The effects of levels-of-processing and organization on conceptual implicit memory in the category exemplar production test

NEIL W. MULLIGAN
Southern Methodist University, Dallas, Texas

and

PAUL S. GUYER and ALLISON BELAND
Illinois State University, Normal, Illinois

The distinction between item-specific and relational information (Hunt & Einstein, 1981) may be relevant to accounts of conceptual priming in the category exemplar production task. In three experiments, the implications of this hypothesis were tested by examining the effects of organization and levels-of-processing (LOP) on this implicit test. Consistent with the hypothesis, the effects of LOP were greater when study lists were organized by category than when they were presented randomly. Furthermore, when subjects claiming test awareness or intentional retrieval were excluded from the analysis, the LOP effect was reduced (and, in fact, eliminated) in the random list condition but remained robust in the categorized condition. Finally, the experimental design of the LOP manipulation (i.e., mixed-list within, blocked-list within, or between subjects) did not moderate the effects of LOP on this priming task.

One important account of dissociations between implicit and explicit tests of memory is based on the principle of transfer-appropriate processing. The central tenet of this view is that performance on a memory task benefits to the extent that the cognitive processes carried out during encoding are re-engaged during retrieval (see, e.g., Kolers & Roediger, 1984; Morris, Bransford, & Franks, 1977). In applying this general principle to implicit and explicit memory tests, Roediger, Blaxton, and colleagues (e.g., Blaxton, 1989; Roediger, 1990; Roediger & McDermott, 1993) have found it useful to distinguish between perceptual (or data-driven) processes, which refer to analyses of a stimulus’s physical or surface-level features, and conceptual processes, which refer to analyses of a stimulus’s semantic content or significance.

The transfer-appropriate-processing (TAP) view of implicit and explicit memory predicts dissociations between memory tests to the extent that they engage different types of retrieval processes. Because many commonly used implicit tests (e.g., perceptual identification, word- and picture-fragment completion) require the identification of perceptually degraded or incomplete stimuli, these tests are assumed to rely heavily on perceptual retrieval processes. In contrast, many of the commonly used explicit tests (e.g., free and cued recall, recognition) are assumed to rely heavily on conceptual retrieval processes (see, e.g., Craik, Moscovitch, & McDowd, 1994; Roediger, Weldon, Stadler, & Riegler, 1992).

Although the most commonly used implicit tests are perceptual in nature, the TAP view proposes that implicit tests are not necessarily perceptual, and that explicit tests are not exclusively conceptual. Given the appropriate combination of memory cues and task instructions, conceptual implicit tests and perceptual explicit tests can be constructed. Furthermore, because the TAP framework focuses on the type of retrieval processes engaged by a test, the TAP view predicts dissociations between perceptual and conceptual tests, whether the tests are implicit or explicit. In addition, the TAP view predicts similar effects of encoding manipulations (i.e., associations rather than dissociations) on an implicit and an explicit test, provided that both tests are of the same type (either perceptual or conceptual).

The TAP framework has had great success accounting for experimental dissociations of implicit and explicit memory tests (see Roediger & McDermott, 1993, for a review). However, recently obtained dissociations among conceptual implicit tests and between conceptual implicit and conceptual explicit tests have produced important challenges to the TAP view. The most compelling of such dissociations were reported by Weldon and Coyote (1996) and Mulligan (1996), in studies comparing performance on matched conceptual implicit and explicit tests—that is, tests that differed only in instructions. Wel-
MULLIGAN, GUYER, AND BELAND

don and Coyote found that study items presented as pictures rather than as words led to superior recall on the conceptual explicit tests of category-cued recall and associated-cued recall. No such picture superiority effect was found on the comparable implicit tests of category exemplar production or word association. Mulligan (1996), in an investigation of the perceptual interference effect, reported a similar dissociation: perceptual interference enhanced recall on the category-cued recall test but had no effect on category exemplar production. Complementary dissociations between conceptual implicit tests (Cabeza, 1994) and conceptual explicit tests (McDermott & Roediger, 1996) have also been reported.

Dissociations within the class of conceptual tests are problematic for the TAP view and suggest that its continued viability requires further refinement (see, e.g., McDermott & Roediger, 1996; Mulligan, 1996; Weldon & Coyote, 1996; see Mulligan, 1998, for a related discussion). One potential direction, suggested by both Mulligan (1996) and Weldon and Coyote, focuses on the distinction between item-specific and relational processing (see, e.g., Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993). As defined by Hunt and Einstein (1981; Einstein & Hunt, 1980; Hunt & McDaniel, 1993), relational information refers to features shared by to-be-retrieved items, whereas item-specific information refers to features unique to an item. Relational information is important in the selection of effective retrieval strategies and the generation of potential responses, whereas item-specific information provides discriminative or distinctive information during retrieval (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993; Hunt & Seta, 1984; Klein, Loftus, Kihlstrom, & Aseron, 1989). Hunt and Einstein (1981, p. 511) describe memory retrieval as "a process of increasingly finer discriminations," beginning at a coarse level in which relational information delimits the class from which to-be-retrieved responses come and progressing to a more fine-grained level in which item-specific information permits discrimination between candidate responses.

Both Mulligan (1996) and Weldon and Coyote (1996) argued that the item-specific–relational distinction provides a plausible account of the difference between category exemplar production and category-cued recall. Specifically, category-cued recall, which requires the production of words that are both members of the category and were on the study list, requires finer discriminations than does category exemplar production, which simply requires producing exemplars from the category. Consequently, we might expect that category-cued recall is sensitive to both item-specific and relational encodings, whereas category exemplar production is differentially reliant on relational encoding. Presumably, the relevant relational information for the category exemplar production task is category-level information (see, e.g., Rappold & Hashtroudi, 1991).

The findings of Mulligan (1996) and of Weldon and Coyote (1996) indicate that, unlike category-cued recall, category exemplar production is relatively insensitive to encoding effects mediated by item-specific information. In addition, priming in this task appears to be sensitive to relational encodings; generating category labels for studied exemplars increases category exemplar priming (Cabeza, 1994), as does blocking studied examples by category (Rappold & Hashtroudi, 1991), a manipulation that makes relational information highly salient (e.g., Hunt & Einstein, 1981). Finally, conceptual replication during study (i.e., following a target item with an associate) enhances conceptual priming under relational encoding instructions, but not when subjects encoded targets and associates in isolation (McDermott & Roediger, 1996). Thus, the present study is motivated by the notion that conceptual priming in the category exemplar production task is differentially sensitive to relational encodings and relatively insensitive to item-specific encodings, a view we label the relational hypothesis.

This paper investigates the effects of the levels-of-processing (LOP) manipulation in light of the relational hypothesis. In particular, we investigate how priming in the category exemplar production task is affected by simultaneous manipulations of semantic elaboration and list organization. To manipulate semantic elaboration, a traditional LOP manipulation was used, whereby some study items were encoded semantically (using a pleasantness-rating task) and others nonsemantically (using a vowel-counting task). It should be noted that typical semantic encoding tasks, such as the pleasantness-rating task, appear to induce item-specific processing. Consequently, within the item-specific–relational framework, LOP is viewed as a manipulation of item-specific encoding (e.g., Hunt & Einstein, 1981; Hunt & McDaniel, 1993). In contrast, the manipulation of list organization, in which study lists were organized to render the categorical structure of the list either more or less salient, affects relational encoding (e.g., Hunt & Einstein, 1981; Rappold & Hashtroudi, 1991).

A primary implication of the relational hypothesis is that effects of semantic encoding will be greater when encoding takes place in a relational context. In particular, for the category exemplar production task, LOP effects should be greater when category-level information is highly salient, as in the categorized study condition.

An even stronger interpretation of this view suggests that priming in this task is not affected by all forms of semantic encoding, but only by those that emphasize relational information—in this case, categorical information. In particular, if the relevant relational information is not salient and a semantic task emphasizing item-specific information is used (such as the pleasantness-rating task), we may observe no effect of semantic elaboration at all. This strong interpretation is at odds with the TAP view, which makes an unqualified prediction of LOP effects on conceptual tests.

