

Category structure and created memories

STEVEN M. SMITH and THOMAS B. WARD
Texas A&M University, College Station, Texas

DEBORAH R. TINDELL
Wilkes University, Wilkes-Barre, Pennsylvania

CYNTHIA M. SIFONIS
University of Illinois at Urbana-Champaign, Urbana, Illinois

and

MERRYL J. WILKENFELD
Texas A&M University, College Station, Texas

Cued recall of categorized lists was used to examine effects of category structure on the creation of false memories. In three experiments, category members that had not been presented on studied categorized lists were nonetheless recalled by participants. Delaying the category cued recall test (Experiment 1) and priming category members that had been omitted from target lists (Experiment 3) increased the frequency of false recall. All three experiments showed that nonpresented category members that were higher in output dominance were more frequent intrusions. The typicality of category members, however, did not uniquely contribute to the predictability of false recall (or accurate recall) once the contribution of output dominance was taken into account, suggesting that item accessibility (related to output dominance) may be more instrumental than item distinctiveness (related to typicality) in causing certain types of false recall. The results show that created memories in category cued recall are strongly biased by prior category knowledge and can be predicted by the graded structure of categories, particularly in terms of the output dominance of category instances.

Category knowledge is often used to guide episodic recall or reconstruction. An example is the situation in which a patient with an allergic reaction is asked to list all of the foods, cleaning agents, animals, plants, and toxic chemicals contacted in any way within the previous 48 h. Likewise, when an eyewitness lists all of the vehicles seen outside a bank on a particular morning, when a manager must appraise a worker's performance by recalling various types of behaviors related to productivity from the previous year, and when a psychiatric patient tries to recall incidents of various types of abuse during childhood, category knowledge may have a guiding effect on the reconstruction of such memories.

The guiding effect of category or other conceptual knowledge can help the retrieval process by providing facilitative cues, but the same guiding knowledge can also lead to systematic memory errors. A false memory is a type of memory error that is said to occur when events that did not actually take place are nonetheless experienced as au-

thentic memories (e.g., Roediger & McDermott, 1995). We maintain that false memories are not random events but, rather, are lawful predictable occurrences, constructed from existing knowledge and memories. Furthermore, as we will demonstrate, the graded structure of category knowledge can influence the creation of false memories. In the present study, we examined false recall as a function of the category knowledge used to guide the recall process.

The internal structure of categories may influence both accurate and false recall. Categories tend to have graded structures such that examples can be seen to lie on a continuum of representativeness or goodness of membership within the category (e.g., Barsalou, 1985), and those variations in representativeness may influence memory performance. Schmidt (1996) pointed out that category typicality is a function of the similarity of the items in a category to the category's "conceptual core" (p. 596) and that retrieval of items from a categorized list is therefore more likely to occur for typical category members than for atypical items. Consistent with this hypothesis, Schmidt found in three experiments that typical category members were recalled better than atypical items from categorized lists that participants had studied. In accounting for category cued recall, Hunt and McDaniel (1993) and Schmidt acknowledged the need to consider both distinctiveness (inversely related to category typicality) and item accessibility in predicting correct recall. In category recall, ac-

The authors thank David Gleaves and Benton Pierce for comments on this research, Amy Blakely, Scott Giffen, Tamara Getz, Michelle Hankins, Kellee Shepherd, Allyson Tole, and Susan Watson for running the reported experiments, and John Finch and Todd Gilliland for help with data analysis. Correspondence should be addressed to S. M. Smith, Department of Psychology, Texas A&M University, College Station, TX 77843-4235 (e-mail: sms@psyc.tamu.edu).

ording to this explanation, similarity of categorized list items to the recall cues (i.e., conceptual information about the category combined with episodic retrieval cues) facilitates retrieval of those items. Hunt and McDaniel, for example, described recall as guided by the similarity (determined in terms of the aligned features among items) of the target items. The aligned properties of items that are recalled are used, in conjunction with the frame of reference provided by retrieval cues (in this case, category cues), to remind the participant of similar items in the target memory set. This account of category cued recall resembles the feature overlap hypothesis (Greenberg & Bjorklund, 1981).

In the present study, we were primarily interested in the effects of category knowledge on the creation of false memories. Typical nonpresented category members should be falsely recalled more frequently than atypical items, because they should be more easily accessed during recall—a prediction supported by Schmidt's (1996) results. According to Hunt and McDaniel (1993), target memories are accessed on the basis of their relation to the "cue environment," defined as a combination of unique episodic cues and relational information, such as category knowledge. Given that nonpresented category members are unlikely to share episodic information with the recall cues given at test, it follows that false recall of nonpresented items should closely reflect the graded structure of the category.

Although category graded structure is often measured in terms of typicality, there are other measures that reflect different aspects of graded structure, such as category ideals (i.e., how well category instances serve a goal associated with a category), central tendency (i.e., an exemplar's average similarity to other category members), and category output dominance (i.e., the frequency with which an instance is given as an example of a category). Whereas typicality has been considered largely in terms of its relationship to item distinctiveness, its effects on memory have also been discussed in terms of item accessibility, on the basis of the idea that less typical category members are less accessible (e.g., Schmidt, 1996). In addition to Schmidt's finding that less typical category members were more poorly recalled, his results also indicated a strong relationship between item typicality and simulated recall, a measure of item accessibility. For simulated recall, participants were asked to pretend that they were recalling categorized lists of words. The frequency with which a category member was "recalled" in this procedure correlated with item typicality and correlated with actual recall of the item when it appeared as a target on a list that had been studied, but this simulated recall measure did not add to the predictability of recall after item typicality had been taken into account.

