Visual Specificity Effects on Word Stem Completion: Beyond Transfer Appropriate Processing?

TIM CURRAN, DANIEL L. SCHACTER, and GAYLE BESSENOFF

Harvard University

Abstract An important, but poorly understood, aspect of memory retrieval concerns the conditions under which priming is influenced by perceptual changes in the form of target items. According to transfer appropriate processing perspectives, perceptual specificity effects on priming require a study task that focuses attention on the perceptual, rather than semantic, features of the items. Other research suggests that perceptual specificity effects are enhanced by conditions yielding high levels of explicit memory. The present experiments manipulated encoding tasks and other variables known to influence explicit memory (repetition and retention interval) in order to gain insight into the determinants of perceptual specificity effects on visual word-stem completion. In Experiment 1 we found that perceptual specificity (letter case) effects on stem completion priming depend on perceptual encoding when subjects' awareness of the study-test relationship is limited. In Experiments 2-4 we found that perceptual specificity effects can be obtained after semantic encoding - especially when the study-test retention interval is short. Perceptual specificity effects after short retention intervals were independent of encoding task, and may reflect a form of involuntary explicit memory.

Numerous studies have established that priming effects on such implicit memory tests as stem completion, fragment completion, word identification, and lexical decision are largely modality specific, are rarely affected by depth of encoding manipulations (but see Brown & Mitchell, 1994; Challis & Brodbeck, 1992), and are typically preserved in patients with organic amnesia. By contrast, performance on standard explicit memory tests is largely modality nonspecific, is greatly affected by depth of encoding, and is profoundly impaired in amnesic patients (for reviews, see Richardson-Klavehn & Bjork, 1988; Roediger & McDermott, 1993; Schacter, 1987; Schacter, Chiu, & Ochsner, 1993; Shimamura, 1986).

An important but as yet poorly understood aspect of memory retrieval concerns the extent to which, and conditions under which, it is influenced by changing the exact perceptual form of target stimuli. Numerous studies have found that priming is reduced by changing the stimulus modality (e.g., visual vs. auditory) from study to test (for a review see, Roediger & McDermott, 1993). Other studies testing within-modality manipulation of specific perceptual features have yielded a wide range of outcomes. Within the domain of visual word priming, some experiments have yielded evidence that priming effects are larger when the typefont or case (i.e., upper or lower) of target items is the same at study and test than when it is changed (e.g., Blaxton, 1989; Jacoby & Hayman, 1987; Roediger & Blaxton, 1987). Other experiments, however, have not obtained such effects (cf. Carr, Brown, & Charalambous, 1989; Rajaram & Roediger, 1993; Scarborough, Cortese, & Scarborough, 1977). For example, Rajaram and Roediger (1993) failed to observe significant effects of changing the typefont of target items between study and test on stem completion, fragment completion, anagram solution, and word identification tasks, even though performance on these tests was significantly higher after visual than auditory presentation.

Additional studies have shown that, within the same experiment, form-specific priming effects may be observed under some conditions but not others. For example, form-specific priming occurs following study of unusual or highly distinctive typefonts or handwriting, but not after study of typical typefonts (Brown & Carr, 1993; Graf & Ryan, 1990). Marsolek, Kosslyn, and Squire (1992) found that form-specific priming occurs when test items are presented in the left visual field but not in the right visual field. Of particular relevance to the present experiments, Graf and Ryan (1990) found that word identification priming was reduced by a study-test change in typefont only when subjects rated the readability of words during study, but not when they rated how much they liked each word. Theorizing from a transfer-appropriate processing perspective (e.g., Morris, Bransford, & Franks, 1977), Graf and Ryan suggested that distinctive perceptual information about targets was only recorded when the encoding task focused processing on sensory or perceptual features of the words. On the other hand, Jacoby, Levy, and Steinbach (1992) advanced the seemingly contradictory suggestion that visual specificity effects are more likely when the encoding task places perceptual analysis in the background of a semantic task rather than overtly drawing attention to perceptual features. Jacoby et al. measured memory for 5- to 9-word questions indirectly through reading times.
Font specificity effects (typed vs. script) were observed when subjects were instructed to read questions silently before answering them (an overtly semantic task that puts perceptual analysis in the background) but were not observed when subjects simply read the questions aloud without answering them (an overtly perceptual task).

Similar issues have arisen in studies of auditory word priming. Jackson and Morton (1984) found that priming effects on a task that required identification of words masked in white noise were larger after auditory than visual presentation, but were unaffected by study-to-test changes in speaker's voice. Schacter and Church (1992) replicated these results, but also reported that voice-specific priming could be observed on an auditory stem completion test without white noise. More recently, Church and Schacter (1994) extended observations of voice-specific priming to an auditory identification test with words degraded by a low-pass filter. They also reported that priming on auditory stem completion and low-pass filter identification tests is affected by study-test changes in linguistic intonation, emotional intonation, or fundamental frequency of a single speaker's voice. Voice-effects have been consistently found after both perceptual and semantic encoding tasks, so encoding task characteristics may only be important in the visual domain.

Recent studies with amnesic subjects raise the possibility that perceptual specificity effects depend on different underlying mechanisms than priming effects that are not perceptually specific. Amnesic patients exhibit severe deficits in explicit memory for recent experiences that are produced by damage to limbic and diencephalic brain structures (Parkin & Leng, 1993; Squire, 1992; Weiskrantz, 1985). Nevertheless, amnesics have consistently shown intact visual word priming on completion and identification tests (for reviews see Bowers & Schacter, 1993; Moscovitch, Vriezen, & Goshen-Gottstein, 1993; Schacter, et al., 1993; Shimamura, 1986), including normal sensitivity to modality change (Carlesimo, Fadda, Sabbadini, & Caltagirone, 1994; Graf, Shimamura, & Squire, 1985).