The strong version of the relational hypothesis is immediately challenged by prior findings of LOP effects on the category exemplar production test (Hamann, 1990;
Mecklenbräuker, Wippich, & Mohrhusen, 1996; Srinivas & Roediger, 1990; Weldon & Coyote, 1996). These studies typically used pleasantness (Srinivas & Roediger, 1990; Weldon & Coyote, 1996) or liking ratings (Hamm, 1990; Mecklenbräuker et al., 1996) as the semantic encoding task. According to the strong interpretation of the relational hypothesis, because these tasks emphasize item-specific encoding, they should be less likely to enhance priming in the category exemplar production task. However, there are two potential reasons, consistent with the strong relational hypothesis, that LOP effects were observed. First, the relevant relational information may have been salient at study, inducing relational rather than item-specific encoding.Second, the observed LOP effects may have been due to explicit contamination. Although it is always difficult to rule out this possibility (see the discussion in Roediger & McDermott, 1993), the results of posttest questionnaires indicate that some level of explicit contamination is likely with the category exemplar production task. Of the young adults in Light and Albertson’s (1989) study, 88% reported being aware of the connection between the study and the test portions of the experiment, and 34% reported using intentional retrieval strategies during the implicit test. Likewise, Mulligan and Hartman (1996, Experiment 1, full attention condition) reported that 93% of their subjects were test aware and 50% claimed intentional retrieval. Although posttest questionnaires were not used in the studies showing LOP effects on this task, it is plausible that some level of explicit contamination was present.

We consider these possibilities in detail in the General Discussion section. At present, we discuss these points to introduce additional features of our experimental design. First, to minimize the salience of the categorical structure in the random list condition, we used relatively long study lists in which no two items from the same category appeared in sequence. Second, in an attempt to reduce the impact of test awareness and intentional retrieval, incidental-learning instructions were used. Finally, a posttest questionnaire was used to determine the extent of explicit contamination and to allow for an analysis excluding subjects reporting test awareness (see, e.g., Bowers & Schacter, 1990; Jelicic, Craik, & Moscovitch, 1996; Light & Albertson, 1989; McKone & Sloc, 1997; Mulligan & Hartman, 1996; Thapar & Greene, 1994).

These considerations raise another implication of the relational hypothesis concerning the probable influence of explicit contamination. With explicit contamination, performance in the category exemplar production task will take on characteristics of its explicit counterpart, category-cued recall, rendering the category exemplar production test more sensitive to item-specific encoding. Consequently, reducing explicit contamination by restricting analysis to test-unaware subjects should largely reduce the influence of item-specific encoding on this task. According to the relational view, this has different implications for LOP effects, depending on whether semantic elaboration occurs under conditions of item-specific processing, as in the random condition, or of relational processing, as in the categorized condition. Specifically, eliminating test-aware subjects should reduce the LOP effect in the random condition, but not in the categorized condition, where relational processing predominates. Note that this implication is related to the strong version of the relational hypothesis, which also predicts that eliminating test-aware subjects will reduce and, in fact, eliminate the LOP effect in the random list condition. However, the present implication is more general and may hold even if the strong version of the relational hypothesis does not.

Experiments 1A and 1B proceeded as follows. The subjects studied words from different taxonomic categories. Study items were either organized by category, rendering the categorical structure of the list salient, or presented in a random order, rendering the categorical structure less salient. In addition, each item was encoded with either a semantic (pleasantness-rating) or a nonsemantic (vowel-counting) task. The subjects were tested with either the implicit test of category exemplar production or its explicit counterpart, category-cued recall.

Finally, the LOP manipulation was implemented in two ways. In Experiment 1A, the LOP-encoding conditions were randomly intermixed in the study list (a mixed-list design). In Experiment 1B, the LOP conditions were blocked, so that a subject encoded half the study list using one encoding task and the other half of the list using the other encoding task (a blocked-list design). Prior research indicates that experimental design determines whether the LOP manipulation affects performance on priming tasks. In particular, LOP typically only affects implicit tests when conditions are blocked, but LOP effects are found in both mixed and blocked conditions for explicit tests (e.g., Challis & Brodeck, 1992; Thapar & Greene, 1994). Prior studies of the LOP effect on the category exemplar production task have used either the within-subjects blocked-list design (Hamm, 1990; Srinivas & Roediger, 1990; Weldon & Coyote, 1996) or a between-subjects design (Mecklenbräuker et al., 1996). To determine whether the effects of LOP are robust on this conceptual implicit test and as a check on the generality of the predicted interaction between LOP and the relational manipulation of list organization, the LOP variable was implemented in a mixed-list design (Experiment 1A) and a blocked-list design (Experiment 1B).

**EXPERIMENTS 1A AND 1B**

**Method**

**Subjects**

One hundred forty-four undergraduates at Illinois State University participated in exchange for extra credit in psychology courses. Seventy-two subjects participated in Experiment 1A and seventy-two in Experiment 1B.

**Design and Materials**

Both experiments used a $2 \times 2 \times 2$ design, in which LOP (semantic: pleasantness rating vs. nonsemantic: vowel counting) was manipulated within subjects and list organization (categorized vs. random)
and type of memory test (category exemplar production vs. category-cued recall) were manipulated between subjects. In addition, in Experiment 1A, in which study trials were blocked by LOP encoding condition, the order of the encoding blocks (pleasantness rating first vs. vowel counting first) was counterbalanced across subjects.

Six examples were chosen from each of 16 categories (sports, fruits, furniture, birds, colors, four-footed animals, articles of clothing, trees, musical instruments, parts of a human body, vegetables, dances, insects, substances for flavoring food, fish, and parts of a building) in the Batig and Montague (1969) norms. The items were not in the top ten most frequently obtained instances, but all were produced by at least 10 of the 400 subjects in the Batig and Montague norms. The average rank of the critical items was 17.4. In addition, eight filler words were selected from previously unused categories and had an average rank of 13.5.

Two master lists of 48 critical items were formed by randomly dividing the 16 categories into two groups. In Experiment 1A, each master list was used to generate a random and a categorized list (note that this arrangement counterbalances items across the categorized–random manipulation). In the random condition, the critical items were randomly ordered, subject to the constraint that no two examples from the same category appeared in sequence. In the categorized condition, all the examples from a category appeared in sequence. Within each list, items were randomly assigned to LOP encoding condition, subject to the constraint that no more than 2 consecutive items appeared in the same encoding condition. Two versions of each of the random and categorized lists were created, to counterbalance items across the two LOP conditions. Four filler words were added to the beginning and four to the end of each list, serving as primacy and recency buffers.

For Experiment 1B, the study lists were created as follows. Each master list was divided into two halves, each containing all the examples from four categories (i.e., each half contained 24 critical items). As in Experiment 1A, each master list was used to generate a random list and a categorized list. In the random list, the critical items within each half-list were randomly ordered, subject to the constraint that no two examples from the same category appeared in sequence. In the categorized list, all the examples from a category appeared in sequence. All of the items within a half-list were assigned to the same LOP condition. Across subjects, the order of the instruction blocks varied. For half the subjects, the pleasantness-rating task was performed on the first half-list, and the vowel-counting task was performed on the second half-list. For the other half of the subjects, the order was reversed. Within each half-list, two filler words were added to the beginning and two to the end, serving as primacy and recency buffers.

For both Experiment 1A and Experiment 1B, the set of items presented at study served as the old items on the subsequent memory test. The nonstudied set of critical items served as the baseline, or new condition. Study lists were distributed across subjects, so that each item was used equally often in the old and new conditions, and when old, equally often in the two LOP conditions. Equal numbers of subjects were in the categorized and random study conditions.

All 16 category names were used as cues in the category exemplar production test. Eight of the category cues corresponded to studied items, and eight corresponded to the critical new items. Only the 8 studied category names were presented in the cued-recall test.

