MEMORY, 1998, *6* (3), 257–275

The Relation of Output Order and Commission Errors in Free Recall and Eyewitness Accounts

Bennett L. Schwartz, Ronald P. Fisher, and Kellye S. Hebert

Florida International University, USA

We explored the relation between output order and the likelihood of a commission error in free recall under both laboratory and eyewitness conditions. In Experiment 1, participants studied a list of 20 unrelated words and, after a five-minute distractor task, were asked to recall those words. Whereas the items that participants recalled were mostly correct, commission errors were more likely to occur at the end of a participant's output. In Experiment 2, participants viewed a police film depicting an armed robbery. Participants described the perpetrators, their truck, and the sequence of events during the robbery. When describing the perpetrators or the truck, commission errors were more likely to occur at the end of the output. However, when describing the sequence of events, commission errors were more likely to occur in the middle of the output. In Experiment 3, we replicated the finding that commission errors are likely to occur at the end of the output order when participants are describing people. We speculate on the potential application of this finding and its theoretical underpinnings.

INTRODUCTION

False memories have recently become a major focus in memory research (see Lindsay & Read, 1994; Loftus & Hoffman, 1989; McDermott, 1996; Roediger, 1996; Roediger & McDermott, 1995; Smith, 1995). The study of when, where, and why false memories are likely to occur is important in applications of memory research such as eyewitness identification. In eyewitness situations, critical aspects of police investigation may often hinge on veridical information from eyewitnesses. Therefore, it is necessary to distinguish when an individual's recollection is veridical and when it is not.

Requests for reprints should be sent to Dr. Bennett L. Schwartz, Department of Psychology, College of Arts and Sciences, Florida International University, Miami FL. 33199, USA. Email: Schwartb@servax.fiu.edu

We thank the Metro-Dade Police Department for providing us with the video used in Experiment 2. We thank Margaret Funnell and Dartmouth College for providing the video used in Experiment 3. We thank the following people for assistance with data collection, analysis, and useful discussion: Jacqueline Corcho, Gaye Huxoll, Iris Mizrahi, and the members of FIU cognition group. We thank J. Richard Hanley, Janet Fraser Parker, Leslie Frazier, and one anonymous reviewer for helpful commentary on an earlier version of this manuscript.

Our analysis of false memories distinguishes between two types of memory failure, errors of omission and errors of commission (see Krinsky & Nelson, 1985). Omission errors occur when people fail to report events that they observed, whereas commission errors occur when people report events that did not, in fact, occur. Historically, memory researchers have focused on errors of omission, or their mirror image, the amount of input that was remembered correctly. More recently, a complementary focus on errors of commission has developed to examine the accuracy of recollection, the degree to which the output correspond to earlier experienced events (see Koriat & Goldsmith, 1994, 1996; Roediger, 1996). The focus here is on errors of commission, and specifically, how we might increase the accuracy of recollection by identifying such errors of commission.

One reason why errors of commission are so interesting is that under natural conditions they occurr infrequently. When subjects have control over what they report, as in a standard free recall task, errors of commission, or intrusions, typically account for less than 10% of all responses (e.g. Dunning & Stern, 1992; Scrivner & Safer, 1988). Even in studies using more realistic materials, as in recall of simulated crime scenes, responses to open-ended questions (e.g. ''Describe what happened'') are highly accurate (less than 10% intrusions) (see Fisher, 1995).

The key to such high accuracy seems to be giving people control over what they report. Koriat and Goldsmith (1994) suggested that people can accurately monitor the veracity of their recollections, thereby enabling them to screen out incorrect responses. When these monitoring criteria are relaxed in response to external pressure, e.g. instructions to guess, more commission errors should occur. For example, in Roediger, Challis, and Wheeler (reported in Roediger, Wheeler, & Rajaram, 1993; see also Roediger & Payne, 1985) subjects saw a series of 60 pictures and then free recalled as many picture names as they could. On the average, they recalled 35.9 pictures, and with few intrusions (0.8). Another group of subjects (forced recall) was requested to make a fixed number of responses (60), guessing if necessary to complete the list. This instruction produced no more correct responses (35.6) than the control group; however, it elicited many more intrusions (5.0).

Whereas it is relatively easy to increase the number of commission errors beyond the level that subjects generate in unforced recall—admittedly, not a terribly useful skill—it is more difficult to increase the accuracy of reports. We know of no successful experimental efforts to decrease the number of commission errors. Koriat and Goldsmith (1994) gave subjects monetary rewards for correct answers and took away money for commission errors. When the penalties for commission errors were severe, there was a slight increase in accuracy (decrease in commission errors), but also a reduction in the number of correct responses. That is, subjects traded-off a considerable amount of information for minimal gains in accuracy, suggesting that they could not

discriminate between correct recollection and errors of commission (see also Koriat, 1995).

If people cannot evaluate the accuracy of their own recollections, how then might we, as external agents, increase the accuracy of recollection? One possible approach is to allow respondents to make incorrect responses—although obviously not to encourage them—and then to identify these errors after the fact so that they may be excluded from the corpus of recollections (see Leippe, 1994; Schooler, Gerhard, & Loftus, 1986; and Smith, 1995, for a similar approach). To accomplish this we need to identify properties that distinguish between correct and incorrect recollections. Several promising candidates have been identified elsewhere (Fisher, 1995; Fisher & Schwartz, 1995). We focus here on using output order to discriminate between commission errors and correct recollections. By output order, we mean the temporal order with which participants report items from their memories.