Amnesics have also exhibited normal auditory word priming on the identification-in-noise test, which does not yield evidence of voice-specific priming (Schacter, Church, & Treadwell, 1994). By contrast, two recent studies indicate impaired form-specific priming in amnesic patients. Kinoshita and Wayland (1993) report that control subjects exhibited form-specific priming on a fragment completion test after studying handwritten words, whereas amnesic patients failed to exhibit form-specific priming. Schacter, Church, and Bolton (1995) found that amnesic patients did not exhibit voice-specific priming on a filter identification test under conditions in which control subjects did exhibit voice-specific effects (for an extended discussion of these findings see Curran & Schacter, in press).

These observations suggest that perceptual specificity effects on priming depend on mechanisms that normally support explicit memory. A number of researchers have argued that spared priming in amnesic patients reflects the operation of a memory system that is distinct from the episodic memory system that is crucial for explicit recollection (cf. Keane, Gabrieli, Fennema, Growdon, & Corkin, 1991; Schacter, 1990; Squire, 1994; Tulving & Schacter, 1990). For example, Schacter has argued that priming depends heavily on a perceptual representation system (PRS) - a collection of cortically-based perceptual systems that process and represent information about the form and structure, but not the meaning and associative properties, of words and objects (Schacter, 1990; Schacter, 1992; Schacter, 1994; Tulving & Schacter, 1990). The observed impairment of form-specific priming in amnesic patients suggests that form-specific priming cannot be based on PRS alone. For example, we have suggested that form-specific priming of the sort observed in Kinoshita and Wayland's (1993) and Schacter et al.'s (1994) experiments may require binding of relatively abstract perceptual word forms with specific features of typography or speaker's voice, and that such binding depends on some of the same limbic/diencephalic structures that ordinarily support explicit memory (Curran & Schacter, in press; Schacter, 1994; Schacter et al., 1995).

Other evidence for a possible link between perceptual specificity and explicit memory comes from studies of priming in normal subjects. In divided visual-field studies, Marsolek, Squire, Kosslyn, and Luleski (1994) noted a number of procedural characteristics that appeared to influence whether or not a case effect is observed in stem completion and stem-cued recall tests. Marsolek et al. (1992) found case effects after implicit stem completion (left-visual field only), but not after explicit stem-cued recall. Marsolek et al. (1994) found left-visual field case effects after both stem completion and stem-cued recall with the following procedural changes (compared to Marsolek et al., 1992): 2 presentations (rather than 1), shorter study-list (45 vs. 15), shorter test-list (80 vs. 20 lists, and shorter study-test retention intervals (6 min. vs. 2 min.). Overall, then, Marsolek et al. (1994) created conditions that yielded higher levels of explicit memory than were observed in Marsolek et al. (1992), and case effects were observed in cued-recall only under the conditions that fostered high levels of explicit memory. Another hint that the creation of a distinctive episodic trace may be important for the occurrence of form-specific priming comes from the fact that typography-specific priming more often occurs when unusual typefaces or hand-written scripts are studied (cf., Brown & Carr, 1993; Graf & Ryan, 1990; Kinoshita & Wayland, 1993).

The present experiments were designed to examine issues arising from the inconsistencies observed in the previously discussed research. Graf and Ryan's (1990) transfer-appropriate processing view suggests that perceptual specificity effects are most likely to occur after perceptual encoding tasks that are known to impede explicit memory (e.g., Craik & Tulving, 1975). This view generally is consistent with predominant theories that posit a link between perceptual or data-driven processing and implicit (e.g., Roediger, 1990; Roediger & McDermott, 1993; Schacter, 1994) or automatic (Jacoby, Toth, & Yonelinas, 1993; Toth, Reingold, & Jacoby, 1994).
memory. A complement of this view holds that explicit or consciously controlled memory predominantly benefits from conceptual processing. However, evidence reviewed above suggests a link between some aspect of explicit memory and perceptual specificity that does not readily fit with these frameworks. The present experiments manipulated encoding tasks and other variables known to influence explicit memory (repetition and retention interval) in order to gain insight into the determinants of perceptual specificity effects in visual word priming.

**Experiment 1**

Experiment 1 was designed to replicate and extend Graf and Ryan's (1990) finding that visual specificity effects occur after perceptual encoding tasks but not after semantic encoding tasks. Graf and Ryan manipulated the font in which words were studied and tested - both fonts were novel and distinctive. In their experiment, words were encoded with liking ratings or readability ratings, and subjects performed a word identification test. In our Experiment 1, subjects studied upper and lower case words in a standard font, and performed either a semantic (liking ratings) or perceptual (t-junction counting) encoding task. Priming was assessed with a visual stem completion test with all stems presented in upper case.

**METHOD**

**Subjects**

Subjects were 32 Harvard undergraduates paid $10.00 for participation. Subjects were tested individually in a 1-hour session.

**Stimuli and Apparatus**

Experimental stimuli were 96 common English words that each began with a different three-letter combination. For counterbalancing purposes, these words were divided into eight, 12-word subsets which were roughly equated for word length ($M = 2.23; SD = .80; range = 5 to 9$), word frequency ($M = 12.22; SD = 8.68; range = 2 to 42$, Kucera & Francis, 1967), and rank frequency among all words starting with the same three letters ($M = 6.07; SD = 3.87; range = 1 to 20$). The number of possible completions for each 3-letter stem with Kucera and Francis (1967) frequency greater than zero was also balanced across subsets ($M = 15.61; SD = 11.22; range = 3 to 59$). Twenty-four words with similar characteristics, but different 3-letter stems, were used as primacy and recency buffers in the study lists. Another unique set of 24 stems served as filler items for the stem completion task.