**Procedure**

The subjects were tested individually. Both experiments consisted of three parts: a study task, a distractor task, and a memory test. Prior to starting the study task, the subjects were told that this was a study of word perception and problem solving. They were told that they would be presented with a series of words on a computer screen. For half of the words, the subjects were asked to rate the pleasantness of the word on a 5-point scale, typing the 1 key for unpleasant, the 2 key for neutral, and the 3 key for pleasant. For the other half of the words, the subjects were asked to count the number of vowels in the word, typing the corresponding number on the computer's keyboard. In Experiment 1A, the subjects were told that the pleasantness-rating and vowel-counting trials would be interleaved, with the task on any particular trial signaled just prior to the display of the study item. In this experiment, each study trial consisted of the following. First, the word PLEASANT or VOWEL was presented at the top of the computer screen, informing the subject of the task for the upcoming word. At the bottom of the screen was printed a reminder of how to respond for this type of trial (i.e., the pleasantness scale or a reminder of the vowel-counting task). This message was displayed for 1 sec prior to the presentation of the study word and stayed on the screen through the presentation of the study word. Next, the study word was presented in the center of the screen for 4.5 sec. The subjects were instructed to respond within the time that the word was on the screen.

In Experiment 1B, the subjects were told that they would perform one of the tasks (either the pleasantness-rating or the vowel-counting task, depending on the order of encoding) for the first half of the words, and the other task for the second half. The study trials of Experiment 1B were identical to those of Experiment 1A, with the exception that, within a block of trials, the instruction signal (either PLEASANT or VOWEL) did not change. At the halfway point of the study list, a message was displayed indicating that the encoding task would now change to the second task. The experimenter reiterated the message before the subject began the second block of study trials.

Immediately following the study task, the subjects were given a 3-min distractor task. The distractor task was used to encourage the subjects to view the subsequent implicit memory test as one in a series of unrelated tasks. In this task, the subjects were presented with a series of arithmetic problems on a sheet of paper. They were asked to solve each problem mentally, without making intermediate calculations on the sheet, and to write the answer upon completion. Following the distractor task, the subjects were given one of two memory tests, either the category exemplar production test or the category-cued recall test. The subjects in the implicit condition were told that the category exemplar production test was part of a norming study and that their responses would be used for the development of materials for upcoming experiments. The subjects were presented with the 16 category names, 1 at a time, on a computer screen. They were instructed to produce eight examples belonging to the category as fast as possible. The experimenter recorded the responses and signaled the subject to continue to the next category after eight examples were produced. The subject controlled the presentation of the category cues by pressing the spacebar. No mention was made of the relationship between this task and the study task. After completing the category exemplar production task, the experimenter asked the subject a series of increasingly specific questions (the awareness questionnaire presented in the Appendix) to assess whether the subject was aware of the relationship between the study and the test tasks and the extent to which he or she engaged in deliberate retrieval. To be classified as test aware, the subject had to indicate that he or she was aware of the study-test relationship during the test phase (if test aware, the subject was also classified as engaging in intentional retrieval or not, although these data were not used in subsequent analyses). The questionnaire and the classification procedure were modeled on that used by Bowers and Schauzler (1990) and Mulligan and Hartman (1996).

In the category-cued recall test, the subjects were presented with the eight category cues corresponding to words from the study list. The cues were displayed one at a time on the computer screen. The subjects were asked to use the cue to try to remember words from the study list and were informed that more than one example from each category had appeared on the study list. Consequently, they were informed to try to remember as many words as possible for each
category cue. The experimenter recorded the subjects' responses. No time limit was imposed, and the number of responses was not fixed.

Results and Discussion

Table 1 summarizes the results of Experiments 1A and 1B. For the category exemplar production task, each of the groups contained 24 subjects. For the category-cued recall test, each group contained 12 subjects.

Category Exemplar Production

The category exemplar production task produced several important results. First, the LOP manipulation produced the expected effect: Semantic encoding increased conceptual priming. Second, the size of the LOP effect was comparable in Experiments 1A and 1B, indicating that the LOP manipulation produced comparable effects whether implemented in a blocked list (Experiment 1B) or a mixed list (Experiment 1A). Finally, the effect of LOP was larger for categorized than for random lists. This finding is consistent with the relational view and indicates that the effects of semantic encoding are enhanced by a relational context.

Statistical analyses confirm the foregoing summary (a significance level of .05 was adopted for all analyses). We began with a preliminary analysis, to determine whether the order of the encoding blocks had an effect on priming in Experiment 1B. Priming scores were submitted to a 2 × 2 × 2 analysis of variance (ANOVA), using LOP (pleasantness rating vs. vowel counting) as a within-subjects factor and list organization (categorized vs. random) and the order of the encoding blocks (pleasantness rating first vs. vowel counting first) as between-subjects factors. Order of encoding blocks produced no significant effect or interaction (all ps > .20). Consequently, the Experiment 1B data were collapsed over this variable in Table 1 and in subsequent analyses.

Primbing scores for both Experiment 1A and Experiment 1B were submitted to a 2 × 2 × 2 ANOVA, using LOP as a within-subjects factor and list organization and experiment as between-subjects factors. The analysis revealed two significant effects: the main effect of LOP \( [F(1,92) = 40.37, M_S^2 = 0.0097] \), indicating greater priming in the semantic than in the nonsemantic task, and the interaction of LOP and list organization \( [F(1,92) = 4.26, M_S^2 = 0.0097] \), indicating a larger LOP effect in the categorized than in the random condition. The main effect of organization approached significance \( (p < .09) \). No other effects approached significance \( (ps > .25) \).

To further investigate the LOP × list organization interaction, priming scores from the categorized and random conditions were analyzed with separate 2 × 2 ANOVAs, using LOP as a within-subjects factor and experiment as a between-subjects factor. These analyses indicated significant effects of LOP in both the categorized \( [F(1,46) = 40.45, M_S^2 = 0.0085] \) and the random \( [F(1,46) = 8.19, M_S^2 = 0.0109] \) conditions. In a pair of complementary analyses, the priming scores from the pleasantness-rating and vowel-counting conditions were analyzed with separate 2 × 2 ANOVAs, using list organization and experiment as between-subjects factors. These analyses indicated that list organization had a significant effect in the semantic condition \( [F(1,96) = 5.54, M_S^2 = 0.0157] \), but not in the nonsemantic condition \( (p > .50) \). Consistent with the initial analysis, there were no significant effects of experiment in any of these analyses \( (ps > .18) \).

The proportions of critical new items produced were submitted to a 2 × 2 ANOVA, using list organization and experiment as between-subjects factors. The analysis revealed no significant effects (all ps > .50). In an additional set of analyses, we examined whether significant priming was obtained. Collapsing across experiments, significant priming was found in each of the four combined LOP organization conditions (i.e., semantic–categorized, semantic–random, nonsemantic–categorized, and nonsemantic–random; all ps < .02).

There are several important implications of these findings. First, consistent with other studies (e.g., Hamann, 1990; Mecklenbrauker et al., 1996; Srinivas & Roediger, 1990), we found that semantic encoding enhanced conceptual priming on the category exemplar production task. Second, the effects of LOP are comparable, whether a

<table>
<thead>
<tr>
<th>Experiment</th>
<th>List Organization</th>
<th>Semantic</th>
<th>Nonsemantic</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category Exemplar Production</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A (mixed-list design)</td>
<td>blocked</td>
<td>.27</td>
<td>.16</td>
<td>.13</td>
</tr>
<tr>
<td>1B (blocked-list design)</td>
<td>blocked</td>
<td>.31</td>
<td>.18</td>
<td>.14</td>
</tr>
<tr>
<td>Category Cued Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A (mixed-list design)</td>
<td>blocked</td>
<td>.41</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>1B (blocked-list design)</td>
<td>blocked</td>
<td>.45</td>
<td>.08</td>
<td></td>
</tr>
</tbody>
</table>
mixed-list design (Experiment 1A) or a blocked-list design (Experiment 1B) is used. This contrasts with findings from other implicit tests (word-fragment completion, perceptual identification, general knowledge questions), in which the experimental design is an important determinant of LOP effects (Challis & Brodbeck, 1992; Thapar & Greene, 1994), a point to which we return in the General Discussion section.

The most important result for present purposes is that the size of the LOP effect is larger for categorized than for random list organization. As is predicted by the relational view, the effects of semantic elaboration on conceptual priming are greater when the encoding takes place in a relational context. The complementary description of the LOP × list organization interaction is that the effects of organization were only found in the semantic encoding condition. Consequently, the semantic condition replicates the finding of Rappold and Hashtroudi (1991), who also found that categorized study lists enhanced conceptual priming. However, the present results imply that the positive effects of organization are limited to semantic encoding conditions and may not appear for non-semantic encodings.