Another measure of item accessibility is output dominance, the frequency with which an item is listed as an instance of a category in a category production norm study (Barsalou, 1985). Although output dominance is similar to simulated recall, there may be differences in the two

measures due to differences in metacognitive knowledge that participants bring to bear on category production versus simulated recall tasks. In simulated recall, participants may try to guess at the composition of study lists they have not seen, whereas no such subject demand characteristic is likely to influence category norm production. In the present experiments, we examined the role of output dominance in false recall.

There is precedent for believing that output dominance will play an important role in cognitive tasks and that its effects may differ from those of typicality. Ward, Sifonis, and Wilkenfeld (1996) showed that the exemplars highest in output dominance for a given category are the ones most likely to influence the form of the new ideas that people create from that category in a task of imagination (e.g., Ward, 1994). Specifically, they had some groups of college students list the first 20 Earth animals and tools that came to mind, and they had other students develop imaginary animals and tools that might exist on other planets. In each case, there was a strong positive correlation between the output dominance, defined as the likelihood of an item being listed as a category member (e.g., a dog being listed as an instance of the category "Earth animal") and the likelihood of that item being used as the basis for an imaginary idea (e.g., subjects patterning their imagined extraterrestrials after dogs). Ward et al. also examined several other aspects of category structure, including the rated typicality of category instances, but output dominance was the best predictor of use in imagination. When people are asked to generate novel ideas in a task of imagination, they clearly engage in constructive cognitive processes. Memory retrieval can also be viewed as a constructive process, and the same variables that affect the development of imaginative new ideas might also be expected to influence the development of false or "created" memories (i.e., memories based on something other than exposure to the relevant item). Thus, in the present study, we examined both output dominance and typicality of category members across a wide range of categories to determine the extent to which those variables relate to both false and accurate recall.

To experimentally examine this question, we used a list-learning method for evoking false memories modeled after a procedure reported by Roediger and McDermott (1995). This method, based on one devised by Deese (1959), evokes a specific false memory of a nonpresented word for each target list. A similar method was used in the present experiments, involving the presentation of categorized lists with common category members omitted from each list. Smith et al. (1996) used a variation of the task reported in the present experiments in which a nonpresented category member had been primed in an incidental task. This primed false recall task was found to be effective for evoking false memories of the primed nonpresented items.

In the first experiment of the present study, we examined the efficacy of the categorized false recall methodol-

Table 1
Mean Proportions of Critical Intrusions
Correct Recall and in Experiment 1

	Delay of Test			
	Immediate		Delayed	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
Critical Intrusions	.09	.023	.23	.053
Correct Recall	.68	.015	.47	.027

ogy for evoking false recall of the targeted nonpresented items, using an immediate or a delayed category cued recall test (no primes were used, as had been in Smith et al.'s 1996 study). The frequency of an item's false recall was examined as a function of its output dominance (Barsalou, 1985) in Experiment 1. In Experiment 2, the frequency of false recall of category members was correlated with the items' output dominance and typicality; both measures were expected to be predictive of false recall. In Experiment 3, category members at different levels of output dominance that were not memory targets were incidentally primed to determine experimentally the effect of this measure of graded structure on false recall.

EXPERIMENT 1

Experiment 1 tested the hypothesis that if common members of categories were selectively omitted from categorized word lists, then false recall of the common category members would occur. We used category typicality (Rosch, 1975) as a means of determining which category members to omit from lists. Of primary interest were cases in which targeted nonpresented items were recalled, to be referred to as *critical intrusions*. It was predicted that substantial levels of false recall of the typical nonpresented items would be observed. Furthermore, although the critical nonpresented items were the most typical members of their categories, they varied according to their output dominance (Barsalou, 1985). It was predicted that, for both critical and noncritical intrusions (i.e., intrusions other than the critical nonpresented item), the frequency of false recall of an item would correlate with the item's output dominance.

In Experiment 1, we also tested the effect of a brief delay on false recall of nonpresented category members. Delays have been found to increase false memories in other list-learning paradigms (e.g., McDermott, 1996) and were likewise predicted to increase the frequency of critical intrusions in Experiment 1.

Method

Participants. The participants in all three experiments were undergraduate psychology students who completed part of a course requirement by participating. The participants were recruited using posted sign-up sheets for group sessions. Student volunteers could enroll for any session of many experiments, including the present experiment. Because unequal numbers of participants enrolled for different sessions, there were unequal numbers in the treatment groups. Sessions were held in groups of approximately 5–15 par-

ticipants at a time. There were 46 students who participated in Experiment 1.

Materials. Nine categorized lists were constructed from Rosch's (1975) semantic category study. Each list corresponded to a different category, such as "Clothing" or "Vehicles." Each list contained 15 items from a category. The most typical member (the critical item) was omitted from each list, and the next 15 most typical exemplars were included on each list.

Stimuli were recorded on videotape from the output of a computer program and were shown to the participants on a large television screen. Each word was shown individually on the screen, using large black uppercase letters on a white background.

Design and Procedure. The participants viewed all nine categorized lists, with a recall test either immediately after each list or with recall tests delayed until all nine lists had been studied. At the beginning of each list presentation, the category name (e.g., "Birds" or "Furniture") was shown for 3 sec, after which the 15 list items were shown. Each of the 15 items on a list was shown for 1.5 sec. The category names were subsequently used to cue recall for each list. On each recall test, a category name appeared on the screen, and the participants were told to recall items that had been shown on that list. Guessing was explicitly discouraged. Each recall test lasted 2.5 min and was followed by a 5-sec "ready" signal preparatory to presentation of the subsequent list.

Delay of test was manipulated between subjects. In the immediate test condition, each list was followed immediately by a recall test. In the delayed test condition, all nine lists were shown, with a 5-sec pause between lists. After all lists had been seen, the nine cued recall tests were given in the same order as the studied lists. A 10-sec pause was given after each cued recall test.

