with the experimenter. They then completed Gordon's Test of Visual Imagery Control.

At this point, subjects were introduced to E2, who asked them to rate the robber's aggressiveness and then to recall the events of the robbery. Instructions for recall emphasised that the subject was required to recall only those events depicted on the video:

> Now can you tell me in your own words, in as much detail as possible, what happened in the robbery incident that you saw on the video right at the start of the experiment. I am only interested in what you saw on the tape at the start of the study, nothing else.

Ten structured questions (the eight asked by E1 and an additional two relating to the mask and the swearing) were then asked by E2. Confidence ratings of all responses (free and structured recall) were also obtained.

Finally, EAT instructions, focusing on subjects' subjective experiences of the robbery item, were administered.

RESULTS

Validity Check

A 3 (susceptibility) × 2 (state instruction) ANOVA of subjects' ratings of hypnotic depth showed significant main effects for susceptibility (F (2, 110) = 42.87, $p < .001$) and state instruction (F (1, 110) = 91.93, $p < .001$), and a significant interaction effect (F (2, 110) = 18.67, $p < .001$). Depth of hypnosis was reported as being greatest by those subjects given hypnosis instructions, with high and medium susceptible subjects reporting greater hypnotic depth than low susceptible subjects.

Measurement of Accuracy

The number of false memories reported was converted, for each subject, to a pseudomemory score, ranging from 0 to 3. Pseudomemory scores were analysed separately for free recall and structured recall, as in Study 1.

Accuracy and Free Recall

ANOVA of pseudomemory scores for free recall showed a significant main effect for state instruction (F (1, 110) = 6.31, $p < .05$), indicating that subjects receiving hypnosis instructions reported significantly more pseudomemories than those who received waking instructions. There was no effect for susceptibility, however, as in Study 1, and no interaction effect between susceptibility and state instruction.

Table 2 shows the number of subjects in each experimental group who demonstrated pseudomemory for each of the three items individually. Logistic regression analysis (chi-square) of the data indicated, for the mask item, a significant main effect for state instruction (χ^2 (1) = 4.59, $p < .05$). Analyses showed that more of those who received hypnosis instructions reported that the robber wore a mask than did those who received waking instructions.

Accuracy and Structured Recall

For structured recall, ANOVA of subjects' pseudomemory scores showed a significant main effect for state instruction (F (1, 110) = 9.14, $p < .01$). Data are reported fully in Table 2. Subjects who received hypnosis instructions reported a greater number of pseudomemories than subjects who received waking instructions. Again, unlike Study 1, there was no significant susceptibility or interaction effect.

Table 2 Number of Subjects Reporting Pseudomemory for Right-Entry, Mask, and Swearing Items in Free Recall and Structured Recall in Study 2

Memory test	Stimulus	Level of susceptibility and state instruction					
		High		Medium		Low	
		Hypnosis	Waking	Hypnosis	Waking	Hypnosis	Waking
Free recall	Right	2(8.0)	1(4.0)	4(23.5)	0(0.0)	0(0.0)	1(6.3)
	Mask	6(24.0)	1(4.0)	4(23.5)	0(0.0)	1(6.3)	2(12.5)
	Swear	2(8.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)	0(0.0)
Structured recall	Right	9(36.0)	2(8.0)	6(35.3)	0(0.0)	2(12.5)	1(6.3)
	Mask	9(36.0)	2(8.0)	3(17.7)	2(11.8)	2(12.5)	3(18.8)
	Swear	7(28.0)	2(8.0)	2(11.8)	1(5.9)	1(6.3)	0(0.0)

Note. Entries in parentheses refer to percentages.

Logistic regression analysis of the data for individual items indicated a significant main effect for state instruction for the entry from the right item (χ^2 (1) = 9.66, $p < .01$) and for the swearing item (χ^2 (1) = 18.72, $p < .001$). A significant interaction effect was also found for the swearing item (χ^2 (2) = 14.71, $p < .001$). Subjects receiving hypnosis instructions were more likely to report that the robber swore, compared to those who received waking instructions, and the more susceptible subjects were more likely to report swearing than less susceptible subjects. High susceptible subjects receiving hypnosis instruction demonstrated the highest incidence of pseudomemory for the swearing item.

Confidence and Free Recall

Analysis of variance of subjects' confidence ratings of responses for free recall showed no significant main or interaction effects. Comparison of the confidence ratings of those subjects who demonstrated pseudomemory (by indicating at least one pseudomemory response in their recall) with those who failed to report pseudomemory also showed no significant differences between the two groups.

Confidence and Structured Recall

For structured recall, ANOVA of subjects' confidence ratings for the three items indicated a significant main effect for the swearing item only (F (1, 110) = 4.58, $p < .05$). Subjects receiving hypnosis instructions were appreciably less confident about their responses than those receiving waking instructions.