**Analysis of test-unaware subjects.** We categorized subjects as test aware or test unaware on the basis of the posttest questionnaire (see the results of Experiment 3 for evidence of high interrater reliability for this classification). Table 2 summarizes these results.

Chi-square analyses indicate that the proportion of test-unaware subjects significantly differed across list organization conditions \[ \chi^2(1) = 7.07 \], indicating a greater proportion in the random list condition, but not across experiments \( (p > .15) \). Priming scores of the test-unaware subjects were submitted to the same analyses as above, beginning with a 2 × 2 × 2 ANOVA, using LOP as a within-subjects factor and list organization and experiment as between-subjects factors. Consistent with the initial analysis, the reanalysis revealed two significant effects: the main effect of LOP \( F(1,41) = 8.23, M_{SE} = 0.0076 \) and the interaction of LOP and list organization \( F(1,41) = 7.87, M_{SE} = 0.0076 \), indicating a larger LOP effect in the categorized than in the random condition. No other effects approached significance \( (ps > .35) \). Follow-up analyses indicated a significant effect of LOP in the categorized list condition \( F(1,14) = 65.85, M_{SE} = 0.0014 \), but not in the random list condition \( (p > .50) \). Furthermore, list organization had a significant effect in the semantic condition \( F(1,41) = 4.35, M_{SE} = 0.0137 \), but not in the nonsemantic condition \( (p > .20) \).

Collapsing across experiments, significant priming was found in the semantic–categorized condition \( F(1,14) = 35.67, M_{SE} = 0.0034 \) and in the nonsemantic–random condition \( F(1,27) = 7.12, M_{SE} = 0.0037 \). In addition, the priming effect approached the traditional significance level in the semantic–random condition \( (p < .08) \), but not in the nonsemantic–categorized condition \( (p > .50) \). An analysis of the new items revealed no significant differences as a function of list organization, experiment, or their interaction \( (all ps > .25) \).

There are several important aspects of the analysis of the test-unaware subjects. First, the larger LOP effect in the categorized condition indicates that the enhanced effects of LOP under relational encoding are probably not due to explicit contamination. In addition, the effects of organization are still confined to the semantic encoding condition. Second, and again consistent with the initial analysis, the effects of LOP are comparable across experiments. Even among the test-unaware subjects, the effects of LOP on category exemplar production are not measurably affected by the experimental design. The analysis of the test-unaware subjects did differ from the initial analysis in one regard: There was no measurable effect of LOP in the random list condition. This intriguing finding must be treated cautiously, because it appears to conflict with reports of LOP effects on category exemplar production under conditions similar to the present random condition (e.g., Hamann, 1990; Srinivas & Roediger, 1990). However, this finding does comport well with the general notion that there are multiple forms of conceptual processing (e.g., Cabeza, 1994; McDermott & Roediger, 1996) and with the more specific notion that conceptual priming in the category exemplar production task is relatively insensitive to item-specific encodings (Mulligan, 1996; Weldon & Coyote, 1996). As was noted in the introduction, the standard LOP manipulation, using pleasantness (or liking) rating

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiments 1A and 1B: Category Exemplar Production Data as a Function of Test Awareness</strong></td>
</tr>
<tr>
<td>Experiment</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>1A (mixed-list design)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1B (blocked-list design)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
as the semantic encoding condition, is viewed as affecting primarily item-specific or distinctive encoding processes (e.g., Hunt & Einstein, 1981; Hunt & McDaniel, 1993). To the extent that the conceptual processes engaged do not involve relational information relevant to the retrieval task (in this case, information about superordinate categories), variation in such conceptual processing may well leave priming unaffected. This raises the possibility that the apparent LOP effects observed in nonrelational encoding conditions (e.g., the present random list condition) may be carried by test-aware subjects.

Given the potential importance of this finding, it is important to assess the power of detecting the LOP effect in the random list condition for the test-unaware subjects. To obtain a reasonable value for the to-be-detected effect size, we estimated the LOP effect size for the test-unaware subjects in the categorized list condition and halved it. Combined across Experiments 1A and 1B, the LOP effect size in the categorized list condition is \( d = 1.69 \). The power to detect an effect half this size, with \( n = 29 \) (the number of test-unaware subjects in the random condition) and \( \alpha = .05 \) (one-tailed), is .93 (Cohen, 1988). Consequently, the analysis of the LOP effect for the test-unaware subjects in the random condition has substantial power to detect an effect of LOP as small as half that observed in the categorized list condition.

**Comparison of test-aware and test-unaware subjects.** According to the relational view, the effects of semantic elaboration are predominately item specific in the random condition and relational in the categorized condition. Furthermore, eliminating explicit contamination should reduce the impact of item-specific information on the category exemplar production test. Consequently, eliminating test-awareness should reduce the size of the LOP effect in the random list condition, but may have no impact in the categorized condition. To examine these predictions, priming data from the categorized and random conditions were submitted to separate 2 × 2 × 2 ANOVAs, using LOP as a within-subjects factor and experiment and test awareness (aware vs. unaware) as between-subjects factors. The results were consistent with the expectations of the relational view. In the categorized condition, the LOP × test awareness interaction was not significant (\( F < 1 \)), indicating comparable LOP effects for test-aware and test-unaware subjects [the main effect of LOP was significant, \( F(1,44) = 33.43, MS_e = 0.0086; \) no others were, \( p > .15 \)]. In the random condition, the LOP × test awareness interaction was significant [\( F(1,44) = 15.03, MS_e = 0.0084 \)], indicating a smaller LOP effect for the test-unaware than for the test-aware subjects (the main effect of LOP was significant [\( F(1,44) = 15.51, MS_e = 0.0084; \) no others were, \( p > .10 \)]. Not only is the LOP effect reduced for test-unaware subjects in the random condition, but as the immediately preceding analysis showed, the effect is rendered nonsignificant.

**Category-Cued Recall**

In a preliminary analysis of Experiment 1B, a 2 × 2 × 2 ANOVA, using LOP as a within-subjects factor and list organization and the order of the encoding blocks as between-subjects factors, revealed that order was not involved in any significant effects (all \( p > .30 \)). Thus, in subsequent analyses, the data were collapsed over this variable.

In the combined analysis of Experiments 1A and 1B, recall proportions were submitted to a 2 × 2 × 2 ANOVA, using LOP as a within-subjects factor and list organization and experiment as between-subjects factors. The analysis revealed a main effect of LOP [\( F(1,44) = 167.58, MS_e = 0.0166 \)], indicating greater recall] in the semantic than in the nonsemantic condition. The effect of list organization was not significant (\( p = .30 \)), nor were any other effects (\( p > .15 \)). The mean number of intrusions for Experiment 1A in the random and categorized conditions was 1.08 and 1.25, respectively; the corresponding means for Experiment 1B were 1.07 and 0.67. A 2 × 2 × 2 ANOVA on the intrusion scores (using experiment and list organization as factors) revealed no significant effects (\( p > .30 \)). The intrusions tended to be higher frequency examples of the category cues. None of the intrusions was a critical new item.

The effects of LOP are pervasive on explicit tests and were expected here (see, e.g., Challis, Velchovsky, & Craik, 1996; Thapar & Greene, 1994). However, the null effect of list organization was unexpected and appears to conflict with the results of Rappold and Hashtroudi (1991), who found robust effects of list organization on category-cued recall.

This is especially surprising because we used similar numbers of subjects, used similar numbers of study items, and implemented the list organization manipulation in a manner similar to that in the Rappold and Hashtroudi (1991) study (their Experiments 1 and 2). The total number of observations (subjects × study items) making up each of the categorized and random conditions were 576 and 432 in Rappold and Hashtroudi’s Experiments 1 and 2, respectively. In the present experiment, the total number of observations in each of the categorized and random conditions (for the semantic encoding condition only, collapsed over Experiments 1A and 1B) was 576. Despite the approximately equal number of observations per condition, Rappold and Hashtroudi found robust effects of organization on category-cued recall (ranging in size from 12 to 15 percentage points), and the present experiment did not. It appears unlikely, therefore, that the difference in results is simply due to power differences. A potentially important difference between the Rappold and Hashtroudi study and the present experiment is that the subjects in the Rappold and Hashtroudi experiments studied all of the words under constant (intentional) encoding instructions, whereas the present subjects studied words under varying encoding instructions. Switching from one encoding task to another might reduce the salience of the categorical organization in the categorized study lists, reducing the strength of the organizational manipulation.