Results

All of the analyses in all of the experiments reported used a significance level of $p < .05$, unless specified otherwise.

Of primary interest were cases in which critical items were falsely recalled (critical intrusions). *Noncritical intrusions* refer to all intrusions other than the targeted critical items. A one-way analysis of variance (ANOVA) was computed to compare the probability of correctly recalling list words with the probability of critical intrusions. The correct answer rate was computed as the proportion of the list words correctly recalled on each test, and the critical intrusion rate was computed as the probability of falsely recalling a critical intrusion on each test. Thus, the correct answer rate is the likelihood that any given list word was recalled on a test, and the critical intrusion rate is the likelihood that a critical intrusion was written on a recall test.

A 2×2 ANOVA was computed to determine the effect of delay of test (immediate vs. delayed test, a between-subjects variable) and type of response (critical intrusion vs. correct answer, a within-subjects variable) on the proportion recalled. There was a significant effect of type of response [$F(1,45) = 160.59$, $MS_e = 0.026$]; correct list words were more likely to be recalled than were critical intrusions (Table 1). The main effect of delay of test was not significant [$F(1,45) = 1.31$, $MS_e = 0.024$]. There was, however, a significant interaction of delay with type of response [$F(1,45) = 27.92$, $MS_e = 0.026$]; more critical intrusions were found in the delayed test than on the immediate test [$t(45) = 2.39$], whereas fewer correct answers

occurred on the delayed test than on the immediate test [$t(45) = 6.91$].

The correlation between the frequency with which a nonpresented word was given as an intrusion (including both critical and noncritical intrusions) and the output dominance of the intruded word was $r = .72$. This correlation was significantly greater than zero [$t(84) = 9.51$]. Including only critical intrusions, the intrusion frequency \times output dominance correlation was $r = .24$, which was not significantly greater than zero [$t(7) = 0.65$].

Discussion

Experiment 1 demonstrated the effectiveness of the category cued recall paradigm in evoking critical intrusions. In Experiment 1, we found greater numbers of critical intrusions when the recall test was delayed than when the recall test was immediate. McDermott (1996) found that critical intrusions increased after a 2-day retention interval, but not after a 30-sec retention interval. In the present experiment, critical intrusions increased when the category cued recall tests were delayed an average of approximately 11 min, with the delays ranging from about 4–18 min. This increase in critical intrusions after a delay of several minutes is consistent with Read's (1996) finding of high levels of critical intrusions using associatively structured target lists, because in Read's experiments retention intervals delayed recall for several minutes (5 min in one experiment and 2 min in another).

The output dominance of nonpresented category members was found to be strongly correlated with the frequency of their occurrence as intrusions. This finding reflects the role of category structure in created memories. Although the correlation was in the predicted direction when the analysis included only critical intrusions, it was not significant, possibly because the range of output dominances for the critical items was too restricted. A set of critical items with a greater range of output dominance scores was therefore used in Experiment 2.

EXPERIMENT 2

Our examination of the relation between category structure and false recall thus far has shown that people often falsely recall highly typical and dominant items that have been omitted from categorized lists. The finding is analogous to that from associative lists in which participants falsely recall the one omitted item that is the strongest associate of the presented items (e.g., Roediger & McDermott, 1995). However, the use of categorized study lists makes it possible to systematically observe the effects of gradations in the strength of items from the category rather than limiting the focus to the single most dominant one. In Experiment 2, we examined a broader range of dominance of the critical omitted items. That is, the categorized target lists in Experiment 2 were more "spread out" to include high, medium, and low output dominance items. To achieve this greater range of items, the members of a category were first rank ordered from

highest to lowest in terms of output dominance. From this rank ordering, alternate items were assigned to two 15-item target lists (Lists A and B), each spanning approximately the same range of items in terms of output dominance. In Counterbalancing 1, Target List A served as the study list, and List B served as the list of potential intrusions, whereas the reverse was true for Counterbalancing 2.

Experiment 2 consisted of two parts. In the first part, we collected norms of category associates for 10 taxonomic categories. In the second part, different participants studied the 10 lists from one counterbalancing and were given category labels as recall cues, one categorized list at a time, as described for Experiment 1. Thus, accurate and false recall for the same category members was assessed in the two counterbalancings of Experiment 2.

Our primary interest was the relation between frequency of false recall of an item and the goodness of that item as a category member, as measured by typicality ratings or output dominance measures. Typicality ratings of category members are considered to measure the similarity of items to the "conceptual core" of the category, as determined, for example, by a category prototype. As such, typicality ratings are often discussed in terms of the distinctiveness of items in relation to other category members (less typical items are more distinctive) and in terms of item accessibility (e.g., Schmidt, 1996). The relation between item accessibility and typicality is not completely clear; although distinctiveness often enhances recall (Hunt & McDaniel, 1993), it has been found that item accessibility is poorer for less typical (more distinctive) category members. Output dominance, another measure of a category's graded structure, is measured as the frequency of a response when participants are asked to list instances of categories (Barsalou, 1985). Output dominance is more clearly a measure of item accessibility, indicating the retrieval fluency or "coming-to-mindedness" of items when they are generated in response to category cues. Because these two indicators of category graded structure measure different aspects of the category members and have been shown to differentially affect performance in some tasks (Ward et al., 1996), we assessed the independent contributions of the two measures in false recall.

Method

Participants. There were 156 participants in Experiment 2. Of these, 36 were used for output dominance norms, 42 were used for typicality norms, and 78 were used in the category cued recall portion of Experiment 2.