Several recent studies address the relation between commission errors and output order. For example, Roediger and McDermott (1995) attempted to induce false recollections in an immediate free recall task by presenting associates of a particular core word but not the core word itself. For instance, subjects were shown the words *stream*, *brook*, *fish*, *lake*, and *bridge*, but not the core word *river*. The participants often incorrectly recalled *river* along with the other associates that were presented. When this intrusion error occurred, it generally was located at the end of the participant's response protocol, within the last fifth of the items output. Indeed, 63% of these related commission errors occurred in the last quintile. Using the same stimuli, McDermott (1996) replicated this effect, although the effect was much smaller (i.e. only 38% of the errors occurred in the last quintile). Payne, Elie, Blackwell, and Neuschatz (1996) also reported this relation when using Roediger and McDermott's stimuli.

Although it is possible that the correlation between commission errors and output order was related to Roediger and McDermott's use of categorised lists and immediate recall (see Payne et al., 1996), we suspect that it is more likely a general feature of performance in a free recall task, and perhaps indicative of a general feature of any unconstrained memory report. We therefore examined here the output position of commission errors in a free recall task with unrelated items. Following this traditional verbal learning study, we examined the output position of commission errors in two studies that used more real-world stimuli (a videotape of a simulated crime and a videotape of a neuropsychological interview).

EXPERIMENT 1

Method

Participants. The participants were 55 introductory psychology students at Florida International University who received course credit. Participants were tested individually in half-hour sessions.

Materials. Some of the participants were tested with Macintosh computers with stimulus words presented in black against a white background. Other participants were tested with DEC PC computers with stimulus words presented in white against a black background. The difference in presentation mode did not affect recall.

Each participant read a list of 20 words chosen randomly from a master list of 252 English words. Therefore, each participant received a unique set of 20 unrelated words. Mean frequency of occurerence in the English language was 66.0 according to the Kuçera and Frances (1967) norms. The words were five to eight letters long, all with two syllables, and were derived from the Toronto word pool (Murdock, 1968).

Procedure. The participants were informed that they would see a list of 20 English words, and that after five minutes of another activity, they would be given a free recall test. They were instructed to make judgements of learning in anticipation of free recall on each word. Judgements of learning were defined as a prediction of the likelihood of recalling that item later (the judgements of learning will not be considered here). Each word appeared on the screen for four seconds. After four seconds of study, a prompt was shown, and participants made their judgements of learning. Participants were allowed five seconds to make the judgement of learning. Then the screen went blank, and a one-second pause occurred before the next word was presented. Following initial list presentation, participants were given a five-minute mental-artithmetic task. Finally, participants recalled as many of the original words as possible. They were given as much time as they wanted to type their responses on a computer keyboard. When they thought they could not recall any more words, they typed in a question mark, and that ended the experiment. Participants were then fully debriefed and thanked for their participation.

The data were scored in the following way. First, obviously misspelled words were counted as correct words (two independent observers verified the misspelled word). If the two observers did not agree, the word was counted as an incorrect response.

We assigned items to quartile in the following manner. We added each participant's correct responses and commission errors (total output). Then we looked at the output position of each commission error among that participant's total output. We divided the position of the error by the total output. If this quotient was between 0 and .249, we assigned it to the first quartile, between .250 and .499, we assigned it to the second quartile, between .500 and .749, we assigned it to the third quartile, and between .750 and 1.0, we assigned it to the fourth quartile. We then tallied the number of errors for each participant in each quartile.

Results

Recall Performance. Overall, participants reported a mean of 8.7 words out of the total of 20. Of these 8.7 words, 7.8 were correct responses for a 89.8% accuracy rate, consistent with the previously reported data from a variety of studies using free report procedures (see Dunning & Stern, 1992). A standard serial position effect was observed with words at the beginning and end of the list best remembered. Statistically this was evaluated by collapsing over quintile, which showed highest recall in the first quintile and the last quintile, $F(4,54) = 2.48$, $MS_e = 0.20$.

Output Order. Each participant's responses were divided into quartiles (all participants made at least four responses). An analysis was performed in which the dependent variable was the proportion of words that a participant recalled that were commission errors in each of the four quartiles. A larger number of commission errors were reported in the fourth quartile. (.60 errors or 26% of responses) than for the first three quartiles (.09 errors or 3% of responses). F (3, 162) = 18.2, $MS_e = 0.20$ (see Fig. 1). Although more errors occurred in the third quartile than in either the first or second, this was not statistically significant. The mean position of a commission error was 6.9 (out of the average 8.7 responses). Interestingly, 20 of the 55 participants (36%) made a commission error on their last response.