Comparison of the confidence ratings of subjects who reported pseudomemory and those who failed to report pseudomemories, for each of the three items, indicated that those who reported pseudomemories were significantly less confident of their response than those who failed to report pseudomemories: For right entry, t (114) = 2.04, $p < .05$; for the mask, t (114) = 2.02, $p < .05$; and for the swearing, t (114) = 2.39, $p < .05$.

Overall, as distinct from accuracy, results for confidence in Study 2 generally duplicated the pattern of data found in Study 1. The confidence of subjects in their memory reports was lowered, not raised, when pseudomemory was evident.

Analysis of EAT Data

Subjects' audiotaped EAT sessions were scored for experiential response to pseudomemory suggestions, for each of the three items, in a manner identical to that used in Study 1.

Analyses of variance of the pseudomemory-experience scores for the right-entry item indicated significant main effects for susceptibility (F (2, 110) = 4.06, $p < .05$) and for state instruction (F (1, 110) = 8.86, $p < .01$), and a significant interaction effect between the two factors (F (2, 110) = 4.06, $p < .05$). High susceptible subjects given hypnosis instructions showed a greater incidence of pseudomemory experience than did all other groups, with high susceptible subjects demonstrating the greatest effect.

For mask item also, significant main effects were found for level of susceptibility (F (2, 110) = 5.35, $p < .01$) and state instruction (F (1, 110) = 8.10, $p < .01$), and a significant interaction effect (F (2, 110) = 5.35, $p < .05$) was found. Again, high susceptible subjects receiving hypnosis instructions showed a greater incidence of pseudomemory experience.

Further, for the swearing item, there were significant main effects for susceptibility (F (2, 110) = 3.54, $p < .05$) and state instruction (F (1, 110) = 11.40, $p < .001$). A significant interaction effect between these two factors (F (2, 110) = 3.54, $p < .05$) indicated that high and medium susceptible subjects receiving hypnosis instructions showed the greatest pseudomemory experience for swearing, with this effect being most evident for high susceptible subjects.

CROSS-COMPARISON OF STUDY 1 AND STUDY 2

Data from Study 1 and Study 2 were combined to provide direct cross-comparison of findings in the two separate experiments. Emphasis was placed on analysis of accuracy, where group differences across studies were evident.

Free Recall

A 3-way analysis of variance (susceptibility × state instruction × type of study: Study 1, Study 2) was conducted on the number of pseudomemories reported by subjects in the two experiments. A significant main effect was found for type of study (F (1, 246) = 5.41 $p < .05$) with significant main effects also being found for susceptibility (F (2, 246) = 7.93, $p < .001$), and state instruction (F (1, 246) = 21.83, $p < .001$). Significant interaction effects were found for susceptibility × state instruction (F (2, 246) = 4.73, $p = .01$) and for susceptibility × type of study (F (2, 246) = 3.77, $p < .05$). Detailed observation of the data showed that high susceptible subjects in Study 1 reported more pseudomemories than those in Study 2, while low and medium susceptible subjects reported more pseudomemories in Study 2 than in Study 1. Across the two studies, highly susceptible subjects who received hypnosis reported a greater number of pseudomemories than other groups.

Structured Recall

Analysis of variance (susceptibility × state instruction × type of study) indicated no main effect for type of study. There were significant main effects for susceptibility (F (2, 246) = 10.46, $p < .001$) and for state instruction (F (1, 246) = 22.46, $p < .001$), and a significant interaction between these two factors (F (2, 246) = 4.34, $p < .05$). High susceptible subjects receiving hypnosis instructions reported a greater number of pseudomemories than all other groups.

DISCUSSION

Data support the main prediction of this program of research that the occurrence of pseudomemory will be significantly less when the ambiguity of communication about the appropriateness of continued responding outside the context of the initial suggestion is reduced. More subtle, but clearly important effects, were evident, however, in the changing pattern of pseudomemory responses across studies and these changes provide useful clues to the nature of the mediating processes underlying pseudomemory responses.