Category norms. Category typicality ratings and output dominance scores were collected in Experiment 2 for the same 10 taxonomic categories used by Barsalou (1985). Each of the 10 category names was given to 36 participants, who were asked to list as many instances of each category as possible. This procedure was self-paced. The output dominance score for each item in each category was computed as the number of participants who listed a particular item as a category member.

Thirty of the items that were highest in output dominance were selected from each of the 10 categories. Another group of 42 par-

Table 2
Simple Correlations and Partial Correlations Between
Typicality, Output Dominance, and the Dependent Measures
Proportion Intrusions, Confidence for Intrusions, Proportion
Correct, and Confidence for Correct Items in Experiment 2

Measure	Zero-Order Correlations		Partial Correlations	
	Typicality	Output Dominance	Typicality	Output Dominance
Proportion Intrusions	.41*	.62*	.08	.50*
Confidence for Intrusions	-.02	-.03	-.02	.01
Proportion Correct	.21*	.46*	-.01	.39*
Confidence in Correct	-.41*	-.42*	-.18*	-.23*

*A significant effect.

ticipants who had not undergone the category instance listing procedure was asked to rate the typicality of each of the 30 instances listed for each of the 10 categories. The participants were instructed to rate each category member on a scale of 1–7 in terms of how good a member of its category the item is, with 1 indicating a *poor example of its category* and 7 indicating an *excellent category member*. This procedure was self-paced.

After rank ordering the 30 exemplars of each category according to output dominance, we assigned alternate items to Counterbalancing Lists A and B, creating two different study lists for each category that were approximately equal to each other in output dominance. The average output dominance of the items was 14.0 for the A lists and 13.6 for the B lists. The average rated typicality for the studied items was 5.7 (out of 7 as a maximum rating) for the A lists and 5.6 for the B lists. The 10 A lists were used as the study materials for 40 of the participants, and the B lists served as study lists for 38 of the participants. In each counterbalancing, the items recalled from the nonstudied lists were counted as intrusions, or false recall.

Design and Procedure. The procedure used in Experiment 2 was essentially the same as that described for the delayed recall condition of Experiment 1, with a few exceptions. Rather than the nine categorized lists, there were 10 lists studied and recalled in Experiment 2. After the participants studied and recalled the 10 categorized lists, as described for Experiment 1, they went back over the words they had recalled and wrote a confidence judgment for each. Confidence scores ranged from 1 (*a complete guess*) to 10 (*absolutely certain the response was on the studied list*). The participants' assignment of confidence scores was self-paced.

Results

Zero-order correlations between typicality, output dominance, and four dependent measures (proportion intrusions, confidence in intrusions, proportion correct, and confidence in correct items) are shown in Table 2. Items higher in typicality and output dominance were both falsely and correctly recalled more often than items lower in typicality and output dominance, and the correlation between output dominance and recall was somewhat stronger than the correlation between typicality and recall. Neither measure was strongly related to confidence in intrusions, but both were negatively correlated to the same extent with confidence in correctly recalled items. These zero-order correlations fail to take into account, however, the high correlation between typicality and output dominance ($r = .52$), which means that the zero-order correlations might overestimate the unique contribution that typicality makes in predicting false recall.

To examine the separate and combined contributions of list, typicality, and output dominance, we ran three separate hierarchical multiple regression analyses for each of the four dependent measures (proportion intrusions, confidence in intrusions, proportion correct, and confidence in correct items). For all three of the regression analyses of each dependent measure, we entered list (coded using nine dummy variables) in the first step. For one analysis of each of the four dependent measures, typicality was entered alone in the second step. For the second analysis of each measure, output dominance was entered alone in the second step. For the third analysis of each measure, both typicality and output dominance were entered simultaneously in the second step. The third analysis provided a measure of the relative importance of typicality and output dominance in uniquely explaining the variance of proportion intrusions, confidence in intrusions, proportion correct, and confidence in correct items.

Proportion intrusions. The reported results are based on intrusions corresponding to the items selected for Experiment 2. The mean number of intrusions on each cued recall test was 0.75 items ($SD = 1.09$).

The incremental contribution of typicality was significant in predicting proportion intrusions [incremental $R^2 = .17$, incremental $F(1,289) = 62.46$], as was the incremental contribution of output dominance [incremental $R^2 = .41$, incremental $F(1,289) = 213.60$]. When typicality and output dominance were simultaneously added to the equation, the incremental contribution of typicality and output dominance was significant in predicting the proportion intrusions [incremental $R^2 = .42$, incremental $F(2,288) = 109.19$]. This was due to the unique contribution of output dominance ($\beta = .651$, $t = 11.33$), a significant partial correlation (Table 2). Items with a high output dominance were more likely to be recalled than items with a low output dominance. The unique contribution of typicality did not significantly predict the proportion of intrusions ($\beta = .095$, $t = 1.79$), as shown by the partial correlation (Table 2).

Confidence in intrusions. Confidence scores associated with intrusions spanned the entire range from 1 to 10. The modal confidence score, which was given for 23% of the intrusions, was a 10, which indicated complete confidence that the item was actually correct. The

mean confidence level for intrusions was 6.1. Only 14% were rated as 1, indicating complete guesses. Neither the incremental contribution of typicality (incremental $R^2 = .00$) nor that of output dominance (incremental $R^2 = .00$) was significant in predicting confidence in intrusions.

Proportion correct. The incremental contribution of typicality was significant in predicting proportion correct [incremental $R^2 = .06$, incremental $F(1,289) = 20.38$], as was the incremental contribution of output dominance [incremental $R^2 = .21$, incremental $F(1,289) = 89.19$].