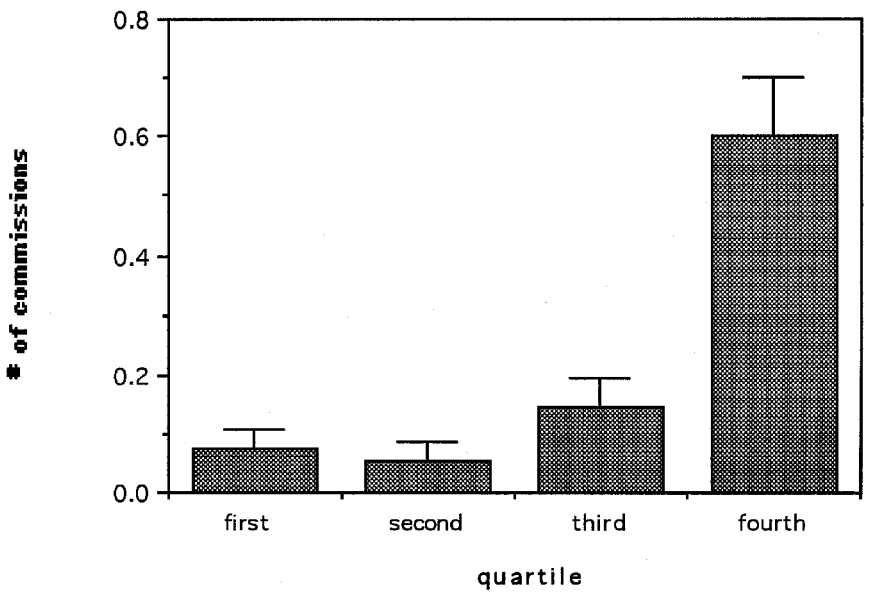

FIG. 1. Commission errors as a function of output order in Experiment 1.

Discussion

The present results replicate the finding of Roediger and McDermott (1995) that commission errors in free recall are most likely to occur towards the end of the participant's report. The effect was quite dramatic. Indeed, in terms of percent correct, participants were correct on 97% of the items reported in the first half of the list (quartiles 1 and 2), but were correct on only 74% of items in the last quartile. The absolute level of commission errors here was considerably less than that found in Roediger and McDermott, but this probably reflects the fact that they deliberately selected stimuli that would induce commission errors whereas we did not. Regardless of the difference in absolute error rates, both studies showed a high correlation between output order and accuracy, suggesting that the relation is a fundamental property of memory, at least for uninfluenced free report.

EXPERIMENT 2

Our goal in Experiment 2 was an applied one. If the correlation between output order and accuracy is a fundamental property of memory in free report mode, then we should see the same relation in memory situations other than a list-learning experiment. Therefore, in Experiment 2, we attempted to study this phenomenon with a different kind of stimulus. Would commission errors be more likely to occur at the end of the output order in an ecologically relevant situation?

In Experiment 2, participants were exposed to a one-minute film (courtesy of the Metro Dade Police Department) depicting an armed robbery. In the film, two perpetrators, one man and woman, approach two men on a street corner. They steal a camera, a pair of binoculars, and a briefcase and then flee in a small pickup truck. The two victims give chase, but to no avail. After viewing the video once, we asked participants to describe the male perpetrator, the female perpetrator, and the getaway truck. We also asked the participants to describe the sequence of events in the video.

There are two important differences between Experiment 1 and Experiment 2. First, we exposed participants to an ecologically relevant stimulus in Experiment 2. Many people will witness or be a victim of a crime in their lifetime and may have to report information to the police. Second, we asked participants both to provide descriptions of people and the vehicle (recalled descriptions) and to recall the sequence of events (recalled sequence). In Experiment 1, participants were not asked to recall sequentially.

The difference between recalled descriptions and recalled sequence may be an interesting one with respect to output order. With recalled descriptions, there is not an immediately obvious pattern that people will use to recall information. Nor are there specific retrieval cues that all participants will use. For example, we think that there is no *a priori* reason why people should report the

perpetrator's hair colour before they report the perpetrator's shirt colour. Thus, there are few intrinsic constraints on the order in which participants report descriptive information from memory. With recalled sequence information, however, there is a clear intrinsic constraint, namely the time in which events occurred during the video. Participants are required to use the serial order of events in the video to guide their memory report.

We suspected that recalled descriptions would lead to a similar pattern to the one observed in Experiment 1 because recall is not constrained. Because there is no set manner in which to retrieve this information, people will use similar strategies to those that they use in free recall of a word list. Therefore, we think the effect of Experiment 1 and Roediger and McDermott (1995) will be generaliseable to the descriptions of people and objects. However, recalled sequence-information is not an unconstrained report. The order in which events occur serves as a natural retrieval cue for those events. Therefore, it is possible that sequentially driven recall will not result in errors occurring late. Indeed, as will become apparent, the results support this hypothesis.

Method

Participants. The participants were 119 undergraduate psychology students at Florida International University who received nominal course credit. Of these participants, 13 failed to give a description of one of the people or the truck, and 8 gave descriptions that we could not code because they wrote paragraphs about the targets rather than listing their descriptions. This left 98 participants who answered all of the questions in the manner instructed. All of the analyses were conducted on these 98 participants.

Materials. We showed a Metro-Dade Police Department training video, as described next.

Procedure. We conducted the experiment in three separate undergraduate classes. Each class received exactly the same instructions, and the experiment was conducted in an identical manner. The participants were asked to volunteer for an experiment concerning memory retrieval. After obtaining consent from all participants, the experimenter set up and played the video. Participants watched the video, and although they expected a memory test, they did not know what form the memory test would take.

The film showed two men, possibly tourists, standing on a street corner, preparing to take a photograph. A small pickup truck pulled up next to them. A man and a woman got out of the truck, both carrying revolvers. The criminals took a camera, a pair of binoculars, and a briefcase from the two men. They ran back to their truck. The male criminal entered on the driver's side, the female

criminal on the passenger side. The two criminals then drove away in the truck while the two victims chased for a few seconds on foot, but to no avail.