For instance, the change in pattern of significant results for total accuracy scores from Study 1 to Study 2 was consistent for both free recall and structured recall. For both these memory test formats, there were appreciable state instruction, susceptibility, and state instruction × susceptibility interaction effects in Study 1; but in Study 2 the effects for level of susceptibility and the interaction of susceptibility with state instruction were eliminated, leaving only state instruction (hypnotic vs. waking) as a significant effect. This changing pattern of data across the experimental settings of Study 1 and Study 2 suggests that pseudomemory response is a multi-faceted phenomenon, likely to involve a mix of mediating processes on particular occasions of testing. Further, when there is an absence of both significant effects for susceptibility, and for susceptibility in interaction with state instruction, as found in Study 2, it is reasonable to infer that the processes mediating pseudomemory reports are not reliably hypnotic ones, or if they are present, their influence may be negated by the operation of other factors.[3]

The manipulation of context attempted in Study 2 reduced the incidence of pseudomemory in the experiment, but more so for structured than for free recall. This suggests that there is a cueing function associated with structured recall that is a relevant factor in explaining the pattern of pseudomemory effects that may be obtained. In Study 2, the effect on susceptible subjects operating under hypnotic instruction was relatively dramatic; for example, the incidence of pseudomemory for the mask dropped from 53.6% to 24.0% and the incidence of the phenomenon for the right-entry item dropped from 35.7% to 8.0%. In structured recall, the same degree of shift was not nearly as evident, the reduction for the right-entry item (for example) being only 3.3% from Study 1 to Study 2. Collectively, data suggest that the specific cues provided by the test stimuli served to reinforce continued pseudomemory responding, even though overt instructions were designed to reduce it.

Looking across the two studies, one of the most interesting aspects of the data was the maintenance of experiential versus behavioural effects. There are many examples in the hypnotic literature (for review, see Sheehan & McConkey, 1982) where behavioural effects do not parallel subjective outcomes. One of these findings is the set of data obtained by Jackson and Sheehan (1986) which demonstrates that low susceptible subjects, while not responding behaviourally to task suggestions, nevertheless show variability in their reported involvement from one suggestion task to another. Present data do not strictly allow one to infer whether the discrepancy between behavioural and experiential data found in the current work is real or artifactual. It could be that behavioural scores are ill-equipped to detect differences that genuinely occur across contexts, and so, more sensitive measures may be required to detect implicit effects. Alternatively, the observed discrepancy may be valid.

Communication effects can be subtle in their influence, and may require sensitive techniques, such as the EAT, to detect them. Results, overall, are provocatively consistent with two major theoretical propositions. The first is that pseudomemory is determined by multiple factors which incorporate hypnotic skill, contextual influences, and state instruction, these variables themselves interacting. One factor (e.g., skill) may operate to raise the incidence of pseudomemory in one environment (e.g., Study 1) — and perhaps more obviously for some subjects than for others — thus emerging as a major mediator of responses in particular settings. In another environment, however (e.g., where the transfer of effect is not as readily suggested or encouraged), other factors may come to the fore in their influence (e.g., state instruction). This view implies that any one factor is not sufficient for explaining the occurrence of hypnotically induced pseudomemory.

The second proposition is that pseudomemory might usefully be considered in some circumstances as an instance of posthypnotic response. This view is reinforced by the finding in Study 2 that state instruction was primarily influential when contextual boundaries were drawn more sharply and needed to be crossed. It is precisely the capacity to transcend contextual boundaries that characterises posthypnotic responding as such a compelling phenomenon. The relative strength of state instruction when boundaries were sharpened in Study 2 highlights the possible relevance of suggestion operating in a posthypnotic setting. Such a model of explanation is also consistent with the stronger indications of pseudomemory in Study 2 for structured (versus free) recall. It is in the stuctured recall situation, for example, that the signal to respond is clearer and, as Weitzenhoffer (1957) has noted, it is the "signal" for excitation that has a special role to play in explaining posthypnotic phenomena.

While these propositions offer a somewhat speculative account of the data that have been gathered, the implications of the evidence collected here suggest there is definite value in researching pseudomemory further by manipulating a wider range of contextual features. It appears to be the case that pseudomemory requires specific contextual support to maximise the frequency of its occurrence, and the ambiguity of the experimenter's communication is one important factor determining the pattern of effects that may be obtained.

REFERENCES

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis*. Cambridge, MA: MIT Press.

Fromm, E., & Shor, R. E. (Eds.). (1979). *Hypnosis: Developments in research and new perspectives*. Hawthorne, NY: Aldine.

Gordon, R. (1949). An investigation into some of the factors that favour the formation of stereotyped images. *British Journal of Psychology, 39*, 156–167.

Jackson, J., & Sheehan, P. W. (1986). Imaginative capacity and individual diferences in hypnotic responsivity. *Australian Journal of Clinical and Experimental Hypnosis, 14*. 139–152.

Jenkins, J. C., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *American Journal of Psychology, 35*, 605–612.

Labelle, L., & Perry, C. W. (1986). *Pseudomemory creation in hypnosis*. Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.

Labelle, L., Laurence, J. R., Nadon, R., & Perry, C. W. (1990). Hypnotisability, preference for an imagic cognitive style, and memory creation in hypnosis. *Journal of Abnormal Psychology, 99*, 222–228.