A clearer picture was obtained when both typicality and output dominance were added together at Step 2, explaining 33% of the variance, a significant effect [$F(11,288) = 12.62$]. The incremental contribution of typicality and output dominance was significant in predicting proportion of correct recall [incremental $R^2 = .21$, incremental $F(2,288) = 44.47$]. This effect was due mostly to the unique and significant contribution of output dominance ($\beta = .508$, $t = 8.01$), indicated by the significant partial correlation (Table 2). Items with a high output dominance were more likely to be recalled than items with a low output dominance. Typicality did not uniquely predict correct recall ($\beta = -.010$, $r = -.01$, $t = -.181$). Although typicality significantly predicted proportion correct when it was added alone, it did not explain any variance that could not be accounted for by output dominance.

Confidence in correct recall. The incremental contributions of both typicality [incremental $R^2 = .13$, incremental $F(1,289) = 50.20$] and output dominance [incremental $R^2 = .16$, incremental $F(1,289) = 60.84$] were significant in predicting confidence in correct items.

When typicality and output dominance were both added to the equation at Step 2, their incremental contribution was significant in predicting confidence of correct responses [incremental $R^2 = .19$, incremental $F(2,288) = 38.10$]. Both typicality and output dominance had significant unique contributions to the confidence of correct responses (typicality, $\beta = -.218$, $t = -3.59$; output dominance, $\beta = .310$, $t = -4.72$), as shown by partial correlations (Table 2). Confidence was higher for items low in typicality and low in output dominance.

Discussion

When considered individually in terms of a zero-order correlation, typicality was a reliable predictor of false recall. Typicality and output dominance were strongly correlated, however, so it is not clear, without considering partial correlations, whether this predictability was due to typicality, output dominance, or the overlap of the two variables. Partial correlations show that the unique contribution of output dominance in predicting false recall was significant, accounting for 25% of the variance, whereas the unique contribution of typicality to predicting false recall was not significant. Findings that typicality is predictive of false recall (e.g., Schmidt, 1996) are not inconsistent with the results of Experiment 2, but those findings may be related more to output dominance than to typicality. The present results suggest that false recall is related

more to accessibility, as measured by output dominance, rather than to distinctiveness, as measured by typicality ratings. The finding parallels that of Ward et al. (1996), who observed a stronger correlation between output dominance and the use of exemplars in imagination than between typicality and use in imagination.

Output dominance, relative to typicality, was also found to be a better predictor of accurate recall. As with false recall, typicality predicted accurate recall only when it was examined as a zero-order correlation. The unique contribution of typicality to predicting accurate recall was not reliable. Again, this finding is not inconsistent with findings that accurate recall of categorized lists correlates with the typicality of the list members (Schmidt, 1996), because zero-order correlations of typicality and correct recall levels were significant. It appears, however, that correlations between typicality and correct recall occur because recall levels are related to output dominance rather than typicality. As with false recall, this finding suggests that category cued recall is related more to item accessibility than to item distinctiveness.

Category typicality has been discussed as an important contributor to an item's distinctiveness (e.g., Hunt & McDaniel, 1993; Schmidt, 1996), whereas output dominance may be more a measure of retrieval fluency (e.g., Jacoby & Dallas, 1981) or how easily an item comes to mind when one is given a category name as a recall cue. The present results suggest that both false and accurate recall in this category cued recall paradigm are strongly determined by the retrieval fluency of items, but not by their distinctiveness. Because the results of Experiment 2 were correlational, the design of Experiment 3 used an experimental manipulation—priming critical items that varied in output dominance—testing the prediction that intrusions of lower output dominance items, relative to higher output dominance items, would be less affected by priming. Nonetheless, the results of Experiment 2 reveal the importance of one dimension of category graded structure, particularly as measured by output dominance, in predicting false recall.

Confidence in falsely recalled items did not correlate with either output dominance or typicality. This finding suggests that neither item accessibility nor distinctiveness contribute to one's confidence that falsely recalled items are accurate. The participants were more confident in correct responses low in output dominance or low in typicality, a finding that suggests that retrieval fluency and/or item distinctiveness is used when participants assess their confidence in accurately recalled categorized list members (i.e., participants are less confident in items that are less fluently retrieved or less distinctive).

EXPERIMENT 3

The findings of Experiments 1 and 2 support the hypothesis that the most accessible incorrect responses, as measured by category output dominance, are the most likely intrusions. In Experiment 3, we sought experimen-

Table 3
Design of Experiment 3, Indicating Output Dominance of
Primed Items for Each List in Each Counterbalancing Group

Counterbalancing	Lists			
	A and E	B and F	C and G	D and H
1	High	Medium	Low	No Prime
2	Medium	Low	No Prime	High
3	Low	No Prime	High	Medium
4	No Prime	High	Medium	Low

tal evidence that false recall of items is more difficult to evoke the lower the output dominance of the items. To test this hypothesis, we used a priming technique for evoking false recall reported by Smith et al. (1996). In that study, it was found in a category cued recall test that false recall of highly typical nonpresented category members increased if the critical items were seen in an incidental affect rating task. Would priming a medium or low output dominance member of a category likewise increase the frequency of that item as an intrusion in category cued recall?

The effects of priming and output dominance were examined in Experiment 3. For each categorized list, a prime corresponded to a nonpresented high output dominance member, a nonpresented medium output dominance member, a nonpresented low output dominance member, or a nonmember (no prime) of the category. It was predicted that higher output dominance nonpresented items would be more likely to occur as intrusions on the category cued recall tests and that incidental priming would increase false recall more for higher output dominance items. Furthermore, on the basis of the correlational results of Experiment 2, it was predicted that confidence for critical intrusions would not vary for items at different levels of output dominance.

Method

Participants. There were 58 participants in Experiment 3.