The audio track of the film was turned off during the showing of the film. We did not include sound because the two victims of the crime, and later, the criminals, were speaking in Spanish, and not all of our participants were bilingual. Although many participants identified both victims and one of the criminals as Hispanic, no participant reported being able to lip-read what the people in the video were saying in Spanish before the criminals approached, nor any of the dialogue between the criminals and the victims.

Immediately after viewing the film, the experimenter handed out the retrieval packs. Each packet contained instructions for the experiment and four questions. When all of the participants received their packet, the experimenter gave the following instructions:

> I would now like for you to imagine that you are the sole witness, being interviewed by a police officer who will use your report as the basis for his or her investigation. Please reach each question carefully and answer as thoroughly as you can.

The questions in the packet were counterbalanced in the following manner. Half of the participants described the sequence of events first, and half described a criminal or the truck first. Therefore, participants either described the events first or last. For the descriptions of the criminals and the truck, the order was fully counterbalanced across participants. Therefore, for those participants who recalled the sequence of events last, one third described the male criminal first, one third described the female criminal first, and one third described the getaway truck first.

Participants were instructed how to report each description. The participants were asked not to answer in narrative form or complete sentences. Rather they were asked, in an effort to ''save time'', to ''list the basic facts, listing one fact per line''. They were given an example of what we expected this to look like. A majority of participants (82%) followed these instructions and were included in the analyses.

Participants then gave their description. They were not allowed to return to an earlier description once they had turned the page and begun answering the next one. The recall phase was self-paced as participants could spend as long as they wished on each description. Moreover, they could report as much or as little as they wished to. When all the participants in each session had completed the packet of questions, they were debriefed and thanked for their participation.

The data were scored in the following way. Two observers went through all of the responses and scored them as correct or incorrect and noted the output position of the incorrect ones. The two observers then went through each participant's packet together and resolved any discrepancies. The two observers

agreed ahead of time on how to score ambiguous responses or responses that were judgements. For example, assessments of ''being overweight'' were considered correct for the male perpetrator, but not the female perpetrator.

We assigned items to quartile in an identical manner to Experiment 1 except when participants made fewer than four responses. Unlike Experiment 1, participants here frequently made fewer than four responses. Using the aforementioned method, an item that was output first, but was only one of three responses would be considered to be in the second quartile (as the quotient would be .33). This scoring, however, is counterintuitive. Therefore, in these cases, the experimenters alternated in the assignment of the item to either the obvious quartile or the quartile as determined by the quotient. If only two items were reported, alternation occurred between quartile 1 and 2, and between 3 and 4. This process was most needed in the question concerning the female perpetrator because many participants reported fewer than four items concerning her. We also scored the data using only the quartile determined by the quotient. We also scored the data without using those participants who reported fewer than four items for any particular item. Neither alternative scoring method affected the general pattern of the results.

Results

Recall Performance. We conducted an analysis of variance on the total amount of information reported in each of the four experimental conditions (i.e. question asked), demonstrating significant differences across the conditions, $F(3,267) = 84.65$, $MS_e = 2.91$ (see Table 1). Post-hoc tests showed that participants reported more in response to the question to describe the sequence of events (7.1 items) than they did to any one of the questions concerning the criminals or the truck (5.0). Post hoc tests also indicated that participants reported more information about the male perpetrator (6.1) than they did about the female perpetrator (3.9) or the truck (4.1).

We also conducted an analysis of variance on the accuracy of participants for each of the questions (see Table 1). We found a significant difference among conditions, $F(3,267) = 32.17$, $MS_e = 0.04$. Post-hoc tests indicated that

TABLE 1
Total Amount Reported and Percent Correct as a Function of Type of Question

Question Type	*Total Reported*	*Percent Correct*
Male Perpetrator	6.1	.79
Female Perpetrator	3.9	.62
Getaway truck	4.1	.85
Sequence of events	7.1	.89

participants were less accurate in their report of the female perpetrator than they were for other kinds of information. Post-hoc tests also indicated that the accuracy for the sequence of events was higher than the accuracy for the male perpetrator.

Output Order. Each participant's responses were divided into quartiles for each question. In the output order analysis, we treated the output of recall of descriptions and recall of sequential information separately. We did this so that we could look for two main effects rather than a complex interaction of type of information with output order. We report the analysis of variance for recall of descriptions first. The analysis of was done on the likelihood of a participant making a commission error for each of the four quartiles for each of the three questions that concerned recall of descriptions (i.e. of the two criminals and the getaway truck). First, there was a significant difference in the numbers of errors for each question, F (2, 194) = 15.00, $MS_e = 0.24$. Post-hoc tests indicated a lower number of errors made in describing the truck. More importantly, fewer commission errors were reported in the first quartile (.16) than for the last three quartiles (.31), F (3,291) = 8.18, $MS_e = 0.22$ (see Fig. 2). The second, third, and fourth quartile were not significantly different from each other. The mean position of errors was 2.6 out of the mean total of 4.7 for any particular