In the next experiment, the subjects studied categorized and random lists under constant (semantic or nonsemantic...
tic) encoding instructions. To the extent that varying the encoding instructions within the study lists reduces the salience of category structure in the categorized condition, Experiment 2 would provide a stronger manipulation of organization. Consequently, we expected the effects of organization to emerge on the category-cued recall test (at least under semantic encoding). Experiment 2 also included the implicit test of category exemplar production, to determine whether the present pattern of results is obtained with a between-subjects manipulation of LOP. In addition, it is important to verify the present results with a stronger manipulation of organization, a manipulation that produces the standard effects on the explicit test of category-cued recall.

EXPERIMENT 2

Method

Subjects
One hundred and ninety-two undergraduates at Illinois State University and Southern Methodist University participated in exchange for extra credit in psychology courses.

Design and Materials
The experiment used a $2 \times 2 \times 2$ design in which LOP (semantic: pleasantness rating vs. nonsemantic: vowel counting), list organization (categorized vs. random), and type of memory test (category exemplar production vs. category-cued recall) were manipulated between subjects.

The 96 critical items from Experiments 1A and 1B were divided into two sets, and each set was used to construct a categorized and a random list. Two filler words were added to the beginning and two to the end of each study list. The resulting four lists were presented to equal numbers of subjects.

Procedure
The subjects were tested in groups of 1–5. Prior to the presentation of the study list, the subjects were informed that the experiment was a study of word perception and problem solving. Study items were presented on $3 \times 5$ in. index cards, one word to a card. The cards were arranged in a stack, with a cover card on top. The subjects were asked to either rate the pleasantness or count the number of vowels in each word. The pleasantness scale varied from 1 (very unpleasant) to 5 (very pleasant). The subjects were given 6 sec per word to perform the task. Time was kept with a tape recording that emitted a beep every 6 sec. The subjects were instructed to flip over the cover card, exposing the first word in the list, upon hearing the first beep. The subjects were asked to spend the entire 6 sec on the present word and not to move on to the next word until hearing the next beep. The subjects were then to turn the card face down, exposing the next word in the list. Thus, the subjects were presented with a single word for the entire 6-sec presentation time. The subjects were instructed to record their rating/vowel count on the provided answer sheet. The experimenter observed the subjects, to ensure that the instructions were followed.

Immediately following the study task, the subjects were given a 3-min distractor task. The distractor task was a stem completion task with city names. Each subject was given a sheet of paper with 80 three-letter word stems that could be completed with the names of U.S. cities (e.g., bos... for Boston). The subjects were asked to complete as many stems as possible with the names of U.S. cities in the allotted 3 min.

After the distractor task, the subjects were presented with a test booklet for either the category exemplar production test or the category-cued recall test. For the category exemplar production test, all 16 categories were used (corresponding to the 8 studied and 8 nonstudied categories). Each page of the booklet contained the name of 1 category and eight numbered blank spaces. As in the prior experiment, the subjects in the implicit condition were told that this was a norming study for the development of experimental materials. The subjects were asked to write down eight examples from the category as rapidly as possible. For the category-cued recall test, only the 8 studied categories were presented. Each page of the test booklet contained a category name; the rest of the page was blank. The subjects were asked to use each category name to remember (and write down) as many examples as possible from the study list. They were informed that more than one example from each category had been presented. For both test conditions, the subjects were given 45 sec to complete each category. The experimenter kept time and signaled the subject (by saying "Next") when it was time to proceed to the next category. The subjects in the implicit condition were then presented with the awareness questionnaire on a sheet of paper.

Results and Discussion
Table 3 summarizes the results. For the category-cued recall test, each of the groups contained 18 subjects. For the category exemplar production test, each group contained 30 subjects.

The category-cued recall data suggest that we were successful at more strongly manipulating list organization: List organization had a pronounced effect on recall in the semantic-encoding condition (but not in the nonsemantic condition). In addition, semantic encoding enhanced recall in both the categorized and the random list conditions. A $2 \times 2$ ANOVA, using LOP and list organization as between-subjects factors, verifies this summary. The analysis revealed a main effect of LOP [$F(1,68) = 183.93$,

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 2: Mean Proportion of Words Recalled or Produced as a Function of Memory Test, List Organization, and Encoding Condition</td>
</tr>
<tr>
<td>Encoding Condition</td>
</tr>
<tr>
<td>Old</td>
</tr>
<tr>
<td>Category exemplar production blocked</td>
</tr>
<tr>
<td>random</td>
</tr>
<tr>
<td>Category cued recall blocked</td>
</tr>
<tr>
<td>random</td>
</tr>
</tbody>
</table>
a main effect of list organization \( [F(1, 68) = 7.06, MS_e = 0.0107] \), and a significant interaction \( [F(1, 68) = 5.68, MS_e = 0.0107] \). Follow-up analyses indicated a significant effect of organization in the semantic \( [F(1, 34) = 8.29, MS_e = 0.0164] \), but not in the nonsemantic \( (p > .50) \) conditions. Furthermore, LOP affected performance in both the categorized and the random conditions \( [F(1, 34) = 130.57, MS_e = 0.0104] \, and \, F(1, 34) = 60.87, MS_e = 0.0110, \) respectively. The mean number of intrusions on the recall test was 2.11, 2.39, 3.56, and 4.67 for the semantic–categorized, semantic–random, nonsemantic–categorized, and nonsemantic–random conditions, respectively. A 2 × 2 ANOVA on the intrusion scores (using LOP and list organization as factors) revealed a significant effect of LOP \( [F(1, 68) = 8.72, MS_e = 7.154] \), indicating more intrusions after nonsemantic than after semantic encoding. No other effects were significant \( (p > .25) \). As in Experiments 1A and 1B, the intrusions tended to be higher frequency examples of the category cues, and none of the intrusions was a critical new item.

Consistent with other research, these results demonstrate a robust effect of organization on category-cued recall (e.g., Rappold & Hashtroudi, 1991). A comparison of the present results with those of Experiments 1A and 1B implies that the strength of the organizational manipulation is enhanced when subjects are not required to vary the encoding task across items. In addition, the effects of organization on category-cued recall may be more readily observed under semantic or intentional encoding instructions (Rappold & Hashtroudi, 1991) than under nonsemantic encoding instructions. This same result was found with the category exemplar production task in the prior experiment and, as is discussed below, in the present experiment.

Turning next to the category exemplar production data, we see that the present study replicates Experiments 1A and 1B in all important regards. First, conceptual priming in this task is enhanced by semantic encoding. Second, the effects of LOP are greater in the categorized than in the random condition. Third, the effects of organization are limited to the semantic encoding condition. Fourth, the analysis of test awareness also replicates the results of Experiments 1A and 1B in detail.

Priming scores were submitted to a 2 × 2 ANOVA, using LOP and list organization as between-subjects factors. All three effects were significant: the main effect of LOP \( [F(1, 116) = 40.62, MS_e = 0.0094] \), the main effect of organization \( [F(1, 116) = 5.21, MS_e = 0.0094] \), and a LOP × list organization interaction \( [F(1, 116) = 6.77, MS_e = 0.0094] \). Follow-up analyses indicate that semantic encoding enhanced priming in both the categorized and the random conditions \( [F(1, 58) = 40.48, MS_e = 0.0093] \), and \( F(1, 58) = 7.08, MS_e = 0.0093, \) respectively. Furthermore, list organization only affected priming in the semantic condition \( [F(1, 58) = 8.60, MS_e = 0.0130] \), and not in the nonsemantic condition \( (p > .50) \). Additional analyses indicate that performance on new items did not vary as a function of LOP, list organization, or their interaction (all \( ps > .10 \)) and that significant priming was found in each experimental group (all \( ps < .02 \).