Stimulus presentation and response collection were controlled by a Macintosh ttx computer. Words were presented in 24-point Geneva font (black on white background).

**Design**

Study condition (nonstudied, t-junction counting, and liking ratings) and study case (upper case vs. lower case) were manipulated within-subjects. Each subject studied two word-lists followed by a single test-list. The study tasks (t-junctions and liking) were blocked, and order was counterbalanced across subjects such that half completed the t-junction task first and vice versa. Study case was manipulated within each study list. Items were completely rotated through the study conditions so that each item was used twice in each condition across subjects.

Study lists for the t-junction and liking tasks were each 36-items long. Twenty-four experimental words were surrounded by six-word primacy and recency buffers. Upper and lower case words were presented in random order, determined separately for each subject with the constraint that no more than 3 consecutive words had the same case. The test list included 120 upper-case, 3-letter stems. The first 24 items were non-studied filler items included for practice and to disguise the fact that many stems could be completed with words from the previous tasks. The remaining 96 stems could be completed with nonstudied target words (48 words), words from the t-junction task (24 words), or words from the liking task (24 words). Each subject received the same test list in which order was randomly determined with no more than 3 consecutive stems from the same presentation condition.

**Procedure**

Subjects were given a number of filler tasks to help obscure the relationship between the study and test tasks. First, subjects were given 5 minutes to write down the names of U.S. States and their corresponding capital cities.

Next, subjects completed the two study tasks in the order determined by the counterbalancing scheme. For both tasks, words were presented for 3 seconds with a 0.5 s inter-stimulus-interval. In the t-junction task, subjects counted the number of instances in which two lines intersect in a t-shaped formation. In the liking task, subjects rated each word on a 5-point scale according to how much they liked its meaning (1 = strongly dislike; 5 = strongly like). Subjects wrote their responses on a numbered form.

After the t-junction and liking tasks, subjects completed two 5-minute filler tasks. First, subjects wrote down names of U.S. Presidents. Second, subjects completed a number-search task in which they were asked to search for specific numbers, each comprised of 5 digits, within a 15 x 15 matrix of digits. The experimenter reminded the subjects that all filler tasks should be performed carefully. The State Capital and Presidents tasks were presented as tests of very long-term memory. The number-search task was described as measuring "baseline response speed" because all of the other tasks were speeded in some manner.

Subjects completed the stem completion task after the number-search task. Subjects were asked to complete each stem with the first word that came to mind. To encourage compliance with this instruction, subjects were put under some time pressure. They were told to hit the space bar on the keyboard as soon as the first completion came to mind. After hitting the space bar the subject could enter the word. They were encouraged to only enter the word that originally came to mind. Reaction times were...
Visual Specificity and Stem Completion

**TABLE 1**
Proportion of Targets Produced on the Stem Completion Task as a Function of Study Task and Letter Case, Experiment 1

<table>
<thead>
<tr>
<th>Study Condition</th>
<th>Study-Test Case</th>
<th>Different</th>
<th>Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liking</td>
<td>M</td>
<td>0.20</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(0.12)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>T-junction</td>
<td>M</td>
<td>0.16</td>
<td>0.22 *</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(0.13)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Nonstudied</td>
<td>M</td>
<td></td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(0.05)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Asterisks denote the significance of a 1-tailed t-test on the case effect. *p < .05

We did not include stems of nonstudied words because the presumed episodic effects should be maximized by using only stems of previously studied words. Because the target materials were the same as in Experiment 1, it is reasonable to assume that the Experiment 1 baseline completion rate (.09) provides an accurate estimate of baseline level in the present experiment.
manipulated test instructions: Subjects were either instructed to respond with the first word that came to mind (unintentional retrieval) or they were instructed to try to recall a studied word (intentional retrieval).

**Experiment 2**

**Method**

**Subjects**

Subjects were 32 Harvard undergraduates paid $10.00 and tested individually in a 1-hour session.

**Design and Materials**

Each subject participated in 4 study-test blocks. Retrieval instructions (unintentional vs. intentional) and study task (t-junctions vs. liking ratings) were manipulated within subjects and between blocks. For half the subjects unintentional instructions were given on the first two blocks and intentional on the last two blocks, and vice versa for the other subjects. Study task was counterbalanced independently of retrieval instructions in an ABBA-BAAB design. Letter case was varied within each study list with test stem always in upper case. Critical words were those used in Experiment 1, and items were rotated through each condition so that each word appeared once in each condition.

Study lists included 30 items, and each list was repeated twice in different random orders. The 24 critical items (half upper case and half lower case) were surrounded by 3-item primacy and recency buffers. Test lists contained 3-letter stems that could be completed with words from the preceding study list. Unlike Experiment 1, test lists did not include stems in a nonstudied condition. Both study- and test-list order were determined randomly for each subject with the constraint that no more than 3 consecutive items were from the same condition.

**Procedure**

First, subjects were given a 10-item practice list with nonstudied visual stems under unintentional retrieval instructions. Instructions were identical to those used in the test phase of Experiment 1. Subjects were asked to press the space bar as soon as the first correction completion came to mind, and then enter the completion into the computer. This practice was intended to familiarize subjects with the test procedure so that instructions given preceding unintentional instructions for half the subjects.

Study and test-list order were determined randomly for each subject with the constraint that no more than 3 consecutive items were from the same condition.

Next, subjects completed 4 study-test blocks: one in each retrieval instruction by study task combination. The interval between study-list repetitions was subject-paced. Each study-test interval included only the test instructions and typically lasted less than 30 seconds. Instructions and stimulus duration for the t-junction and liking tasks were identical to those in Experiment 1. In the unintentional test blocks, as in the practice test, subjects were told to respond quickly with the first completion that came to mind. Immediately following the last unintentional test list, subjects completed the questionnaire described in Experiment 1.