Laurence, J. R., & Perry, C. W., (1983). Hypnotically created memory among highly hypnotisable subjects. *Science, 222*, 523–524.

Lynn, S. J., Weekes, J. R., & Milano, M. (1989). Reality versus suggestion: Pseudomemory in hypnotisable and simulating subjects. *Journal of Abnormal Psychology, 98*, 137–144.

Mayes, A. (1983). The development and course of long-term memory. In A. Mayes (Ed), *Memory in animals and humans* (pp. 133–176). Wolmingham, England: Van Nostrand Reinhold.

McCann, T., & Sheehan, P. W. (1987). The breaching of pseudomemory under hypnotic instruction: Implications for original memory retrieval. *British Journal of Experimental and Clinical Hypnosis, 4*, 101–108.

McCann, T., & Sheehan, P. W. (1988). Hypnotically induced pseudomemories — sampling their conditions among hypnotizable subjects. *Journal of Personality and Social Psychology, 54*, 239–346.

McConkey, K. M., Labelle, L., Bibb, B. C., & Bryant, R. A. (1990). Hypnosis and suggested pseudomemory: The relevance of test context. *Australian Journal of Psychology, 42*, 197–206.

Orne, M. T. (1959). The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology, 58*, 277–299.

Orne, M. T. (1962). Hypnotically induced hallucinations. In L. J. West (Ed), *Hallucinations* (pp. 211–219). New York: Grune & Stratton.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

Sheehan, P. W. (1967). A shortened form of Betts' Questionnaire Upon Mental Imagery. *Journal of Clinical Psychology, 23*, 386–389.

Sheehan, P. W. (1980). Factors influencing rapport in hypnosis. *Journal of Abnormal Psychology, 89*, 263–281.

Sheehan, P. W., & McConkey, K. M. (1982). *Hypnosis and experience: The exploration of phenomena and process*. Hillsdale, NJ: Erlbaum.

Sheehan, P. W., Statham, D., & Jamieson, G. (in press). Pseudomemory effects and their relationship to level of susceptibility to hypnosis and state instruction. *Journal of Personality and Social Psychology*.

Shor, R. E., & Orne, E. O. (1962). *Harvard Group Scale of Hypnotic Susceptibility, Form A*. Palo Alto, CA: Consulting Psychologists Press.

Spanos, N. P., & McLean, J. (1986). Hypnotically created pseudomemories: Memory distortions or reporting biases? *British Journal of Experimental and Clinical Hypnosis, 3*, 155–159.

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology, 83*, 268–277.

Weitzenhoffer, A. M. (1957). Posthypnotic behaviour and the recall of the hypnotic suggestion. *Journal of Clinical and Experimental Hypnosis, 5*, 41–58.

Weitzenhoffer, A. M., & Hilgard, E. R. (1962). *Stanford Hypnotic Susceptibility Scale, Form C.* Palo Alto, CA: Consulting Psychologists Press.

Wessells, M. G. (1982). *Cognitive psychology.* New York: Harper & Row.

Yuille, J. C. (1982). *Video as a medium for eyewitness research (1).* Unpublished manuscript. University of British Columbia, Canada.

1 It is possible with this item for subjects to interpret "right entry" in two ways, "right" meaning different things depending on the perspective of the observer. Testing was made unambiguous in this study by asking subjects: "Watching the scene from where you sit now, did the robber enter from the right or the left?"

2 The design of this program of research permitted the detection of any responses by subjects to pseudomemory suggestion that were ambiguously tied to reports of what was seen on the videotape. No subject was included in Study 1 (or Study 2) who reported before pseudomemory suggestion was given that the robber wore a mask, swore, or entered from the right.

3 What alterations in procedure are precisely responsible for effects on subjects' posthypnotic responses is not certain. In order to achieve the treatment manipulation under test, changes needed to be made to both E1's communication during pseudomemory suggestion and E2's communication prior to testing for free and structured recall. Future research might usefully explore the exact contributions to the total set of results of the different stimulus features that were employed.

Australian Journal of Clinical and Experimental Hypnosis
Vol. 19, No. 1, 1991, pp. 19–30

LOGLINEAR ANALYSIS: A PRIMER FOR HYPNOSIS RESEARCHERS

Bruce W. Carlson, Joseph P. Green, and Steven Jay Lynn

Ohio University

This paper presents an introduction to loglinear analysis. Two examples from the hypnosis literature are used to demonstrate this statistical technique.