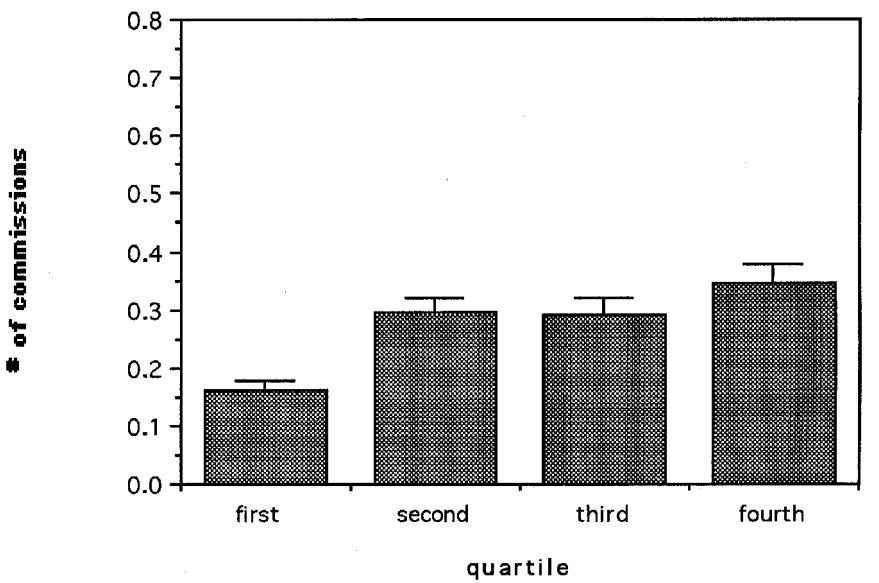

FIG. 2. Commission errors as a function of output order for recall of descriptions in Experiment 2.

description. The interaction between kind of descriptive question and output order was not statistically significant ($F < 1$).

The relation between output order and commission errors differed between the description of sequential information and the description of people and objects. A statistical analysis was also done on the relation between output order and commission errors for recall of the sequence of events. Quartiles were determined in the same manner as for recall of the perpetrators and the truck. We found that there was a significant difference in the number of errors reported as a function of quartile, $F\ (3,291) = 12.16$, $MS_e = 0.19$. Post-hoc tests indicated that more errors were observed in the second and third quartile than were observed in the first and fourth quartile (see Fig. 3). The mean position of commission errors was 4.2 out of a mean total of 7.1 items.

Discussion

In both experiments we found that there was a significant relation between output position in free recall and the likelihood of a commission error. When no constraints were placed on the order of recall, the likelihood of an error increased monotonically as a function of output position. This was observed for recall of the list words (Experiment 1), and for the recall of descriptions (Experiment 2). However, when some constraints were placed on the free recall

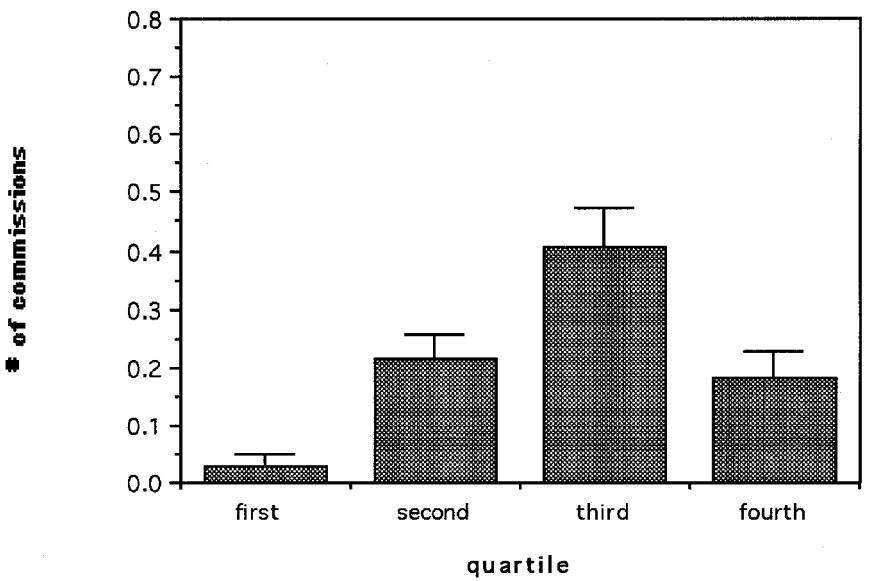

FIG. 3. Commission errors as a function of output order for recall of events in Experiment 2.

procedure, as in the recall of the sequence of events (Experiment 2), the relation between output order and commission errors changed. When participants were required to report the *sequence* of events seen in the film, then errors were most likely to be seen in the middle of a person's report.

Interestingly, when we asked participants to report the sequence of events in the film, errors accumulated in the middle of the reports. The pattern of data in Fig. 3 looks like a serial position curve, with lower errors in the primacy and recency portions of the curve. Perhaps participants lower their criterion because they expect memory to be worse for events in the middle. However, in the crime they witnessed on the video, the most salient events, and hence the most memorable occurred in the middle (the actual hold-up).

Experiment 2, like Experiment 1, demonstrated an association between the likelihood of a commission error and output order. However, in Experiment 1, the significant increase in reported errors occurred between the third and fourth quartile, whereas in Experiment 2 it occurred between the first and second quartile. Although the direction of the association is the same, the timing is different. What can account for this difference? There were two major differences between the experiments that may shed some light on the observed differences in output order. First, in Experiment 1, participants output nearly 10 items, whereas in Experiment 2, participants output 5 items per description. Therefore, the difference may have had to do with the number of items reported. Second, in Experiment 1, we used laboratory-based word lists, and Experiment 2, we used an eyewitness crime scene. This difference may also have created differences in the way people reported items.