In the intentional test blocks, subjects were instructed to “Take as much time as you need to search through your memory for the previously studied words in order to recall a word that fits the test stem. If you cannot recall the studied word, take a guess by writing down the first word that comes to mind that completes the 3-letter stem.”

Regardless of retrieval instructions, subjects were informed that stems could be completed with studied words. A number of aspects of the method—two presentations of studied words, no filler tasks, negligible retention interval, no nonstudied test stems, and intentional instructions preceding unintentional instructions for half the subjects—made it unrealistic to expect subjects to remain unaware of the study-test relationship. Therefore, we emphasized compliance with the unintentional retrieval instructions while keeping awareness of the study-test relationship uniform across subjects. Informing subjects of the study-test relationship may make them more likely to respond with the first word that came to mind than would allowing subjects to discover the relationship themselves (Bowers & Schacter, 1990). At the end of the experiment, subjects completed a questionnaire that included all but the first question of the Experiment 1 questionnaire.

**Results and Discussion**

The target completion rates are presented in Table 2. The intentional and unintentional conditions were analyzed separately. These mixed-model ANOVAs included encoding task (t-junctions vs. liking) and case (different vs. same) as repeated measures and test order (intentional first vs. unintentional first) as a between-subject factor.

The intentional retrieval instructions yielded a straightforward pattern of results—completion rates were significantly higher after the liking task than after the t-junction task, \( t(1,30) = 77.10, M_5 = .04 \). No effects involving study case or test order approached significance.

In the unintentional retrieval condition, completion rates were significantly higher after the liking task than after the t-junction task, \( t(1,30) = 16.23, M_5 = .027 \); and significantly higher when the study-test case was the same

<table>
<thead>
<tr>
<th>Test Instruction</th>
<th>Study Case</th>
<th>Different</th>
<th>Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unintentional</td>
<td>Liking</td>
<td>0.57</td>
<td>0.67 *</td>
</tr>
<tr>
<td></td>
<td>T-junction</td>
<td>0.46</td>
<td>0.55 *</td>
</tr>
<tr>
<td>Intentional</td>
<td>Liking</td>
<td>0.82</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>T-junction</td>
<td>0.52</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Note: Asterisks denote the significance of a 1-tailed t-test on the case effect. * \( p < .05 \)
than when it was different, $F(1,30) = 10.48$, $M_{SE} = 0.027$. There was also a significant encoding task by test order interaction, $F(1,30) = 6.52$, $M_{SE} = 0.027$.

To understand this encoding by order interaction, separate encoding by case ANOVAs were performed on the subjects who had been given each test order. The encoding task had a significant effect only when the unintentional lists came second, $F(1,15) = 15.92$, $M_{SE} = 0.037$, but not when they came first, $F(1,15) = 1.70$, $M_{SE} = 0.017$. These ANOVAs also revealed that the case effect was highly significant when the unintentional condition was tested first, $F(1,15) = 7.15$, $M_{SE} = 0.029$, but not when tested second, $F(1,15) = 3.51$, $M_{SE} = 0.024$. Thus, the case effect appeared to be stronger in conditions in which there was no effect of encoding task.

We believe this order by encoding interaction is attributable to subjects being less likely to comply with the unintentional instructions (to respond with the first word that comes to mind) after they had already completed the intentional phase. Questionnaire responses confirm that 50% of subjects who had the unintentional instructions second indicated that sometimes they intentionally tried to remember words from the previous lists. Only 31% admitted to intentional retrieval strategies when the unintentional condition was first. Subjects generally better complied with unintentional retrieval instructions when that test was first, and this is a likely explanation for why only subjects who performed unintentional retrieval second benefited from the liking task. Similarly, Richardson-Klavehn, Lee, Joubran, and Bjork (1994b) showed that the presence of such LOP effects is largely determined by whether or not subjects use intentional retrieval strategies.

To compare with the results of Experiment 1, planned comparisons examined the case effect within each encoding condition under unintentional retrieval instructions. Completion rates were significantly higher when the study-test case remained constant after both the t-junction task, $t(31) = 227$, $SE = 0.04$, and the liking task, $t(31) = 2.71$, $SE = 0.04$. Thus, unlike Experiment 1 in which the case effect was only significant after the t-junction task, the case effect was significant after both encoding tasks in Experiment 2.

This experiment demonstrates that perceptual specificity effects can be obtained after a semantic encoding task. Differences between Experiment 1 and Experiment 2 were intended to increase explicit memory, so one obvious potential explanation is that case effects are attributable to intentional retrieval. This idea is undermined by the absence of significant case effects under intentional retrieval instructions. Both retrieval conditions lead to explicit memory in the sense that subjects were aware that their completions were study-list members, but the presence or absence of the case effects depended upon whether or not subjects intentionally attempted to recall list items. Not only does this dissociation meet the retrieval intentionality criterion (Schacter, Bowers, & Booker, 1989), but it also satisfies Merikle and Reingold’s (1991) more stringent test of demonstrating a greater effect on the unintentional test than the intentional test. This dissociation is particularly convincing in the t-junction conditions because the intentional and unintentional completion rates were very similar; hence the comparison is not compromised by baseline differences or ceiling effects.