This paper is an introduction to loglinear analysis presented for researchers in hypnosis who have gathered categorical data, as frequently reported in the hypnosis literature and as evidenced by studies of trance logic, age-regression and simulation effects (see Spanos, 1986). Although chi-square analyses are by far the most widely reported, loglinear analysis has definite advantages over chi-square analysis. Unlike chi-square, loglinear analysis permits the examination of interactions among three or more variables. Also, whereas the chi-square analysis permits the examination of the marginal associations between two variables — that is, the association between the two variables after collapsing over all other variables — loglinear analysis permits the examination of the partial association between two variables; their association in the presence of other variables. To better appreciate the advantages of loglinear analysis, we will first present a discussion of bivariate chi-square analysis.

BIVARIATE CROSS-CLASSIFICATION TABLES: CHI-SQUARE TEST

Consider the bivariate cross-classification table shown in Table 1. These data were collected by Green et al. (1990). Our study was designed to test Erickson's (1980) assertion that literalism is a marker of the hypnotic state. To test literalism, subjects were asked six questions including, "Do you mind telling me your name?" Literalism was scored if subjects answered "yes" or "no" without further

An earlier version of this paper was presented at the Annual Meeting of The Society for Clinical and Experimental Hypnosis, St. Louis, November 2, 1989.

Requests for reprints should be sent to Bruce W. Carlson, Department of Psychology, Ohio University, Athens, OH 45701, U.S.A.

Table 1 Number of Individuals in Each Treatment Group who did and did not Give Literal Responses (Green et al., 1990)

Group	Literal response		Total
	Yes	No	
Hypnosis	2	10	12
Simulating	7	5	12
Library	3	26	29
Total	12	41	53

comment. This question was addressed to hypnotised and simulating subjects during hypnosis (Stanford Scale of Hypnotic Susceptibility; SSHS:C, Weitzenhoffer & Hilgard, 1962). Hypnotised subjects passed at least 11 suggestions on both the 12-item Harvard Group Scale of Hypnotic Susceptibility (HGSHS:A, Shor & Orne, 1962) and on the 12-item SSHS:C. Simulating subjects, who were instructed to simulate hypnosis, scored 3 or below on the HGSHS:A. A third group of subjects were tested informally outside the campus library.

The Pearson chi-square test is the traditional method for analysing bivariate cross-classification tables, such as Table 1. Its purpose is to determine whether two variables are independent of one another. In this example, it provides a test of the hypothesis that group membership and literalness of responses are independent.

Conceptually, the Pearson chi-square test statistic is a function of the difference between the observed cell frequencies from the cross-classification table and the expected cell frequencies under the null hypothesis that the two variables are independent. If the discrepancy between the observed and the expected cell frequencies is great, then we conclude that the independence hypothesis is untenable and that the variables are related. In this example, group and literalism are related, χ^2 $(2, n = 53) = 11.48, p < .05)$.

BIVARIATE CROSS-CLASSIFICATION TABLES: LOGLINEAR ANALYSIS

Now let us consider how these same data may be analysed using loglinear analysis. The aim of loglinear analysis is to construct a model with few parameters that fits the data well. To accomplish this goal it is necessary to identify potential models, estimate parameters, assess fit, select the best model, validate the model, and interpret the results.

Identifying Potential Models

The hypothesis that the variables are independent can be represented by the following loglinear model:

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y$$

where m_{ij} is the expected frequency for the cell corresponding to row i and column j, μ is a parameter representing an overall effect, and λ_i^X and λ_j^Y are parameters representing the effect of the row variable and the effect of the column variable, respectively. This model is often called the loglinear model of independence because it assumes that the row and column variables are independent of one another.

The above model is just one possible model that could describe the data in a bivariate cross-classification table. The simplest model supposes that the expected frequencies are a function of only the overall effect:

$$\log m_{ij} = \mu$$

That is, the expected frequencies are a function of neither the row nor the column variable. Alternatively, the expected frequencies might be a function of only the row variable:

$$\log m_{ij} = \mu + \lambda_i^X$$

Or, they might be a function of only the column variable:

$$\log m_{ij} = \mu + \lambda_j^Y$$

Finally, the expected frequencies might be a function not only of the row and column variables, as in the independence model, but also of the joint effects of the two variables. This loglinear model of dependence can be expressed as follows:

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

A model which contains all possible main effects and all possible interactions is called a *saturated model.*

Loglinear analyses are generally restricted to hierarchical models which, for a bivariate cross-classification table, consist of those models described above. An hierarchical loglinear model is a model in which the presence of an effect implies the presence of all lower order effects that can be constructed from the variables contained in the effect. For example, the presence of the XY interaction in a model demands the presence of both the X and Y effects in the model also. Note that this excludes certain models such as the following:

$$\log m_{ij} = \mu + \lambda_j^Y + \lambda_{ij}^{XY}$$

Because of their structure, hierarchical loglinear models can be denoted by the highest order interaction involving each of the variables. Table 2 lists the models for the present example. Note that [GL] represents the saturated model, [G][L] represents the independence model, [G] represents the row-effects model, and [L] represents the column-effects model.