As in Experiments 1A and 1B, the subjects were categorized as test aware or test unaware on the basis of the posttest questionnaire (Table 4). Chi-square analyses indicated significantly fewer test-unaware subjects in the semantic–categorized condition than in any of the other conditions \( (ps < .04) \); the other conditions did not significantly differ \( (p > .20) \). Priming scores of the test-unaware subjects were submitted to a 2 × 2 ANOVA, using LOP and list organization as between-subjects factors. This analysis revealed significant main effects of LOP \( [F(1, 58) = 19.35, MS_e = 0.0046] \), and of list organization \( [F(1, 58) = 11.69, MS_e = 0.0046] \). In addition, the interaction of LOP and list organization was significant \( [F(1, 58) = 15.15, MS_e = 0.0046] \), indicating a larger LOP effect in the categorized than in the random condition. Follow-up analyses indicated a significant effect of LOP in the categorized list condition \( [F(1, 26) = 58.36, MS_e = 0.0022] \), but not in the random list condition \( (p < 1) \). Furthermore, list organization had a significant effect in the semantic condition \( [F(1, 20) = 15.29, MS_e = 0.0059] \), but not in the nonsemantic condition \( (p < 1) \).

As in Experiments 1A and 1B, it is quite surprising that the test-unaware subjects in the random list condition showed no LOP effect. Again, it is important to determine the power of this analysis. A LOP effect size was estimated from the test-unaware subjects in the categorized list condition, yielding \( d = 3.57 \). The power to detect an effect half this size with \( n^* = 16.8 \) (the harmonic mean of the test-unaware samples sizes of the semantic–random and nonsemantic–random groups; see Cohen,

### Table 4

**Experiment 2: Category Exemplar Production Data as a Function of Test Awareness**

<table>
<thead>
<tr>
<th>List Organization</th>
<th>Test Awareness</th>
<th>Old</th>
<th>New</th>
<th>Priming (Old–New)</th>
<th>Old</th>
<th>New</th>
<th>Priming (Old–New)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocked</td>
<td>aware</td>
<td>.31</td>
<td>.12</td>
<td>.19</td>
<td>.13</td>
<td>.07</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 23)</td>
<td></td>
<td></td>
<td>(n = 9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>unaware</td>
<td>.29</td>
<td>.11</td>
<td>.18</td>
<td>.13</td>
<td>.11</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 7)</td>
<td></td>
<td></td>
<td>(n = 21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td>aware</td>
<td>.26</td>
<td>.09</td>
<td>.17</td>
<td>.16</td>
<td>.11</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 15)</td>
<td></td>
<td></td>
<td>(n = 11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>unaware</td>
<td>.14</td>
<td>.10</td>
<td>.04</td>
<td>.14</td>
<td>.11</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 15)</td>
<td></td>
<td></td>
<td>(n = 19)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1988, Equation 2.3.1, p. 42) and alpha = .05 (one-tailed) exceeds .99. The power to detect an effect one-third this size is still very high at .95.

Performance on new items did not vary as a function of LOP, list organization, or their interaction (all ps > .50). Significant priming was found in the semantic-categorized group \(F(1,6) = 154.64, MS_\epsilon = 0.0010\). Because priming in the other three groups did not significantly differ (pairwise comparisons yielded ps > .50), these groups were combined to determine whether the low levels of priming significantly exceeded zero. Combining across these groups yielded significant priming \(F(1,54) = 12.04, MS_\epsilon = 0.0024\).

Next, we assessed the prediction that eliminating test-aware subjects significantly reduces the LOP effect in the random condition, but not in the categorized condition. Priming scores from the categorized and random conditions were submitted to separate 2 x 2 ANOVAs, using LOP and test awareness as between-subjects factors. In the categorized condition, the LOP x test awareness interaction was not significant \(F < 1\), indicating comparable LOP effects for test aware and test-unaware subjects [the main effect of LOP was significant, \(F(1,56) = 26.70, MS_\epsilon = 0.0095\); the effect of test-awareness was not, \(F < 1\)]. In the random condition, the LOP x test awareness interaction was significant \(F(1,56) = 5.75, MS_\epsilon = 0.0076\), indicating a smaller LOP effect for the test-unaware than for the test-aware subjects [the main effects of LOP and test awareness were also significant, \(F(1,56) = 7.92, MS_\epsilon = 0.0076\); and \(F(1,56) = 9.82, MS_\epsilon = 0.0076\), respectively].

**EXPERIMENT 3**

Experiment 3 was conducted to determine whether the predictions of the relational view are replicated using a variant of the category exemplar production task developed by Srinivas and Roediger (1990). In this version of the task, a single example from each category is presented during encoding. Consequently, there is no categorial structure of the study materials to bias encoding toward relational information (i.e., encoding is assumed to be dominated by item-specific processing). Under these conditions, the relational view suggests that test awareness should interact with LOP, so that the LOP effect is reduced for test-unaware subjects. In addition, the strong version of the relational view posits that, because semantic elaboration is item specific, the LOP effect should be non-significant for test-unaware subjects. Experiment 3 is modeled after Srinivas and Roediger’s Experiment 2 (visual study condition).

**Method**

**Subjects**

Sixty undergraduates at Southern Methodist University participated in exchange for extra credit in psychology courses.

**Design and Materials**

The present experiment used a blocked-list manipulation of LOP. Consequently, the experimental design was 2 x 2, in which LOP (semantic: pleasanness rating vs. nonsemantic: vowel counting) was manipulated within subjects and the order of the encoding blocks (pleasanness rating first vs. vowel counting first) was manipulated between subjects.

The materials were those of Srinivas and Roediger (1990, Experiment 2), listed in their Appendix B, and consisted of 60 critical examples from 60 different categories, drawn from the norms of K. P. Hunt and Hodges (1971) and Battig and Montague (1969). The examples were selected to be low- to medium-frequency associates of the category name. Following Srinivas and Roediger, the critical words were randomly divided into five sets of 12 examples each. Five study lists were created by assigning four of the sets to studied conditions (two sets each to the semantic and nonsemantic encoding conditions) and the fifth set to the new (nonstudied) condition. This last set was not presented at study and was used to measure baseline responding at test. Across study lists, each set was assigned to the semantic condition twice, the nonsemantic condition twice, and the new condition once. Study lists were presented to equal numbers of subjects. The category exemplar production task consisted of 70 category names. Sixty of the names corresponded to the critical examples, 24 in the semantic condition, 24 in the nonsemantic condition, and 12 in the new condition. The remaining 10 category names were filler items. Thus, the ratio of studied:nonstudied items was 48:22 (Srinivas & Roediger, 1990, p. 398).

**Procedure**

The procedure was the same as that in Experiment 2, with the following exceptions. At study, the subjects were told that they would perform one task (either pleasanness rating or vowel counting) for the first half of the words and then switch to the other task for the second half. At the midpoint of the list, the subjects were informed that it was time to switch encoding tasks and were reminded which encoding task to perform for the second half of the list. The pleasanness scale varied from 1 (very unpleasann) to 7 (very pleasann). After the study task, the subjects were presented with a test booklet for the category exemplar production task (as in Srinivas & Roediger, 1990, there was no distractor task). The test booklet consisted of the 70 category names, each numbered and followed by four blank lines. The subjects were asked to write down as many examples of the category as possible. The subjects were given 20 sec for each category name. Time was kept with a tape recording that emitted a beep every 20 sec. After the test, the subjects were presented with a sheet containing the awareness questionnaire.

**Results and Discussion**

A preliminary analysis of the category exemplar production data indicates that the order of the encoding blocks produced neither a main effect nor an interaction with LOP \(F < 1\). Consequently, subsequent analyses collapse over this factor. The mean proportions of critical items produced in the semantic, nonsemantic, and new conditions were .30, .19, and .15, respectively. An analysis of the priming scores indicates a significant effect of LOP \(F(1,59) = 42.37, MS_\epsilon = 0.0086\). In addition, priming was significantly greater than zero in both the semantic \(F(1,59) = 61.40, MS_\epsilon = 0.0107\) and the

<table>
<thead>
<tr>
<th>Encoding Condition</th>
<th>Test Awareness</th>
<th>Semantic</th>
<th>Nonsemantic</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aware (n = 38)</td>
<td>.34</td>
<td>.20</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Unaware (n = 22)</td>
<td>.23</td>
<td>.19</td>
<td>.14</td>
<td></td>
</tr>
</tbody>
</table>
nonsemantic \( F(1.59) = 5.15, MS_e = 0.0083 \) encoding conditions. The present experiment closely replicates the results of the Srinivas and Roediger (1990) study on which it was based, both in finding an effect of LOP and in the quantitative levels of priming. The semantic and nonsemantic conditions produced mean priming of .15 and .04 in the present experiment and of .14 and .03 in Srinivas and Roediger (Experiment 2, averaged over presentation modality).