Perhaps the unintentional retrieval conditions of Experiment 2 encouraged a form of involuntary explicit memory (Richardson-Klavehn, Gardiner, & Java, 1994a; Richardson-Klavehn, et al., 1994b; Schacter et al., 1989), whereas performance in Experiment 1 was primarily attributable to implicit memory. If so, the results of Experiments 1 and 2 suggest that implicit case effects may depend on perceptual encoding (Experiment 1, Graf & Ryan, 1990), but encoding-independent case effects can arise under conditions that favor involuntary explicit memory (Experiment 2). The next two experiments attempted to isolate the parameters that differed between Experiments 1 and 2 that were critical for the appearance of the encoding-independent case effect in Experiment 2. Specifically, we examined the possible contributions of retention interval and number of study-list presentations to the observation of encoding-independent case effects.

Experiment 3

METHOD

Subjects

Subjects were 128 Harvard undergraduates paid $10.00 and tested individually in a 1-hour session.

Design and Materials

Retention interval (short [< 30 sec] vs. long [10 minutes]) and number of study list presentations (1 vs. 2) were manipulated between subjects (32 per group). Study task (t-junction vs. liking) was blocked within subjects with order counterbalanced across subjects. Study case (upper case vs. lower case) was varied within each study list. Composition of the study and test lists were identical to Experiment 2, but study lists were only repeated for half the subjects. The critical words were those used in Experiments 1 and 2, and were counterbalanced across subjects so that each appeared twice in each condition.

Procedure

Subjects began with the practice stem completion task described in Experiment 2. Next, subjects completed 2 study-test blocks (t-junction and liking) with retrieval instructions that were identical to the unintentional test instructions of Experiment 2. Subjects again were informed that stems could be completed with studied words, but were encouraged to quickly respond with the first completion that came to mind.

Subjects in the 10 minute retention interval groups were given a filler task after each study list. The filler task was a serial reaction time task (Curran & Keele, 1993; Nissen

---

1 We thank Kathleen McDermott for bringing this point to our attention.
TABLE 3
Proportion of Targets Produced on the Stem Completion Task as a Function of Retention Interval, Study Task, Number of Presentations, and Letter Case, Experiment 3

<table>
<thead>
<tr>
<th>Retention Interval</th>
<th>Study Task</th>
<th>Presentations</th>
<th>Different</th>
<th>Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td>Liking</td>
<td>1 M</td>
<td>0.19</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SD)</td>
<td>0.13</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>T-junction</td>
<td>1 M</td>
<td>0.14</td>
<td>0.20 *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SD)</td>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>Short</td>
<td>Liking</td>
<td>1 M</td>
<td>0.54</td>
<td>0.62 *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SD)</td>
<td>0.19</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>T-junction</td>
<td>1 M</td>
<td>0.42</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SD)</td>
<td>0.17</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Note: Asterisks denote the significance of a 1-tailed t-test on the case effect. *p < .05, **p < .01

& Bullemer, 1987) in which subjects were asked to press 4 keys that corresponded to 4 spatially-defined targets on the computer screen. Subjects were asked to simultaneously monitor a series of high- and low-pitched tones and keep a running count of the high-pitched ones. To help equate the total time that each subject was in the session as well as to control for any general effects that the filler task may have on stem completion performance, subjects in the no retention interval group were given the filler task before each study list rather than within each study-test interval. These subjects, like those in Experiment 2, were merely reminded of the test instructions after the study list. At the end of the experiment subjects completed the same questionnaire as in Experiment 2.

RESULTS AND DISCUSSION

The target completion rates are presented in Table 3. First, these completion rates were entered into a retention interval by repetition by study task by case ANOVA. The main effects of retention interval, $f(1,124) = 193.41$, $MS_e = .059$; study task, $f(1,124) = 21.86$, $MS_e = .023$, and case, $f(1,124) = 15.38$, $MS_e = .019$, were all significant. A significant study task by retention interval by repetition interaction, $f(1,124) = 9.22$, $MS_e = .023$, reflected the fact that in the short retention interval condition the study task effect was significant only after one presentation, but in the long retention interval condition the study task effect was significant only after two presentations. We have no explanation for this odd pattern.

As in the previous experiments, we are primarily interested in the case effects in the various conditions, so 1-tailed t-tests compared the same and different case completion rates in each condition (see Table 3 for summary). In the long retention interval and 1 presentation condition, the case effect was significant after the t-junction task, $t(31) = 1.78$, $SE = .03$, but not after the liking task, $t(31) = .91$, $SE = .03$. This same pattern held for the long retention interval and 2 presentation condition: t-junctions, $t(31) = 2.62$, $SE = .02$; liking, $t(31) = 1.40$, $SE = .03$. In both conditions with a short retention interval, the case effect was significant after the liking task: 1 presentation, $t(31) = 2.33$, $SE = .04$; 2 presentations, $t(31) = 1.79$, $SE = .03$; but not significant after the t-junction task: 1 presentation, $t(31) = .89$, $SE = .04$; 2 presentations, $t(31) = .65$, $SE = .04$.

Thus, although the case by encoding task by retention interval interaction did not reach significance in the overall ANOVA, $f(1,124) = 1.81$, $MS_e = .016$, these t-tests suggest that with a long retention interval case effects are observed only after the t-junction task, but with a short retention interval case effects are observed only after the liking task. This pattern does not appear to be attributable to differences in adherence to the unintentional retrieval instructions because a similar percentage of subjects admitted to sometimes intentionally try to recall completions in the short (27%) and long (25%) retention interval conditions. These percentages include a large number of subjects who said that they intentionally tried to recall words only one or two times, so they represent a somewhat liberal estimate of the use of intentional retrieval.

Unlike Experiments 1 and 2, significant case effects were not obtained in the short/t-junction condition. The absence of a nonstudied baseline condition complicates interpretation of this null case effect. Such null effects would be uninformative in conditions without any memory influence on stem completion (i.e., conditions with performance that is not above baseline). However, there is good reason to believe that memory influenced stem completion in the t-junction condition of Experiment 3. The same items were used in Experiments 1 and 3, so the baseline completion rate of Experiment 1 (09) provides a rough estimate. It is apparent that all Experiment 3 completion rates are well above this baseline estimate, so we do not believe that our null case effects reflect null memory effects.