Estimating Parameters

To anticipate our discussion, the fit of a loglinear model is based on a comparison of the observed and estimated expected cell frequencies, just as it is in a chi-square test. The observed cell frequencies are defined by the data. To compute the estimated expected cell frequencies, estimates of the parameters of a loglinear model must be obtained. Normally, maximum likelihood estimates of the parameters are obtained using one of several numerical algorithms. An understanding of these algorithms is not essential to an understanding of loglinear analysis. Since a detailed discussion of these methods is beyond the scope of this paper, the interested reader is referred to Feinberg (1977) and Agresti (1990).

Assessing Fit

Testing a model. Once the parameter estimates for a model have been computed, the expected frequencies under the model can be calculated. By comparing the expected frequencies to the observed frequencies, the fit of the model to the data can be assessed. This is accomplished by computing the likelihood-ratio chi-square statistic:

$$G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \log \left(\frac{n_{ij}}{\hat{m}_{ij}} \right)$$

where I and J are the number of rows and columns of the cross-classification table, respectively, n_{ij} is the observed number of observations in each cell, and $\hat{m}_{ij}$ is the expected number of observations in each cell. The likelihood-ratio chi-square statistic, like the Pearson chi-square statistic, has an asymptotic chi-square distribution. Further, both statistics have the same number of degrees of freedom:

$$df = \# \text{ of cells} - \# \text{ of parameters.}$$

Generally, unless a cross-classification table contains many cells with small expected frequencies, the two chi-square statistics will give comparable results. For the example in Table 1, $\chi^2(2, n = 53) = 11.48$ and $G^2(2, n = 53) = 10.30$ for the independence model. Both statistical tests lead to the conclusion that the independence model does not fit the observed frequencies well. The likelihood-ratio chi-square statistics and the degrees of freedom for the set of models described earlier are contained in Table 2.

Testing an effect. It is possible to test the significance of individual effects as well as complete models. Consider two nested models, that is, two models such that one contains all of the effects or parameters that are in the other. For example, the independence model is nested within the saturated model

Table 2 Loglinear Models to fit to the Green et al. (1990) Data

Model	df	G^2	*p*
[GL]	0	0.00	1.00
[G][L]	2	10.30	.01
[G]	3	27.07	.00
[L]	4	20.48	.00

(i.e., every effect in the independence model is also in the saturated model). The difference in the chi-square statistics of two nested models provides a test of the statistical significance of the effects that appear in the more complex model but not in the simpler model. This difference will be distributed as a chi-square statistic, and it will have degrees of freedom equal to the difference in the degrees of freedom of the two models. For example, if the chi-square statistic for the independence model is subtracted from the chi-square statistic for the row-effect model, the resulting difference can be used to test the significance of the effect representing the column variable. The degrees of freedom for this test will equal the difference in the degrees of freedom of the two models.

Selecting the Best Model

The aim of loglinear analysis is to find the best-fitting model. A significant chi-square test means that a model fits the data poorly, that is, the observed and expected frequencies differ. Thus, the data analyst searches for models that produce nonsignificant chi-square tests. In the present example, choice of the best fitting model is easy. The saturated model appears to be the only model that fits the data well (see Table 2). Thus, we conclude that group membership is related to literal responding. While we now know that the groups differ with respect to literal responding, we don't know how they differ. In an analysis of variance, a significant effect is often followed by a series of contrasts to more closely examine differences among the treatment means. Contrasts may also be used in loglinear analysis.

Contrasts. As in the analysis of variance, $k - 1$ orthogonal contrasts can be formed for a variable that has k levels or, in other words, a contrast can be formed for each degree of freedom associated with any effect. An examination of the data in Table 1 suggests that the responses of the hypnosis and the library group are similar and that both differ from the simulating group. This suggests a pair of orthogonal contrasts: (a) compare the relative frequency of literal responding in the hypnosis and library groups; (b) compare the relative frequency of literal responding in the combined hypnosis and library groups to the relative frequency of literal responding in the simulating group. Each of these contrasts corresponds to a parameter in a loglinear model. Because we expect the hypnosis and the library groups to be similar, we should be able to omit the parameter representing the first contrast. If all of this seems unduly complicated, it should be noted that the standard statistical software packages easily permit the construction of contrasts in loglinear analysis. The model that contains the second contrast only fits the data well, $G^2(1, n = 53) = .30, p > .05$. Thus, we conclude that the response of simulating subjects differs from the response of the hypnosis and library subjects.

