Rice University

Selective Remembering in an Orienting-Task Paradigm

by

Madeline Campbell

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Master of Arts

Approved, Thesis Committee:

[Signatures]

Michael J. Watkins, Professor, Chair
Psychology

David M. Lane, Associate Professor
Psychology & Statistics

Denise Chen, Assistant Professor
Psychology

Houston, Texas
May 2004
ABSTRACT

Selective Remembering in an Orienting-Task Paradigm

by

Madeline Campbell

The levels-of-processing effect is extraordinarily, even puzzlingly, robust. It occurs even when a memory test is expected, ample study time is given, and deep processing is encouraged regardless of orienting task. Thus, processing appears to be “fixed” by the requirements of the orienting task. This enigma is explored in a selective remembering procedure involving the recall of words of arbitrarily varying values. After verifying that selectivity is substantially localized at the encoding rather than the recall stage of the remembering process (Experiment 1), recall was found to be selective despite the imposition of an orienting task and regardless of whether item values were assigned according to the items’ semantic category (Experiment 2) or entirely at random (Experiment 3). Indeed, no evidence was found for any effect at all of orienting tasks on selectivity (Experiment 4). Orienting tasks do not, after all, universally dominate the encoding process.
Acknowledgments

Many, many thanks to my advisor, Mike Watkins, for his guidance, wisdom, humor, support, and sheer hard work, and to Yanliu Huang, for being the best lab-mate a girl could hope for.

Thank you to my committee: David Lane, for his brilliance and patience in explaining sphericity just one more time, and Denise Chen, for her insightful comments and fresh perspective.

Thanks to everyone who helped with this project: Meredith Jenkins, Briony Schnee, Cristina Rubino, and Allison Dill.

Big love and thanks to all my family and friends, who have always been there for me.

Special love and thanks to Kevin, for your love, encouragement, and belief in me…it means so much.

This thesis is dedicated to my wonderful, loving parents, Pauline & Michael, and my fantastic brother, Steven, for their endless support across the miles…I couldn’t have done it without you.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>A Puzzle</td>
<td>4</td>
</tr>
<tr>
<td>Volitional Control of Memory</td>
<td>7</td>
</tr>
<tr>
<td>The Selectivity Index</td>
<td>12</td>
</tr>
<tr>
<td>Localization of Selectivity</td>
<td>15</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>16</td>
</tr>
<tr>
<td>Method</td>
<td>17</td>
</tr>
<tr>
<td>Results</td>
<td>19</td>
</tr>
<tr>
<td>Discussion</td>
<td>21</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>21</td>
</tr>
<tr>
<td>Method</td>
<td>21</td>
</tr>
<tr>
<td>Results</td>
<td>25</td>
</tr>
<tr>
<td>Discussion</td>
<td>30</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>31</td>
</tr>
<tr>
<td>Method</td>
<td>32</td>
</tr>
<tr>
<td>Results</td>
<td>33</td>
</tr>
<tr>
<td>Discussion</td>
<td>37</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>38</td>
</tr>
<tr>
<td>Method</td>
<td>39</td>
</tr>
<tr>
<td>Results</td>
<td>42</td>
</tr>
<tr>
<td>Discussion</td>
<td>47</td>
</tr>
<tr>
<td>General Discussion</td>
<td>48</td>
</tr>
<tr>
<td>References</td>
<td>52</td>
</tr>
<tr>
<td>Appendix A</td>
<td>58</td>
</tr>
</tbody>
</table>
Appendix B ........................................................................................................66

Appendix C ........................................................................................................67
List of Tables

Table 1  Mean Selectivity Indices (SI) and Mean Percentage of Words Recalled (Experiment 1)

Table 2  An example of the material presented in the control, shallow, and deep encoding conditions

Table 3  Mean Selectivity Indices (SI) and Percentage of Words Recalled (Experiment 2)

Table 4  Mean Selectivity Indices (SI) and Percentage of Words Recalled (Experiment 3)

Table 5  Mean Selectivity Indices (SI) and Percentage of Words Recalled, as appropriate, for the orienting-tasks-plus-selectivity (LOP&SEL) group, the orienting-tasks-only (LOP) group, and the selectivity-only (SEL) group (Experiment 4)
List of Figures

Figure 1  Boxplots of selectivity indices for study-test and test conditions (Experiment 1)

Figure 2  Selectivity indices for control and orienting task conditions (Experiment 2)

Figure 3  Selectivity indices for control, deep, and shallow conditions (Experiment 2)

Figure 4  Recall in control, deep, and shallow conditions (Experiment 2)

Figure 5  Mean recall of 10-point words (Experiment 2)

Figure 6  Mean recall of 1-point words (Experiment 2)

Figure 7  Selectivity indices for control, deep, and shallow conditions (Experiment 3)

Figure 8  Recall in control, deep, and shallow conditions (Experiment 3)

Figure 9  Recall of 10-point words in control, deep, and shallow conditions (Experiment 3)

Figure 10  Recall of 1-point words in control, deep, and shallow conditions (Experiment 3)

Figure 11  Selectivity indices for the control, deep, and shallow conditions of the orienting-tasks-plus selectivity (LOP&SEL) group and for the selectivity-only (SEL) group (Experiment 4)

Figure 12  Percentage of words recalled in each condition for the orienting-tasks-plus-selectivity (LOP&SEL) group, the orienting-tasks-only (LOP) group, and the selectivity-only (SEL) group (Experiment 4)

Figure 13  Percentage of 10-point words recalled in control, deep, and shallow conditions (LOP&SEL, Experiment 4)
Figure 14 Percentage of 1-point words recalled in control, deep, and shallow conditions (LOP&SEL, Experiment 4)
Selective Remembering in an Orienting-Task Paradigm

The levels-of-processing framework proposed in 1972 by Craik and Lockhart is widely regarded as one of the most influential perspectives in cognitive psychology. Certainly, their paper is among the most cited in all of psychology. Roediger and Gallo (2001, p. 44) have called it, “one of the most valuable developments in the psychology of memory in the 115-year history of its empirical study.” And yet it has been heavily criticized (e.g., Baddeley, 1978; Craik & Tulving, 1975; Eysenck, 1978; Morris, Bransford, & Franks, 1977; Nelson, 1977) and has left some puzzling findings in its wake. This paper focuses on one such puzzle and describes a series of experiments to explore its scope.