Materials. Eight categorized lists were modified for use in Experiment 3. In Experiment 1, only the highest output dominance member of each category had been omitted from the study lists, whereas, in Experiment 3, there were three critical items identified for each category. In Experiment 3, a high output dominance item, a medium output dominance item, and a low output dominance item were identified and omitted from each categorized list, using Barsalou's (1985) norms to determine output dominance. The three critical items omitted from each categorized list are shown in the Appendix.

Four different priming lists of 18 words were constructed, each corresponding to a different counterbalancing. The same 12 filler words were used in all four counterbalancings. The critical primes were six words in each counterbalanced list of primes. Two of the critical primes on each of the four versions of the priming list were high output dominance members of categories, two were medium output dominance members, and two were low output dominance members.

Design. The variable type of list refers to the eight categorized lists that the participants studied and recalled and was manipulated within subjects. Six of the categorized lists for each counterbalancing (e.g., Lists A, B, C, E, F, and G) corresponded to a critical word on the priming list, which was either the nonpresented high output

dominance item for that list, the nonpresented medium output dominance item, or the nonpresented low output dominance item. The remaining two lists (e.g., D and H) did not correspond to any related words on the priming list (the nonprimed lists). Thus, there were four different types of list in each counterbalancing. Two lists corresponded to a high output dominance prime (e.g., Lists A and E), two to a medium-output dominance prime (e.g., Lists B and F), two to a low output dominance prime (e.g., Lists C and G), and two lists had no related primes (e.g., Lists D and H).

The four lists of primes were counterbalanced between subjects, with each group of participants viewing only one of the priming lists. That is, in Counterbalancing 1, for example, Lists A and E corresponded to high dominance primes; in a second counterbalancing, Lists A and E corresponded to medium dominance primes; in a third, they corresponded to low dominance primes; and, in a fourth, they corresponded to no related primes. This experimental design is shown in Table 3, which indicates the critical items that were primed for each of the eight categorized lists in each of the four counterbalancing groups. In Counterbalancing Group 1, for example, list type would be high dominance primed for Lists A and E, low dominance primed for Lists B and F, medium dominance primed for Lists C and G, and no prime for Lists D and H.

Procedure. The priming procedure involved an incidental learning task in which the participants rated the pleasantness of words. The 24 words on the pleasantness rating task, consisting of six priming words and 18 unrelated filler words, were shown one at a time for 3 sec in black uppercase letters on a white background on a television screen. The participants rated the pleasantness of each word on a scale of -3 (*very bad*) to $+3$ (*very good*).

The remaining procedure after the pleasantness rating task was the same as that described for Experiment 1, except that all of the eight cued recall tests were delayed (see Experiment 1 Procedure section). The participants were instructed to recall words only if they had been shown on the categorized lists that had been studied.

Results

A 3×2 (type of list \times type of response) ANOVA was computed to examine the difference between correct recall and critical intrusion rates when critical nonpresented words had been primed (nonprimed lists were not included in this analysis). Type of list, a within-subjects variable, was low dominance primed, medium dominance primed, or high dominance primed. Type of response, a within-subjects variable, was the proportion of words correctly recalled on the primed lists versus the probability of falsely recalling a critical intrusion on tests of the same primed lists. The means for this analysis are shown in Table 4. There was a significant effect of type of list [$F(2,114) = 40.93$, $MS_e = 0.034$]. There was also a significant effect of type of response [$F(1,57) = 33.57$, $MS_e = 0.103$]. These main effects were mediated by a significant type of list \times type of response interaction [$F(2,114) = 26.11$, $MS_e = 0.039$]. Simple main effect analyses of type of response were computed for each type of list. The effect of type of response was significant for the low dominance primed list [$F(1,57) = 128.03$, $MS_e = 0.039$] and for the medium dominance primed list [$F(1,57) = 9.34$, $MS_e = 0.039$]. The effect of type of response in the high dominance primed list was not significant [$F(1,57) = 3.84$, $MS_e = 0.039$]; that is, the correct recall rate was no different from the critical intrusion rate for primed high dominance items.

Table 4
Mean Proportions of Tests with Critical Intrusions
and Mean Proportion of Correct Recall as a
Function of Priming and List Type in Experiment 3

	List Type					
	Low		Medium		High	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
Unprimed Intrusions	.02	.012	.14	.030	.23	.045
Primed Intrusions	.06	.025	.38	.041	.44	.049
Correct Recall	.48	.021	.49	.018	.51	.018

Note—"List Type" refers to the output dominance of the critical intrusion that was primed in that condition.

Another 3 × 2 (output dominance × priming) ANOVA was computed using proportion of intrusions as the dependent measure. Output dominance was either high, medium, or low, and priming was either primed or unprimed. Both were within-subjects variables. There significant effects of output dominance [$F(2,114) = 44.05$, $MS_e = 0.063$] and priming [$F(1,57) = 23.65$, $MS_e = 0.099$]. These effects were mediated, however, by a significant priming × output dominance interaction [$F(2,114) = 5.69$, $MS_e = 0.057$]. Simple main effects analyses showed significant output dominance effects in both the primed condition [$F(2,114) = 42.27$, $MS_e = 0.057$] and the unprimed condition [$F(2,114) = 11.87$, $MS_e = 0.057$]. Pairwise comparisons for the unprimed condition showed that critical intrusions for high and medium dominance category members occurred more often than low dominance items [$ts(114) = 4.84$ and 3.63 , respectively].¹ The high and medium dominance intrusions for the unprimed condition marginally differed [$t(114) = 1.70$, $p = .094$]. Pairwise comparisons for the primed condition also showed that critical intrusions for high dominance items and medium dominance items were more frequent than those for low dominance items [$ts(114) = 7.88$ and 7.58 , respectively] and that the difference between high and medium dominance intrusions was not significant [$t(114) = 1.31$].