Validating a Model

Validation of the best fitting model should be part of any loglinear analysis, just as it should be part of any linear regression analysis. While a detailed discussion of validation techniques is not possible here, it is essential that residual or outlier analyses be conducted and that cross-validation procedures be done (e.g., Agresti, 1990).

Interpreting the Parameters of the Best Fitting Model

In an analysis of regression, a statistical test of a regression coefficient can indicate whether or not a predictor variable is related to a response variable, but it does not indicate the strength of that relationship. To accomplish the latter, a measure of the proportion of variance accounted for is often computed. A similar problem exists in loglinear analysis. The likelihood-ratio chi-square test can indicate whether an effect is statistically significant (e.g., whether a set of variables are related) but it does not provide a measure of the strength of the relationship between those variables.

A commonly used measure of the strength of the relationship between two categorical variables is the odds ratio. Consider the bivariate cross-classification table in Table 3. The odds ratio is defined as follows:

$$\hat{\theta} = \frac{\hat{m}_{11}/\hat{m}_{12}}{\hat{m}_{21}/\hat{m}_{22}} = \frac{\hat{m}_{11}\hat{m}_{22}}{\hat{m}_{12}\hat{m}_{21}}$$

Table 3 2 × 2 Cross-Classification Table

	Column	
Row	1	2
1	$\hat{m}_{11}$	$\hat{m}_{12}$
2	$\hat{m}_{21}$	$\hat{m}_{22}$

Table 4 Expected Cell Frequencies for the Best Fitting Loglinear Model for the Green et al. (1990) Data

Group	Literal response		Total
	Yes	No	
Hypnosis	1.46	10.54	12
Simulating	7.00	5.00	12
Library	3.54	25.46	29
Total	12	41	53

In other words, the odds ratio reflects how many times more likely it is that an observation in the first row falls in the first column than it is that an observation in the second row falls in the first column.

The estimated expected cell frequencies for our best fitting model are displayed in Table 4. Given the nature of the model we fit in our example, it seems reasonable to report the odds ratio for a literal response by the hypnosis group relative to the simulating group, and by the library group relative to the simulating group. Because of the nature of the contrast we formed, these two odd ratios must be equal. More specifically:

$$\hat{\theta} = \frac{1.46 \times 5.00}{10.54 \times 7.00}$$

$$= .10$$

That is, hypnotised subjects are .1 as likely to report literal response as simulating subjects are. Similarly, it can also be shown that library subjects are .1 as likely to report literal responses as simulating subjects are.

MULTIVARIATE CROSS-CLASSIFICATION TABLES: LOGLINEAR ANALYSIS

Consider the multivariate cross-classification table shown in Table 5. These data were collected in a study conducted by Lynn, Weekes, and Milano (1989). The purpose of this study was to examine the pseudomemory phenomenon in hypnosis. The design was a $2 \times 2 \times 2$ factorial design with three between-subjects factors: group (whether subjects were assigned to a hypnosis or a simulating group), telephone (whether or not a telephone actually rang during the experimental session), and suggestion (whether or not a suggestion was made that a telephone rang during the experimental session). For the purposes

Table 5 Observed Cell Frequencies for the Cross-Classification Table Formed from the Data Reported by Lynn et al. (1989)

Group	Telephone	Suggestion	Reported ring Yes	Reported ring No	Total
Hypnosis	Ring	Ring	9	1	10
		No ring	11	1	12
	No Ring	Ring	0	16	16
		No ring	1	17	18
Simulating	Ring	Ring	9	0	9
		No ring	12	2	14
	No Ring	Ring	1	11	12
		No ring	4	5	9
Total			47	53	100

of this paper, subjects were grouped into one of two categories depending on whether or not they reported hearing a telephone ring during the experimental session. Thus, these data can be represented in a 2 × 2 × 2 × 2 cross-classification table.

At the outset, it should be recognised that many of the expected cell frequencies will be small, perhaps too small for the analysis, although the guidelines offered by Wickens (1989) suggest that this may not be a problem. Usually, when a cell in a cross-classification table contains no observations, a small amount — typically .5 — is added to each cell in the table. This is necessary because log 0 is undefined. The standard statistical packages permit the addition of small values to the frequencies in a cross-classification table, so this need not be a concern of the data analyst.

Identifying Potential Models

To conceptualise the set of possible models, an analyst must answer the following question: Can a distinction be drawn between predictor variables and response variables? If not, the analyst will be interested in the relationship among all variables. If so, the analyst will be interested in the relationship of the predictor variables with the response variables. Because relationships among the predictor variables will not be of interest, the set of potential models can be restricted to those in which all possible interactions among the predictor variables are included in the model, even though some of them may be nonsignificant.[1]

In the present example, whether or not subjects reported hearing a phone ring can be thought of as a response variable, and the remaining variables — group, telephone, and suggestion — can be thought of as predictor variables. Thus, the potential models all contain the group-by-telephone-by-suggestion interaction (see Table 6).