EXPERIMENT 3

In Experiment 2, we used only one video as the stimulus. Therefore, we are unsure whether the results we observed in that experiment were a function of the particular video used, or a general characteristic of memory. Therefore, in Experiment 3, we used a different video, both in content and in the people shown. The video depicts an interview between a neuropsychologist and an aphasic patient. The setting was indoors, and both people were calmly seated throughout the interview. Thus, this video allowed us to examine the commission error/output order relation under a different set of circumstances than Experiment 2. The patient's aphasia is quite acute, and thus the video was a highly interesting stimulus for psychology students. Indeed, many of the students who participated in the experiment wanted to see the rest of the video after the experiment was over.

Participants were asked to describe both the psychologist (female) and the patient (male). However, unlike Experiment 2, they were not asked to describe any sequential information because the conversation focused on demonstrating the extent of the patient's aphasia, and did not lend itself to sequential

descriptions. Therefore, we hypothesised that we would see a monotonic increase in the number of errors as a function of output quartile.

Method

Participants. The participants were 68 undergraduate psychology students at Florida International University who received nominal course credit. One subject was excluded from the analyses for failing to follow directions.

Materials. We showed a video of a neuropsychologist interviewing a patient with acute anomic aphasia.

Procedure. We conducted the experiment as a group in an undergraduate psychology class. The participants were asked to volunteer for an experiment concerning memory retrieval. After obtaining consent from all participants, the experimenter set up and played the video. Participants watched the video, and although they expected a memory test, they did not know what form the memory test would take.

The film showed a dark-haired female psychologist and an older white-haired male patient, seated in front of the camera. The psychologist asked the man a number of questions about his past employment, education, and residences. The patient tried his best to answer the questions but had severe difficulties finding the correct word to use (anomia). The participants were shown three minutes of the video. The sound was left on, so the participants could appreciate and learn about aphasia. Immediately after viewing the film, the experimenter handed out the retrieval packets. Each packet contained instructions for the experiment and two questions. When all of the participants received their packet, the experimenter gave the following instructions:

> Now you will be asked to describe both the psychologist and the patient in this interview. Please list as much as you can remember about the patient, first, and then the psychologist.

All of the participants described the man first and then described the woman. Participants were instructed how to give each description. The participants were asked not to report in narrative form or complete sentences. Rather they were asked, in an effort to ''save time'', to ''list the basic facts, listing one fact per line''. They were given an example of what we expected this to look like. A majority of participants (98.5%) followed these instructions and were included in the analyses.

Participants then gave their descriptions. They were not allowed to return to an earlier description once they had turned the page and begun giving the next description. The recall phase was self-paced as participants could spend as long

as they wished on each item. Moreover, they could report as much or a little as they wished to. When all the participants in each session had completed the packet, they were debriefed and thanked for their participation. We assigned items to quartile in an identical manner to Experiment 2.

Results

Recall Performance. There were no differences in the total amount of correct information ($F<1$) or incorrect information ($F = 2.6$) for the description of the man and the woman. Overall, participants reported 4.7 correct items per person and reported .73 errors per person. Thus, the overall accuracy was 86.6%, comparable to the earlier studies.

Output Order. An analysis was done on the likelihood of a participant making a commission error for each of the four quartiles. More commission errors were reported in the fourth quartile (.33) than for the first three quartiles (.14), F (3,194) = 10.29, $MS_e = 0.14$ (see Fig. 3). The first, second, and third quartile were not significantly different from each other (the second and third quartile were marginally different, $P = .06$). The interaction between kind of question (male vs. female) and output order was not statistically significant ($F = 1.07$).

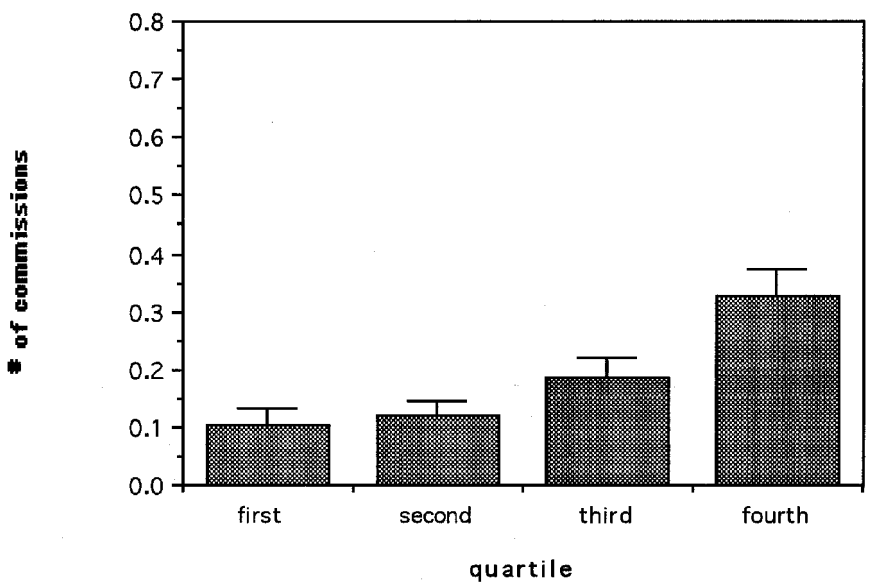

FIG. 4. Commission errors as a function of output order for descriptions of people in Experiment 3.