As in the earlier experiments, the subjects were categorized as test aware or test unaware on the basis of the posttest questionnaire (Table 5). Before analyzing the data, we first assessed the interrater reliability of the classification. The questionnaire was designed to minimize potential effects of experimenter demands on reported test awareness. Consequently, the questionnaire consisted of a series of increasingly specific questions, from which the experimenter inferred whether the subject was aware, at the time of the test, of the study–test relationship. Two of the experimenters independently classified each of the 60 subjects as test aware or test unaware from their written responses. There was 100% agreement on the classifications.

Priming scores from the categorized and random conditions were submitted to separate \( 2 \times 2 \) ANOVAs, using LOP and test awareness as between-subjects factors. This analysis revealed a significant main effect of LOP \( F(1.58) = 34.65, MS_e = 0.0076 \) and a significant LOP \( \times \) test awareness interaction \( F(1.58) = 8.79, MS_e = 0.0076 \), indicating a reduced LOP effect for test-unaware subjects. The effect of LOP was significant for the test-aware subjects \( F(1.37) = 26.64, MS_e = 0.0308 \). Although there may be a trend for a LOP effect with the test-unaware subjects, it was not significant \( (F = 2.20, p > .15) \). In addition, the effect of test awareness was significant in the semantic condition \( F(1.58) = 4.33, MS_e = 0.0068 \), but not in the nonsemantic condition \( (F < 1) \). Finally, for the test-unaware subjects, priming (averaged over the semantic and nonsemantic conditions) was significantly greater than zero \( F(1.21) = 9.94, MS_e = 0.0029 \).

The critical results from the prior experiments have been replicated, using a different variant of the category exemplar production task. In this variant, only one example per category is presented during encoding. Consequently, item-specific encoding should predominate. Consistent with the relational view, removing test-aware subjects not only reduces, but eliminates the effects of LOP. To assess the power of this last comparison, a to-be-detected effect size was estimated from the test-aware subjects \( (d = 2.02) \). For the test-unaware subjects, the power to detect an effect half this size was .95. Combining the test-unaware subjects from the present experiment and the random conditions of Experiments 1A and 1B (which also used a within-subjects manipulation of LOP) produces a sample of \( n = 51 \). In this group, the average difference in priming scores between the semantic and the nonsemantic conditions is 2 percentage points, for a nonsignificant LOP effect \( (F = 1.04) \). The estimated size of the LOP effect for the corresponding test-aware subjects is \( d = 2.17 \). The power to detect an effect just one quarter this size \( (d = 0.54) \) in the test-unaware subjects is .85. Even if there is some small LOP effect for test-unaware subjects, the present data indicate that, under conditions of low category salience, the LOP effects reported in the literature are largely produced by test-aware subjects.

Finally, we note that the results of the present experiment do not by themselves unambiguously support the relational account. It could be argued that the LOP \( \times \) test awareness interaction simply indicates that removing explicit influences impacts most heavily the condition with the highest level of explicit recall—in this case, the semantic encoding condition. Such an account is consistent with the present results and with the results of the random condition in the prior experiments. However, this account predicts the same LOP \( \times \) test awareness interaction for the categorized list, a prediction at odds with the results of Experiments 1A, 1B, and 2. Thus, the complete set of findings is not consistent with this account, whereas it is consistent with the relational hypothesis.

**GENERAL DISCUSSION**

The present study was motivated by the suggestion that conceptual priming in the category exemplar production task is sensitive to relational encodings and relatively insensitive to item-specific encodings (Mulligan, 1996; Weldon & Coyote, 1996). The relational view produces several predictions. First, semantic encoding should have greater impact on priming when relational information is salient. Second, removing explicit contamination should differentially remove the influences of item-specific elaboration, reducing LOP effects for the random, but not for the categorized list presentation. Finally, a strong interpretation of this view suggests that conceptual priming in the category exemplar production task is not affected by all forms of semantic encoding, but only by those that emphasize relational information—in this case, categorical information.

Across three studies encompassing three experimental designs and two variants of category exemplar production, consistent results were found. First, the effects of LOP on conceptual priming were larger in the categorized than in the random list condition, indicating that semantic elaboration is more effective in a relational context. Second, the effects of LOP were consistent across experimental designs, indicating that, unlike other implicit tests, the effects of LOP on category exemplar production are not dependent on the details of the experimental design (i.e., mixed-list vs. blocked-list vs. between-subjects). Third, removing test-aware subjects reduced LOP effects in the random condition, in which item-specific encoding dominates, but not in the categorized condition, in which relational encoding predominates. Fourth, the analysis of the test-unaware subjects provides two important results: (1) Consistent with the full analysis, the LOP effect was larger in the categorized than in the random list condition, indicating that it is likely that the
LOP × list condition interaction is not an artifact of explicit contamination; and (2) among test-unaware subjects, there was no effect of LOP in the random list condition, indicating that there may be merit in the strong as well as in the weaker version of the relational encoding hypothesis.

The results are in accord with the relational view of the category exemplar production task. Recall that the motivation for the relational view was findings of dissociations between conceptual implicit and explicit tests (Mulligan, 1996; Weldon & Coyote, 1996). These and other dissociations (see, e.g., Cabeza, 1994; McDermott & Roediger, 1996) are problematic for the TAP view of implicit and explicit memory and indicate that distinctions within the class of conceptual processes need be drawn (e.g., McDermott & Roediger, 1996; Mulligan, 1996; Weldon & Coyote, 1996). The item-specific-relational distinction represents one way to expand the notion of conceptual processing (Mulligan, 1996; Weldon & Coyote, 1996). The present results support this approach.

It should be noted that the preceding critique applies most directly to the TAP view of implicit and explicit memory (focusing on the perceptual/conceptual processing distinction; e.g., Blaxton, 1989; Roediger, 1990; Roediger & McDermott, 1993). As was noted in the introduction (and in note 1), this theory represents only one application of the more general principle of transfer-appropriate processing, which focuses on encoding-retrieval compatibility (e.g., Kolars & Roediger, 1984; Morris et al., 1977; see, also, Hintzman, 1990). This general principle has broad applicability, and our primary results generally support it. Specifically, categorized lists encourage category processing at encoding that transfers well to a memory test using category cues (i.e., the category exemplar production task). Thus, the finding of larger LOP effects with categorized than with random lists is consistent with the general TAP principle. However, without further elaboration, this principle does not provide an account of the results from the test accuracy analyses or of the reported dissociations between category exemplar production and category-cued recall (Mulligan, 1996; Weldon & Coyote, 1996). For example, why should the perceptual interference or picture superiority effect dissociate two tests that both used the same nominal cues and differed only in instructions (see Mulligan, 1996, for a discussion)? Without additional assumptions, the TAP principle does not provide a fine-grained account of the relevant results in this domain. From the present perspective, it would be of great interest if such an account is developed that does not rely on the item-specific-relational distinction.

Returning to the empirical results, an alternative description of the LOP × list organization interaction also deserves discussion, because it implies that organization has little or no effect on conceptual priming (or cued recall) under nonsemantic encoding instructions. However, when items are encoded semantically (or under intentional learning instructions; Rappold & Hashtroudi, 1991), the effects of organization emerge. This is a sensible finding, because the relevant relational information for the category exemplar production task is semantic in nature (i.e., categorical information). Apparently, category-level information is not encoded in a manner that impacts later performance unless meaning is the focus of the encoding task. Merely blocking items by category at study will not necessarily lead to enhancements in conceptual priming. This indicates that the conceptual information (i.e., categorical information) supporting performance on this task is not encoded automatically, in the sense that the beneficial effects of organization require attending to the semantic features of the stimuli. This is consistent with findings that divided attention during encoding produces low levels of conceptual priming in this task and, in fact, can eliminate priming entirely (Mulligan, 1997; Mulligan & Hartman, 1996).