The results of Experiment 3 suggested that retention interval was most likely the critical difference between Experiments 1 and 2. Completion rate increased with presentation frequency, but this effect did not interact with the case effect. Furthermore, the results of the long retention interval conditions of Experiment 3 clearly replicated those of Experiment 1 (i.e., case effects after the t-junction encoding task, but not after the liking encoding task), even though nonstudied items were tested in Experiment 1 but not in Experiment 3. Although t-tests suggest a case effect by encoding task by retention interval interaction in Experiment 3, the interaction failed to reach significance in the mixed-model ANOVA. In Experiment 4 we sought a more powerful test of this interaction by manipulating all variables within subjects, rather than between subjects.
Experiment 4

METHOD

Subjects

Subjects were 64 Harvard University undergraduates who participated in a one-hour session for $10.00.

Materials and Design

Retention interval (short vs. long), study task (liking vs. t-junctions), and case (same vs. different) were manipulated within subjects. Each subject completed 4 study-test blocks - 2 short retention interval (S) and 2 long retention interval (L) - in one of four orders: SLSL, SLLS, LSLS, LSSL. In contrast with the previous experiments, t-junctions and liking ratings were randomly intermixed within each study list. In each list, 12 words were assigned to the t-junction task and 12 words were assigned to the liking task. Each study list also included a 2-word primacy and 2-word recency buffer (1 liking and 1 t-junction). Half the words were studied in upper case (same as test) and half tested in lower case.

The stimuli were mostly those used in the previous experiments. Some words were replaced because of extremely high baseline completion rates in Experiment 1 and other experiments in our laboratory. The items were divided in 16, six-item sublists. These sublists were roughly matched for length (M = 6.21, SD = 0.82, range = 5 to 9), word frequency (M = 17.83, SD = 23.72, range = 1 to 130, Kucera & Francis, 1967), number of possible completions of the 3-letter stem with Kucera and Francis frequency of greater than zero (M = 18.52, SD = 13.75, range = 4 to 70), and baseline completion rate (M = .05, SD = .05, range = 0 to .15). These sublists were rotated across subjects so that each item appeared equally often in each condition.

Procedure

The stem completion procedure was changed from that used in Experiments 1 through 3 to further discourage the use of intentional retrieval. To this end, subjects were explicitly told that they should not try to intentionally remember words because we were primarily interested in speed of responding, and warned that trying to intentionally remember would only slow them down. Also, the “Try to Respond Faster !!!” feedback was presented after responses greater than 800 ms, rather than after 2 s as in the previous experiments.

Each session began with a 24-item practice stem completion task. Next, subjects were given study task instructions. Prior to the presentation of each word, the study task (“Liking” or “T-Junctions”) was identified on the bottom of the computer screen for 1 second. The task identity remained on the screen throughout the four-second presentation of the word. Subjects were instructed to enter a number into the computer that corresponded to the number of t-junctions or to the liking rating. The inter-trial interval was 1 second. Study lists were separately randomized for each subject with the constraint that no more than three consecutive items were from the same study task.

RESULTS

An ANOVA on the target completion rates (Table 4) showed significant main effects of retention interval, F(1, 60) = 34.79, MS = .07; study task, F(1,60) = 10.42, MS = .03; and case, F(1,60) = 12.43, MS = .03. No interactions approached significance, but planned one-tailed t tests examined the case effect in each condition separately. In the long retention interval condition case effects were not significant: liking task, t(63) = 1.02, SE = .02; t-junction task, t(63) = 1.37, SE = .02. In the short retention interval condition the case effect was significant after both the liking task, t(63) = 2.57, SE = .02, p < .01, and the t-junction task, t(63) = 1.86, SE = .03.

Seventeen percent of the subjects indicated that they had occasionally intentionally tried to recall studied words.

Experiment 4 failed to detect the effect by encoding task by retention interval interaction that was suggested in Experiment 3. Unlike Experiments 1 and 3, the case effect was not significant in the t-junction condition with a long retention interval. The short retention interval condition gave significant case effects after both encoding tasks - a result that is similar to Experiment 2 but different from Experiment 3. A possible explanation for these inconsistent results may be the mixture of encoding tasks within each list. The mixture of encoding tasks may make encoding of perceptual and semantic attributes more automatic in both conditions.

<table>
<thead>
<tr>
<th>Retention Interval</th>
<th>Study Task</th>
<th>Study-Test Case</th>
<th>Same</th>
<th>Different</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>Liking</td>
<td>M</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(0.13)</td>
<td>(0.12)</td>
<td></td>
</tr>
<tr>
<td>T-junction</td>
<td>M</td>
<td>0.15</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(0.13)</td>
<td>(0.12)</td>
<td></td>
</tr>
<tr>
<td>Long</td>
<td>Liking</td>
<td>M</td>
<td>0.47</td>
<td>0.53**</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(0.17)</td>
<td>(0.21)</td>
<td></td>
</tr>
<tr>
<td>T-junction</td>
<td>M</td>
<td>0.44</td>
<td>0.49*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(0.17)</td>
<td>(0.18)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Asterisks denote the significance of a 1-tailed t-test on the case effect. *p < .05, **p < .01.
As in Experiment 3, we must be concerned with the possibility that null case effects are the uninteresting consequence of null memory effects. This possibility is especially tenable in the long retention interval conditions of Experiment 4 which had rather low completion rates. Without the appropriate nonstudied condition, our best baseline estimate for the items in this experiment is .05 (from Experiment 1 and similar experiments in our lab). All Experiment 4 completion rates were at least three times this value, so we believe that memory influenced performance in all conditions.