Discussion

In Experiment 3, we again demonstrated that commission errors are most likely to occur at the end of the output order when participants are describing non-sequential information. Although not as large as the effect observed in Experiment 1, in Experiment 3 participants made twice as many errors during the last quartile of their output than any of the first three quartiles when describing either an older male neuropsychological patient or a younger female neuropsychologist. This is similar to the pattern observed in Experiment 2 for the description of the two criminals. Commission errors and output order were associated in that experiment as well. Therefore, in two widely differing scenarios in which participants describe people, we have found a positive association between commission errors and output order.

GENERAL DISCUSSION

What underlying cognitive mechanisms are responsible for the output order–accuracy relation? We did not test any specific explanations of the relation here, so we can only speculate what these mechanisms might be. Two hypothetical mechanisms come to mind: (a) response inhibition and (b) relaxing the accuracy criterion. Response inhibition states that the act of searching through memory decreases the effectiveness of ensuing memory searches (Anderson, Bjork, & Bjork, 1994; Bjork, 1989). Thus, having searched for the first *N*-1 target items renders the search for the *N*th item less effective. Presumably there is relatively little response inhibition at the beginning of recall, and hence few errors appear at early output positions. As the search process continues, response inhibition increases, so that errors occur towards the end of the output, when response inhibition is greatest. Traditionally, response inhibition has been used as an explanation for the failure of the retrieval process to access the searched-for memory (in signal detection terms, a *miss*). The concern here, however, is not so much a failure to find the searched-for item, but rather producing an incorrect recollection (a *false alarm*). So, the retrieval inhibition mechanism, by itself, should not account for the finding of commission errors. Nor can it explain why errors increase in the middle of the output for sequential information.

The second explanation, relaxing the accuracy criterion, accounts better for commission errors, as it suggests that as one progresses further into the output, the accuracy criterion for producing a response is relaxed. As a result of relaxing the output criterion, the rememberer may produce a response that he or she might have withheld earlier in the output, with a more conservative criterion. One possible explanation of why the rememberer might lower his or her output criterion as more responses are produced is that, because of the response inhibition mechanism, the difficulty of retrieving ensuing items increases. Lowering the response criterion may then serve as a way to

counteract the decreased effectiveness of the search process. Adjusting the response criterion may also explain why errors increase then decrease in reporting sequential information. Participants may compensate for their poor memories of items in the middle of the sequence by lowering the criterion in the middle then raising it again for the more memorable items towards the end. Thus, changing the response criterion may be done to adjust for changing retrievability.

Alternatively, or perhaps in concert with this suggestion, lowering one's response criterion may come as a result of (perceived) social pressure on the respondent to generate additional responses. Such a socially mediated process is compatible with Roediger and Payne's (1985) study. In this study, the experimenter explicitly instructed participants to generate a specific number of responses (60), well beyond the amount they would normally recall with conventional (unforced) recall instructions (ca. 25–30). Not surprisingly, the number of commission errors was drastically increased by the experimenter's explicit request for additional output.

An alternative explanation of the relation between output order and commission errors was recently advanced by Payne et al. (1996) to explain the phenomenon with related lists (i.e. Roedger & McDermott's 1995 stimuli). Their explanation is based on the fuzzy-trace theory of Brainerd and Reyna (1990). Fuzzy-trace theory stipulates two kinds of memory representations, a verbatim or literal representation, and a gist or meaning-based representation. The theory predicts that verbatim memories are reported first, and slower gist memories are reported later. Therefore, fuzzy-trace theory accounts for the output order/commission error for related lists. The critical lures in the Roediger and McDermott procedure are commission errors but they are related to every item actually on the list. Fuzzy-grace theory can also account for the results of Experiments 2 and 3 here. Retrieving the gist may lead some participants to report errors concerning the perpetrators. People may have schemata for what thieves should look like, what they may wear, and what pickup trucks may look like. Indeed, a frequent error was reporting Florida licence tags on the truck, which did not have a licence plate at all.

That the output order-accuracy relation holds across both the eyewitness and verbal learning paradigms should probably not be surprising, as there is no compelling reason to believe, *a priori*, that the two paradigms differ in a way that should alter the relation (Fisher, 1995). However, one puzzling difference did exist in the results from Experiments 1, 2, and 3. In Experiments 1 and 3, commission errors increased most from the third to the fourth quartile, whereas in Experiment 2 the greatest increase in commission errors was from the first to the second quartile. In both experiments, there was a monotonic increase in errors as a function of output position, but the steepest increase in the curve was different in each experiment (contrast Figs. 1 and 2). Perhaps, because of the more stressful stimuli—an armed robbery—participants were more likely to

lower their response inhibition early. Consistent with the idea of an earlier relaxing of the response criterion in Experiment 2 is the lower accuracy of the information reported in that experiment.

We suspect that the relation between output order and accuracy has important implications for both research and application in eyewitness recollection. First, it sounds a note of caution when interpreting the effects of experimental variables that are correlated with output order. Suppose, for instance, that some Variable X, which is predictive of accuracy, is also correlated with output order. The relationship between Variable X and accuracy may occur simply because X affects output order, i.e. the X-accuracy relation is accounted for (mediated) by the intervening variable of output order. The variable X, in and of itself, may have minimal or no effect on accuracy outside its ability to manipulate output order. Indeed, in Experiment 2, the change in output order as a function of type of question (recalled descriptions vs. recalled event sequence) was accompanied by a difference in accuracy (participants were more accurate for recalled sequences). Although this may not mean that the two variables are linked, it does leave the possibility open that changes in accuracy may simply reflect changes in output order.