The effects of organization can also be related to the negative spacing effect (see, e.g., Glanzer, 1969), in which the recall of associates is negatively related to the spacing between them on the study list. The effect of organization can be conceived of as a negative spacing effect in that the space between associates (i.e., members of the same category) is greater for random than for categorized lists, leading to lower recall and conceptual priming in the former condition. Neither the effect of negative spacing nor that of organization is observed with perceptual identification, a perceptual priming task (Greene, 1990; Rappold & Hashtroudi, 1991).

A subsidiary goal of the present experiments was to examine the effects of LOP on category exemplar production, using various experimental designs. As were reviewed in the introduction, several implicit tests (word-fragment completion, perceptual identification, general knowledge questions) show smaller LOP effects in mixed-list designs than in blocked-list designs (Challis & Brodebeck, 1992; Thapar & Greene, 1994). In addition, the LOP effects are often nonsignificant in the mixed-list condition.

The present results limit the generality of this finding. The effects of LOP on the implicit test of category exemplar production are not moderated by the experimental design. The effects of LOP are robust, whether implemented in a mixed-list (Experiment 1A), a blocked-list (Experiments 1B and 3), or a between-subjects (Experiment 2) design. These results are most striking when contrasted with findings using the general knowledge test, which, like category exemplar production, is a conceptual implicit test. For the general knowledge test, the LOP effect was not only reduced but eliminated in a mixed-list design (Thapar & Greene, 1994). For category exemplar production, the LOP effect remains robust in this condition. Although the reasons for this difference are unclear, this pair of results does represent a dissociation between two conceptual implicit tests (see Cabeza, 1994, for another such dissociation). This dissociation adds to the set of dissociations among conceptual tests (reviewed in the introduction), which collectively indicate the need for re-
vision of the TAP view of implicit and explicit memory (see Mulligan, 1998, for a related discussion).

Next, we turn to the results of the test-unaware subjects. Use of the test awareness questionnaire was motivated by prior research indicating relatively high levels of test awareness in the category exemplar production task (Light & Albertson, 1989; Mulligan & Hartman, 1996). As is typical, questionnaire data were used to minimize the potential for explicit contamination by examining priming for subjects claiming no test awareness (see, e.g., Bowers & Schacter, 1990; Jelicic et al., 1996; Light & Albertson, 1989; McKone & Sliw, 1997; Mulligan & Hartman, 1996; Thapar & Greene, 1994). The re-analysis of the test-unaware subjects revealed the same pattern of results as that obtained with the full sample: there was a larger LOP effect in the categorized than in the random list condition, and the design of the experiment did not moderate LOP effects. Consequently, the primary results of the experiments are probably not jeopardized by explicit contamination. The more surprising finding was that, among the test-unaware subjects, the effect of LOP was found in the categorized, but not in the random list condition. This result was obtained across three experiments, indicating that the result is replicable.

Results from test-unaware subjects probably represent a more pure measure of implicit memory than do the results of the entire sample (see, e.g., Bowers & Schacter, 1990; Jelicic et al., 1996; Light & Albertson, 1989; Mulligan & Hartman, 1996; Thapar & Greene, 1994). We note, however, that test unawareness may be a conservative criterion for implicit memory. Richardson-Klavehn, Gardiner, and Java (1994; Richardson-Klavehn & Gardiner, 1996) have argued that conscious awareness may arise as a result of implicit memory retrieval, a phenomenon called involuntary conscious memory. To the extent that involuntary conscious memory can occur without involving intentional retrieval strategies, test unawareness may be a conservative criterion for implicit memory. It would be of great interest to apply the methods of Richardson-Klavehn, Gardiner, and their colleagues (e.g., Richardson-Klavehn & Gardiner, 1996; Richardson-Klavehn et al., 1994), to see whether the same results obtain.

The results from the test awareness reanalysis are suggestive in terms of the relational hypothesis. Recall that the most common semantic encoding task, pleasantness rating, is also considered the prototypical item-specific encoding task (e.g., Hunt & Einsten, 1981; Hunt & McDaniel, 1993). The strong version of the relational hypothesis suggests that semantic encoding should enhance priming in the category exemplar production task only if the semantic processing is relational in nature, and not if it focuses on item-specific information, as in the pleasantness-rating task. As was noted in the introduction, the strong version of the relational hypothesis appears to be inconsistent with earlier reports of LOP effects in category exemplar production (Hamann, 1990; Mecklenbrauker et al., 1996; Srinivas & Roediger, 1990; Weldon & Coyote, 1996). These studies use semantic tasks similar to those in the present study, yet report robust effects of LOP. However, because the results of the test-unaware subjects conform to the strong version of the relational hypothesis, it may be worth considering these prior findings in light of this hypothesis. In particular, this hypothesis suggests that observed LOP effects in category exemplar production are driven by relational (or categorical) semantic processing at encoding, or explicit contamination of the priming task. To anticipate, a consideration of the relevant research does not rule out the strong relational hypothesis and suggests additional research.

For present purposes, we ask two questions about studies that have found LOP effects in the category exemplar production task. First, was the categorical nature of the study list salient? Second, was an attempt made to assess the likelihood of explicit contamination on the implicit task? With regard to the first question, in two of the studies (Hamann, 1990; Weldon & Coyote, 1996), it is likely that the categorical structure of the study list was salient. In Weldon and Coyote (1996, Experiment 4), each study block consisted of four words from each of four categories. Although the words were randomly ordered, it seems likely that the categorical structure was salient within such a short study block. In Hamann (1990, Experiment 1), again, the order of study items was randomized, but the number of words per category (five) and the total length of the study blocks (15) also made it likely that some amount of category clustering occurred on the study lists, rendering the categorical nature of the study list salient. The same is true of the Hamann (1989, reported in Hamann, 1990) study, which reports a LOP effect, using amnestic subjects. Consequently, it is possible that the observed LOP effects in all of these studies were at least partially due to relational–semantic processing.

This account is less applicable to the other studies, which either structured study lists so that the category organization was not salient (e.g., relatively long study lists in which no more than one example from a category appeared in sequence; Hamann, 1990, Experiment 2; Mecklenbrauker et al., 1996) or used a version of the category exemplar task in which only a single example from each category was presented at study (Srinivas & Roediger, 1990). These studies found effects of LOP under conditions similar to our random list condition. However, in these studies, no attempt was made to assess the level of explicit contamination. Because these studies used similar methods and subject populations (i.e., young adult university students), it is reasonable to assume that relatively high numbers of subjects were test aware. The results of these experiments are consistent with this critique. When the categorical structure of the study list is not salient, test awareness contributes greatly to the ob-
served size of the LOP effect. This result holds for both variants of the category exemplar production task that have been used in the literature.

Of course, there are candidate methods other than awareness questionnaires for assessing explicit contamination on implicit tests and for rendering more pure measures of implicit memory, including the retrieval-intentionality criterion and the process dissociation procedure (see Roediger & McDermott, 1993, for a thorough discussion). One of the prior studies of LOP and category exemplar production (Weldon & Coyote, 1996) satisfied the retrieval-intentionality criterion, indicating that, for this study, it is reasonable to conclude that explicit contamination did not greatly influence the implicit test. But, as has been noted, the observed LOP effect in this study may be due to relational-semantic processing. Mecklenbrauker et al. (1996) used the process dissociation procedure in their study (recall that, in the Mecklenbrauker et al. study, the categorical structure of the study list was salient). The process dissociation analysis indicated that LOP affected conscious (explicit) memory but not unconscious (implicit) memory. Although there are many unresolved questions and criticisms of the process dissociation methodology (e.g., Curran & Hintzman, 1993; Komatsu, Graf, & Uttl, 1995; Mulligan & Hirshman, 1997; Richardson-Klavehn et al., 1994), this result is consistent with the strong version of the relational hypothesis: Semantic elaboration only affects priming in this implicit test when it occurs in a relational context.

REFERENCES