Table 6 Loglinear Models for Lynn et al. (1989) Data

Number	Model	df	G^2	*p*
M1	[GTSR]	0	0.00	1.00
M2	[GTS][GTR][GSR][TSR]	1	0.56	.46
M3	[GTS][GTR][GSR]	2	5.99	.05
M4	[GTS][GTR][TSR]	2	1.74	.42
M5	[GTS][GSR][TSR]	2	4.16	.12
M6	[GTS][GTR][SR]	3	5.95	.11
M7	[GTS][GSR][TR]	3	8.52	.04
M8	[GTS][GSR][GR]	3	4.72	.19
M9	[GTS][GTR]	4	7.23	.12
M10	[GTS][GSR]	4	79.41	.00
M11	[GTS][TSR]	4	9.09	.06
M12	[GTS][GR][PR][SR]	4	8.59	.07
M13	[GTS][GR][TR]	5	9.59	.09
M14	[GTS][GR][SR]	5	80.09	.00
M15	[GTS][TR][SR]	5	12.24	.03
M16	[GTS][PR]	6	13.06	.04
M17	[GTS][GR]	6	81.79	.00
M18	[GTS][SR]	6	84.88	.00
M19	[GTS][R]	7	86.43	.00

Note G = Group; T = Telephone; S = Suggestion; R = Response.

Selecting the Best Model

Table 6 contains the likelihood-ratio chi-square statistics, the degrees of freedom, and the significance levels for the 19 loglinear models that were considered in this analysis. An examination of this table suggests that 10 models may fit the data: M1, M2, M4, M5, M6, M8, M9, M11, M12, M13. From among these models, we must select the best one.

Unfortunately, there is no single criterion that will guarantee selection of the best fitting model. Benedetti and Brown (1978) discuss a number of alternative criteria. The criterion we will use is described by Feinberg (1977). Starting with the saturated model, we will delete effects from the model, one at a time. At each stage, we will test the significance of the deleted effect and the significance of the full model. When one of these reaches significance, including a poor fit, we will select the model from the previous stage.

The model-selection steps for this analysis are summarised in Table 7. Rejecting Model 1, the saturated model, in favour of Model 2 is straightforward. Let us consider how we might proceed next. Model 2 contains all possible three-variable interactions. Is it possible to find a model that eliminates one of these interactions, yet still fits the data well? The answer is yes. There are actually two such models, M4 and M5. From an examination of the likelihood-ratio chi-square test statistics for both models, we select M4, because

Table 7 Tests of Effects and Models for Lynn et al. (1989) Data

	Effect test		Model test		
Effect	df	G^2	Model	df	G^2
G × T × S × R	1	0.56	2	1	0.56
G × S × R	1	1.18	4	2	1.74
G × T × R	1	2.98	8	3	4.72
T × S × R	1	3.87*	12	4	8.59
G × R	1	4.37*	11	5	9.09

* $p < .05$.

it provides the best fit to the data. Model 4 was constructed from Model 2 by eliminating the group-by-suggestion-by-response interaction. So, even though Model 4 fits the data well, we need to ask, did eliminating this interaction significantly worsen the fit of the model to the data relative to Model 2? The difference in the likelihood-ratio chi-square statistics of models M2 and M4 provides a test of this three-variable interaction. That is:

$$G^2\,(M4) - G^2\,(M2) = 1.74 - 0.56 = 1.18$$

The degrees of freedom for this chi-square test statistic are the following:

$$df\,(M4) - df\,(M2) = 2 - 1 = 1$$

Since the critical value for a chi-square test statistic with one degree of freedom is 3.84, we can conclude that the three-variable interaction can be eliminated from the model. In other words, deleting the three-variable interaction does not significantly worsen the fit of the model to the data. Following this procedure, we can then compare models M4 and M8 to test the group-by-telephone-by-response interaction, and so on. Following this procedure leads us to conclude that the best fitting model is M8.

The procedure that we just described is simple if the number of potential models is small. However, when a cross-classification table consists of several variables, the number of potential models can be very large, particularly if no distinction is drawn between response and predictor variables. Application of this procedure may not be feasible. Fortunately, many statistical software packages contain stepwise selection methods, similar to those available for regression analyses, to assist in screening models.

Contrasts. Because every variable in the present example has only two levels, there is only one degree of freedom and hence one contrast associated with each effect. Thus, no post-hoc tests need to be performed.