Another possible instance of confounding output order with another variable is the often-cited claim that open-ended questions lead to higher response accuracy than do direct-probe questions (see Fisher 1995). Note, however, that in a typical eyewitness interview, open-ended questions are asked first, with direct-probe questions used primarily to follow up on information not elicited by the open-ended question (Fisher & Geiselman, 1992). If this questioning order is followed, then responses to open-ended questions are made before those to direct-probe questions, i.e. question type is correlated with output order. Question type, itself, may have no direct effect on accuracy outside its association with output order.

Practically, the output order-accuracy relation may be valuable to a police investigator as a means of assessing the accuracy of individual items within an eyewitness's report. When an investigator asks questions that elicit long answers, he or she may use output order—specifically those items that appeared at the very end of the respondent's protocol—as a potential means of discriminating between accurate and inaccurate recollections. We do not mean to imply that all information given at the end of the output is incorrect. Clearly, our data show that even most final items in the output are still correct. Nonetheless, the probability that they are correct is lower. This type of procedure of differentially weighting recollections based on markers that are reliably associated with accurate (early to middle output) versus inaccurate recall (late output) may lead to increasing the functional accuracy of the witness's recollection, and with no obvious cost. This is particularly valuable, as other efforts to increase the accuracy of recollection by preventing incorrect responses (e.g. by severely penalising respondents for wrong answers) incur a

heavy cost, of severely reducing the amount of correct information recalled (Koriat & Goldsmith, 1994).

We can easily imagine that other, non-police interview contexts might make use of the output order–accuracy relation. For instance, therapists often elicit long, detailed descriptions of autobiographical experiences. To the extent that it is important for therapists to distinguish between accurate and inaccurate recollections, output order may again be of assistance. Another possible outlet for using the output order–accuracy relationship is in journalism and history, in which researchers who conduct oral histories frequently elicit lengthy responses that typically contain primarily correct recollections, but which also include a smattering of incorrect recollections. In the task of piecing together these mixed recollections to formulate an historically accurate event, these interviewers might also rely on output order to promote accuracy. We leave it for those more knowledgeable in these areas than we are to develop potential uses for the relationship.

Manuscript received 11 October 1996
Manuscript accepted 20 February 1997

REFERENCES

Anderson, M.C., Bjork, R.A., & Bjork, E.L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063–1087.

Bjork, R.A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H.L. Roediger III & F.I.M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). Hillsdale, NH: Lawrence Erlbaum Associates Inc.

Brainerd, C.J. & Reyna, V.F. (1990). Gist is the gist: The Fuzzy-trace theory and new intuitionism. *Developmental Review, 10*, 3–47.

Dunning, D., & Stern, L.B. (1992). Examining the generality of eyewitness hypermnesia: A close look at time delay and question type. *Applied Cognitive Psychology*, *6*, 643–657.

Fisher, R.P. (1995). Interviewing victims and witnesses of crime. *Psychology Public Policy, and Law*, *1*, 732–764.

Fisher, R.P., & Gieselman, R.E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, Ill: Charles Thomas.

Fisher, R.P., & Schwartz, B.L. (1995, November). *Increasing the accuracy of recollection*. Paper presented at the meeting of the Psychonomic Society, Los Angeles.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*, 311–333.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*, 297–315.

Koriat, A., & Goldsmith, M. (1996). Memory metaphors and the real-life laboratory controversy: Correspondence versus storehouse conceptions of memory. *Brain and Behavioral Sciences*, *19*, 167–188.

Krinsky, T., & Nelson, T.O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica*, *58*, 141–158.

Kuçera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Leippe, M.R. (1994). The appraisal of eyewitness testimony. In D. Ross, J. Read, & M. Toglia (eds.), *Adult eyewitness testimony: Current trends and developments*. London: Cambridge University Press.

Lindsay, D.S., & Read, J.D. (1994). Psychotherapy and memories of childhood sexual abuse: A cognitive perspective. *Applied Cognitive Psychology*, *8*, 281–338.

Loftus, E.F., & Hoffman, H.G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General*, *118*, 100–114.

McDermott, K.B. (1996). The persistence of false memories in list recall. *Journal of Memory & Language*, *35*, 212–230.

Murdock, B.B. (1968). Modality effects in short-term memory: Storage or retrieval? *Journal of Experimental Psychology*, *77*, 79–86.

Payne, D.G., Elie, C.J., Blackwell, J.M., & Neuschatz, J.S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*, 261–285.

Roediger, H.L. (1996). Memory illusions. *Journal of Memory and Language*, *35*, 76–100.

Roediger, H.L., & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 803–814.

Roediger, H.L., & Payne, D.G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition*, *13*, 1–7.

Roediger, H.L., Wheeler, M.A., & Rajaram, S. (1993). Remembering, knowing, and reconstructing the past. In D. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory*. (Vol. 30, pp.97–134). New York: Academic Press.

Schooler, J.W., Gerhard, D., & Loftus, E.F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *12*, 171–181.

Scrivner, E., & Safer, M.A. (1988). Eyewitnesses show hypermnesia for details about a violent event. *Journal of Applied Psychology*, *3*, 371–377.

Smith, S.M. (1995, November). *Distinguishing recovered memories from false ones: The comparative memory paradigm*. Paper presented at the meeting of the Psychonomic society, Los Angeles.