The above analysis of the output dominance × priming interaction did not completely explain the pattern of results giving rise to the interaction, showing only that simple effects of output dominance occurred for both primed and unprimed conditions. To further explicate this interaction, priming effects were assessed at each level of output dominance. Simple main effects analyses showed significant priming effects in both the high output dominance condition [$F(1,57) = 21.78$, $MS_e = 0.057$] and the medium output dominance condition [$F(1,57) = 29.64$, $MS_e = 0.057$], but not in the low output dominance condition [$F(1,57) = 0.95$, $MS_e = 0.057$].

Confidence. A one-way ANOVA computed to compare confidence in intrusions versus confidence in correctly recalled items found a significant difference [$F(1,57) = 48.54$, $MS_e = 76.40$]. Confidence was greater for correct items ($M = 9.3$, $SD = .67$) than for intrusions ($M = 7.8$, $SD = 1.95$). Other results, however, indicate that confidence in intrusions was relatively high. For example, the modal confidence score, which was given for 48% of the

intrusions, was a 10, indicating complete confidence that the item was correct. Less than 5% of the intrusions were rated as complete guesses.

Another one-way ANOVA was computed to compare confidence in intrusions across the three levels of output dominance (collapsing across primed and unprimed conditions). There was no effect of output dominance on confidence in intrusions [$F(2,12) = 0.04$, $MS_e = 6.46$] (Table 5).

Discussion

The results of Experiment 3 replicate and extend the findings of the first two experiments of the present study. The higher the output dominance of the category member, the greater the proportion critical intrusions observed for that item, replicating the same patterns in Experiments 1 and 2. This output dominance effect was observed for both primed and unprimed intrusions. The priming effect was greater for high and medium dominance category members than for low dominance members, showing further that it is more difficult to induce false recall for category members that are lower in output dominance.

Because of the correlations of output dominance × frequency of intrusions found in Experiments 1 and 2, we selectively primed and omitted from the target lists critical items at three levels of output dominance for the nonpresented category members in Experiment 3. This experimental treatment of output dominance led to the same result as was suggested by the correlational evidence generated by Experiments 1 and 2, that higher output dominance category members were more likely to be given as intrusions than were lower output dominance items.

The overall level of confidence that intrusions were actually correctly recalled items was lower than the mean confidence given for accurately recalled items. Nonetheless, the participants indicated that, for nearly half of the intrusions, they were completely confident that the items were accurate. As in Experiment 2, this finding demonstrates that, rather than believing that they are blindly guessing, participants feel confident that most of their falsely recalled responses are, in fact, accurate memories. Also replicating a result found in Experiment 2 was the finding that output dominance was not related to confidence in intrusions.

For high and medium output dominance category members, priming increased false recall by an average of 22%. Furthermore, as was found in the corresponding condition in Smith et al.'s (1996) study, the primed false recall rate for high output dominance items in Experiment 3 was

Table 5
Mean Confidence in Critical Intrusions as a Function
of Output Dominance of Intrusion in Experiment 3

	Output Dominance					
	Low		Medium		High	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
	7.64	0.93	7.74	0.40	7.37	1.15

no different from the correct recall rate. Again, this result indicates the potency of the primed false recall effect.

GENERAL DISCUSSION

All three of the reported experiments found evidence of false memories in the categorized list recall paradigm. That is, when recalling categorized lists of words from which common category members had been omitted, the participants often falsely recalled the nonpresented words even though they had been instructed to recall words only from the categorized lists that had been presented. Delaying the cued recall test a few minutes substantially increased the frequency of false cued recall (Experiment 1). Priming critical items by way of an incidental task also increased the frequency of critical intrusions. The false recall effect was so powerful in some conditions that the probability of responding with a critical intrusion did not differ from the probability of correctly recalling a presented target word. The confidence scores given for the falsely recalled items indicate that the participants did not see most of their intrusions as guesses. The modal confidence rating given to intrusions in Experiments 2 and 3, for example, was a perfect 10, indicating complete confidence.

The frequency of false recall of any particular omitted category member was found to be correlated with the output dominance of that item in Experiments 1 and 2. In Experiment 2, in which items with a range of output dominances were omitted from the target lists, it was found that output dominance uniquely accounted for one fourth of the total variance in false recall scores. In Experiment 3, in which the output dominance of omitted items was experimentally manipulated, it was found to influence the frequency of critical intrusions, with more intrusions occurring for higher output dominance items. For critical items high and medium in output dominance, there was a reliable priming effect in Experiment 3, but, for items low in output dominance, the priming effect was not significant. These findings show clearly that output dominance, a measure of the graded structure of categories, is a reliable predictor of false recall in the category cued recall paradigm used in the present experiments.

The relation between category typicality and false recall is not as clear as the relation between output dominance and false recall. Although typicality, by itself, could be used to predict the frequency of false recall, a more careful analysis shows that when the effects of output dominance are taken into consideration, the unique contribution of typicality for predicting false recall is not significant (Experiment 2). This pattern of results supports the notion that item accessibility, rather than item distinctiveness, is the better predictor of false recall in the category cued recall paradigm used in the present experiments. A similar pattern of results in Experiment 2 showed that the unique contribution of category typicality for predicting accurate recall was also not significant,

indicating that the same aspects of category structure influence both accurate and false recall.

Consistency With Prior Research

The results of our three experiments are highly consistent with several converging lines of research, including studies of false memory, knowledge consistency, and creative cognition. In each of these areas of research, it has been shown that prior knowledge biases the outcomes of cognitive processing, whether the result is accurate and appropriate for a particular task or otherwise, consistent with Ratcliff and McKoon's (1995, 1996) bias principle.