**GENERAL DISCUSSION**

The present experiments were designed to help understand the conditions under which perceptual specificity effects arise on a visual stem completion task. In particular, we intended to test the generality of the notion that perceptual specificity effects depend upon encoding tasks that focus attention on visual attributes of target items (e.g., Graf & Ryan, 1990), and to explore the extent to which case effects are facilitated by conditions that yield high levels of explicit memory. Results from individual experiments were somewhat variable. The most consistent and important outcome was observed in Experiments 2 through 4: Case effects were observed after a semantic encoding task when the retention interval was short. Across experiments, 4 of 4 conditions with a short retention interval and unintentional retrieval instructions showed significant case effects after liking ratings. The status of case effects after t-junction counting and a short retention interval was more variable. The short retention, t-junction conditions in Experiments 2 and 4 showed significant case effects, but these conditions in Experiment 3 did not. In long retention interval conditions, significant case effects were never observed after the liking task, but observed in 3 of 4 t-junction conditions. The case effect was only marginally significant in the long retention and t-junction condition of Experiment 4 (p = .09).

Given the small size of the case effects observed in our and other experiments, it is quite likely that these inconsistencies merely reflect a lack of power to detect small differences within individual experiments. Thus, meta-analyses were conducted to estimate the significance of the case effect in each study task by retention interval condition across experiments. Following the recommendations of Rosenthal (1991), the p values obtained from each of the one-tailed t tests on the case effect were converted to z scores. A combined z score was obtained for each condition by summing z across the individual comparisons (n = 4 for both the long and short retention intervals) and dividing by the square root of n. Across the long retention conditions, the case effect was significant after the t-junction task, z = 3.78, p < .0001, and after the liking task, z = 2.06, p < .05. Across the short retention conditions, the case effect was also significant after both the t-junction task, z = 2.76, p < .01, and the liking task, z = 4.49, p < .0001. These meta-analytic z’s can be compared to assess the influence of retention interval and study task on the case effects (z = (z1-z2)/√2, Rosenthal, 1991). Within each retention interval, the case effects did not differ between the encoding tasks: long: z = 1.22; short: z = 1.31. However, long retention intervals lead to a significantly larger case effect than short retention intervals after the liking task, z = 1.73, p < .05, but not after the t-junction task, z = 0.72.

The meta-analytic results suggest that, across experiments, case effects were significant in all conditions. This pattern is inconsistent with a strong interpretation of transfer appropriate processing that would hold that perceptually focused encoding tasks are necessary for the emergence of case effects (e.g., Graf & Ryan, 1990). They are also inconsistent with the suggestion that perceptual specificity effects are most likely to occur when the encoding task pushed perceptual analysis into the background (Jacoby et al., 1992). Rather, the present results suggest that the perceptual versus semantic emphasis of the encoding task is not always a critical determinant of the presence of perceptual specificity effects when case is manipulated in a stem completion paradigm.

Meta-analytic results suggest an interaction between encoding tasks and retention interval. The case effect increased as the retention interval decreased after liking judgments, but the case effect occurred independently of the retention interval after t-junctions. From a psychometric perspective, it is not surprising that case effects would be larger after the short retention interval. Chapman and Chapman (1988; see also Chapman, Chapman, Curran, & Miller, 1994) have shown that the expected value of an accuracy difference (e.g., the difference between same-case and different-case conditions) varies with overall accuracy (e.g., mean completion rate across same case and different case) in an inverted U-shape function that peaks at an overall accuracy near 50%. Thus, difference scores (i.e., case effects) are artifactually inflated when overall accuracy is near 50% compared to when overall accuracy is closer to 0 or 1.

In the long retention interval conditions, average accuracy was 20% (not including the nonstudied condition of Experiment 1). In short retention interval conditions, average accuracy was 52% (not including the intentional retrieval conditions of Experiment 2). Thus, in the liking conditions, the larger case effect after short retention intervals can be understood as a mathematical artifact of accuracy being closer to 50%. The inverted U-shape was empirically observed in the present experiment when case effects were plotted against overall accuracy for individual subjects in Experiments 1 through 4, so this appears to be the best explanation for the results from the liking task. Following the same logic, the case effect also should be larger in the short than long retention conditions after the t-junction task. However, the meta-analysis detected no difference between these conditions, and the trend was actually in the opposite direction. Given the mathematical relationship between accuracy (i.e., completion rate) and difference scores (i.e., case effects), the real mystery of the present results is why the case effect was not influenced by retention interval manipulation after the t-junction task. In the t-junction conditions, there must be some
Visual Specificity and Stem Completion

countervailing – and presumably psychologically more interesting – force that boosts the case effect after long retention intervals compared to short retention intervals.

Whatever the explanation for these findings, it is important to consider the implicit versus explicit nature of the observed effects. In Experiment 1 many subjects reported no awareness of the study-test relationship (even though most subjects noticed completing some stems with words from previous tasks), but the study-test relationship was purposefully revealed to subjects in Experiments 2 through 4. Graf and Ryan’s (1990) transfer-appropriate processing result was replicated in Experiment 1, so it remains possible that perceptual encoding tasks are crucial for obtaining perceptual specificity effects when subjects are unaware of the study-test relationship. By contrast, the major findings of Experiments 2–4 – case effects after semantic encoding and larger case effects with short retention intervals – do not reflect unconscious priming. Subjects were made aware of the study-test relationship, but were nevertheless given instructions for unintentional retrieval (i.e., to respond with the first word that came to mind). Therefore, we interpret the results from Experiments 2–4 as reflecting a kind of unintentional retrieval that is different from intentional, explicit memory.