False memory studies using associative lists (e.g., McDermott, 1996; Payne, Elie, Blackwell, & Neuschatz, 1996; Read, 1996; Roediger & McDermott, 1995) have yielded results very similar to the present findings, showing that simple experimental methods can be used to evoke and investigate false memories. One difference, however, which has been highlighted in the present study, is that the category cued recall method allows for more than one targeted intrusion per list. This quality allowed us to investigate false recall of more than one item per list, providing a means for studying contributions of category graded structure to the creation of false memories.

The results of the present experiments are also in line with what may be referred to as *knowledge consistency* studies of memory. These studies show that episodic memory is often guided by long-term semantic knowledge structures, such as categories and schemata. Consistency of events with participants' schemata can facilitate episodic memory for those events, but schema-guided recall can also produce intrusions that are consistent with knowledge structures (e.g., Brewer & Treyens, 1981; Rabinowitz & Mandler, 1983). The role of category knowledge in episodic memory has also been demonstrated (e.g., Schmidt, 1996). The present results add to this body of research, showing that when recall is guided by category structures, category consistent intrusions can occur.

The present experiments have shown a consistent relationship between the output dominance of a category member (Barsalou, 1985) and the frequency of that item as an intrusion in category cued recall. Output dominance is a measure of the accessibility of an exemplar when a category name is given as a semantic cue. It has been shown to play a stronger role than typicality in previous studies (Ward et al., 1996). Ward et al. interpreted their findings in terms of "coming-to-mindedness." That is, when developing a new idea, people tend to pattern their creations after those exemplars that most readily come to mind from the relevant domain. The present results suggest that those items that come to mind most readily also play a major role in accurate recall, and they are often misinterpreted as exemplars that were actually presented. The present finding that output dominance predicts created memories is also consistent with findings that output dominance likewise predicts the features of ideas reported in creative idea generation studies (e.g., Ward et al., 1996).

An understanding of the contributions of category graded structure to predictions of false recall is also relevant to certain practical concerns about the accuracy of memories. Examples of situations in which false memory is an important concern include eyewitness memory, recovery of memories in psychotherapy, and memories of instances of organizational behaviors during job performance appraisal. Each of these examples can involve recall of categories of memories. An understanding of the effects of category graded structure on recall in these situations might eventually lead to methods for predicting, avoiding, or identifying false memories.

Conclusions

The present experiments show that nonpresented category members are often falsely recalled on category cued recall tests and that delaying recall tests and priming critical items increase the frequency of such false or created memories. The results also show that output dominance, a measure of category graded structure, is a powerful predictor of both accurate and created memories and may account for some reported recall test results that have been attributed to typicality.

REFERENCES

BARSALOU, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 629-654.

BREWER, W. F., & TREYENS, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, **13**, 207-230.

DEESE, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, **5**, 305-312.

GREENBERG, M. S., & BJORKLUND, D. F. (1981). Category typicality in free recall: Effects of feature overlap or differential category encoding? *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 145-147.

HUNT, R. R., & MCDANIEL, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory & Language*, **8**, 81-87.

JACOBY, L. L., & DALLAS, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.

MCDERMOTT, K. B. (1996). The persistence of false recall. *Journal of Memory & Language*, **35**, 212-230.

PAYNE, D. H., ELIE, C. J., BLACKWELL, J. M., & NEUSCHATZ, J. S. (1996). Memory illusions: Recalling, recognizing and recollecting events that never occurred. *Journal of Memory & Language*, **35**, 261-285.

RABINOWITZ, M., & MANDLER, J. M. (1983). Organization and information retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 430-439.

RATCLIFF, R., & MCKOON, G. (1995). Bias in the priming of object decisions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 754-767.

RATCLIFF, R., & MCKOON, G. (1996). Bias effects in implicit memory tasks. *Journal of Experimental Psychology: General*, **125**, 403-421.

READ, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review*, **3**, 105-111.

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

ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.

SCHMIDT, S. R. (1996). Category typicality effects in episodic memory: Testing models of distinctiveness. *Memory & Cognition*, **24**, 595-607.

SMITH, S. M., SIFONIS, C. M., TINDELL, D. R., WILKENFELD, M. J., PIERCE, B. H., & WARD, T. B. (1996, November). *Priming and category dominance in created memories*. Paper presented at the annual meeting of the Psychonomic Society, Chicago.

WARD, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, **26**, 1-40.

WARD, T. B., SIFONIS, C. M., & WILKENFELD, M. J. (1996, November). *Graded category structure and generative thinking: The humans and hammers effects*. Paper presented at the annual meeting of the Psychonomic Society, Chicago.

NOTE

1. Using a Bonferroni correction with familywise error = .05, the significance level for each of the three pairwise comparisons was set at $\alpha = .017$.

APPENDIX

Categories, Critical High, Medium, and Low Output Dominance Nonpresented Items, and Output Dominances of Critical Items Used in Experiment 3

Category	Output Dominance		
	High	Medium	Low
Birds	robin (16)	dove (5)	penguin (2)
Clothing	pants (28)	dress (13)	belt (2)
Fruit	apple (32)	watermelon (5)	lemon (2)
Furniture	chair (30)	sofa (10)	rug (2)
Sports	football (33)	softball (4)	wrestling (2)
Vegetables	carrot (18)	cucumber (6)	sprouts (2)
Vehicles	car (31)	train (8)	tractor (2)
Weapons	gun (33)	sword (4)	cannon (2)

Note—Output dominances for critical items were taken from the Barsalou (1985) norms and are shown in parentheses.

(Manuscript received January 14, 1998; revision accepted for publication July 14, 1998.)