Only Experiment 2 included a manipulation that directly assessed the relationship between retrieval intentionality and case effects. Significant case effects were observed when subjects were given unintentional retrieval instructions, but not when they were given intentional retrieval instructions. There was a 2-3% trend in the direction of intentional case effects, so it is possible that a more powerful experiment, or meta-analyses of multiple experiments, would reveal a significant case effect. Though perceptual specificity effects are clearly possible on explicit tests (e.g., Graf & Ryan, 1990; Marsolek et al., 1994), Experiment 2 shows that, under equivalent conditions, they are more likely when retrieval is unintentional. This conclusion is most clearly supported by the t-junction conditions where the nonsignificant case effect after intentional retrieval cannot be attributable to ceiling effects or overall accuracy differences compared to the unintentional condition.

We think it is likely that the use of short retention interval especially fostered the emergence of a form of involuntary explicit memory in which unique study episodes are brought to awareness even though the subject is not trying to retrieve these episodes (Richardson-Klavehn et al., 1994a; Richardson-Klavehn et al., 1994b; Schacter, 1987; Schacter et al., 1989). Combined with findings that amnesics do not show normal perceptual specificity effects (Kinoshita & Wayland, 1993; Schacter et al., 1995), it is likely that many published reports of perceptual specificity effects on priming reflect a form of involuntary explicit memory that depends on the medial temporal lobe and diencephalic structures that are damaged in amnesia. These brain mechanisms may be necessary to bind different memorial attributes into a coherent episodic trace (Curran & Schacter, in press; Schacter, 1994; Schacter & Church, 1995). By contrast, priming effects that do not involve binding different memory attributes likely depend on cortically-based perceptual representation system(s) that can operate independently of medial temporal lobe and diencephalic structures. Further investigations into the nature of and relation between these different forms of priming represents an important task for future research.

This research was supported by Grant R01-MH45398 from the National Institute of Mental Health and National Institute on Aging Grant R01-AG08441. We thank Kathy Angell for testing some of the subjects in Experiment 1. We also thank Brad Challis, Shashi Mathew, and Kathleen McDermott for helpful comments and suggestions. Tim Curran is now at the Department of Psychology, Case Western Reserve University. Correspondence concerning this article should be addressed to Daniel L. Schacter, Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138.

References


Sommaire

Les effets de spécificité perceptuelle lors d’épreuves de codage

Un des aspects importants, mais très peu compris de la récupération mnémonique concerne les conditions dans lesquelles l’amorçage est influencé par les changements perceptuels dans la forme des items cibles. Selon la perspective du traitement adapté au transfert, les effets de la spécificité perceptuelle sur l’amorçage requièrent une épreuve d’étude qui oriente l’attention sur les attributs perceptuels des items plutôt que sur leurs attributs sémantiques. D’autres recherches suggèrent que les effets de la spécificité perceptuelle augmentent dans des conditions produisant un haut niveau de mémoire explicite. Dans les présentes expériences, les épreuves de codage (préférence contre décompte des intersections structurales) ont été manipulées ainsi que d’autres variables connues pour influencer la mémoire explicite (répétition et intervalle de rétention) afin d’obtenir des indices sur les déterminants des effets de la spécificité perceptuelle sur le complétèment du trigramme visuel.

Au cours de la première expérience, en utilisant un schéma spécifique aux expériences sur la mémoire implicite, nous avons constaté que les effets de la spécificité perceptuelle (casse de la lettre) sur l’amorçage du complète ment du trigramme sont tributaires du codage perceptuel si les sujets sont plus ou moins conscients de la relation étude-test. La deuxième expérience a été conçue pour créer plus facilement des traces épisodiques accessibles, en donnant deux présentations de stimuli étudiés, sans intervalle de rétention, et en ne testant pas de stèmes non étudiés. Dans ces conditions, les sujets à qui on avait donné des consignes de récupération non intentionnelle démontrèrent des effets de spécificité perceptuelle après des épreuves de codage sémantique et de codage perceptuel.

Cependant, les sujets à qui on avait donné des consignes de récupération intentionnelle ne démontrèrent aucun effet de spécificité perceptuelle. Par conséquent, les effets que démontrèrent les sujets non intentionnels semblaient avoir été causés par la récupération intentionnelle. Dans la troisième et la quatrième expérience, nous avons manipulé les conditions de base différentes dans l’expérience 1 et l’expérience 2 – le nombre de présentations (1 contre 2 dans l’expérience 3) et la durée de l’intervalle de rétention (< 30 s contre 10 min dans les expériences 3 et 4) – pour isoler les variables qui déclenchent les effets de spécificité perceptuelle après un codage sémantique. Ces expériences ont révélé des effets de spécificité perceptuelle seulement lorsque l’épreuve de codage sémantique est combinée à un intervalle de rétention court. Après les épreuves de codage perceptuel, les effets de spécificité perceptuelle étaient moins constants.

Une méta-analyse des conditions de récupération non intentionnelle de toutes les expériences a permis de déceler des effets de casse (contraste majuscule minuscule) significatifs dans toutes les épreuves de codage dans des conditions d’intervalle de rétention, mais les effets de casse sur les plus marqués suivissaient l’épreuve de codage sémantique et étaient possiblement attribuables aux taux de complètement généralement plus élevés après le codage sémantique. On conclut que les épreuves de codage perceptuel ne sont pas nécessairement un prérequis pour l’observation des effets de spécificité perceptuelle. Toutefois, les effets de spécificité perceptuelle qui suivent les épreuves de codage sémantique pourraient refléter une forme de mémoire explicite involontaire, plutôt qu’une mémoire implicite réelle.

Revue canadienne de psychologie expérimentale, 1996, 50:1, 33