Level of processing of target items is traditionally manipulated by the use of orienting tasks. Typically, semantic orienting tasks, such as rating pleasantness, judging category, or judging sentence congruity, are assumed to induce deep processing; tasks that focus attention on physical characteristics of the word, such as identifying case (upper or lower) or deciding whether it includes a given letter, are assumed to induce shallow processing. Semantic tasks, even when imposed under incidental learning conditions (i.e., with the upcoming memory test presumed to be unanticipated), lead to roughly equal, and occasionally slightly better, performance than the absence of any orienting task. By contrast, memory following a shallow task is impaired relative to memory following unfettered study of equal duration (e.g., Hyde & Jenkins, 1973). This
pattern has been replicated many times even with considerable procedural variation (Craik & Tulving, 1975). In short, the levels-of-processing effect has proved to be a remarkably robust phenomenon both in magnitude and generality.

The levels-of-processing approach has not, however, been without its critics. Craik and Lockhart have responded to these critics, conceding some points and disputing others (e.g., Lockhart, Craik, & Jacoby, 1976; Lockhart & Craik, 1978; Lockhart & Craik, 1990; Lockhart, 2001). For example, Nelson (1977) argued that the lack of an independent measure of depth renders the concept circular and unfalsifiable: deeper processing is presumed to yield better memory, and better memory is presumed to arise from deeper processing. Lockhart and Craik (1990) countered that depth, operationally defined by the orienting tasks, may be specified by an a priori analysis of the processes involved in performing the task. Thus, low-level structural analysis of the sort called for in a “shallow” task is a prerequisite of semantic analysis, but not vice versa. Further, even when the notion of levels of processing is invoked as a post hoc analysis, and therefore exhibits a degree of circularity, it is in the company of other significant scientific theories, such as the Darwinian theory of natural selection (Watkins, 2002).

Others have stressed the limitations of the levels-of-processing framework, particularly its failure to account for retrieval effects. For example, Morris et al. (1977) found that retrieval context was an important determinant of performance. Specifically, they showed that a rhyming recognition test, in which subjects were asked at test whether
a test word rhymed with a target word from a previously presented list, was more
effective for items studied in a phonetic (rhyme) context than for items studied in a
semantic context. Lockhart and Craik (1990, p. 101) “strongly endorse” the notion of
transfer-appropriate processing advocated by Morris et al., although they still maintain
that “some types of initial processing are superior to others, and these types involve
deeper, semantic types of analysis.” In support, they cite Fisher and Craik’s (1977)
finding that semantic cues are more effective for words studied in a semantic context than
rhyme cues are for words studied in a rhyme context. On the other hand, Lockhart and Craik (p. 109) freely concede that the levels-of-processing construct is far from a
complete theory of memory, and should be regarded as merely a “heuristic framework,”
useful for stimulating further research.

Concerns about the metatheoretical status and limited applicability of the levels-
of-processing construct are one thing, but concerns about its adequacy as an account of
the effects of orienting tasks are quite another. And yet there are a number of findings
from the orienting task procedure that are not readily accounted for by the levels-of-
processing construct. These findings have been catalogued in a thoughtful review by
Roediger and Gallo (2001). Prominent among the intriguing questions raised by
Roediger and Gallo is why the levels-of-processing effect persists under conditions of
intentional learning. This enigma is the focus of this thesis.
A Puzzle

Craik and Lockhart (1972, p.681) recommended that levels-of-processing experiments be administered under incidental learning conditions, thus optimizing the experimenter's control over the subjects' encoding operations. It is reasonable to presume that, when alerted to an imminent memory test (intentional learning), subjects would seek to maximize their recall by deeply processing all words, not just those subjected to a deep orienting task. Such a strategy should, according to the levels-of-processing framework, mask the effect of a shallow orienting task and thereby eliminate the differential effects of orienting tasks. But this does not happen. The levels of processing effect occurs even under intentional learning instructions (e.g., Chow, Currie, & Craik, 1978; Craik & Tulving, 1975; Hyde & Jenkins, 1969).

This finding still puzzles researchers: subjects know that their memory for the stimulus words will be tested, and presumably they could easily process words that had been assigned a shallow orienting task to a deeper level and thereby enhance their likelihood of being recalled in the test. This puzzle has been widely noted. As Craik and Tulving (1975, p.279) have remarked, "surely under intentional learning conditions the subject would analyze and perceive the name and meaning of the target word with all three types of question. In this case equal retention should ensue." Chow et al. (1978, p.109) put it this way: "Why is it that subjects performing the shallow structural task under intentional learning do not carry out further processing in order to enhance their recall?"
And Roediger and Gallo (2001, p.34) asked, “When subjects are given so-called shallow orienting tasks under intentional learning instructions, why doesn’t recall increase dramatically?”

Craik and Tulving (1975, Experiment 2) wondered whether such failure is due to insufficient time, since shallow tasks typically take less time to complete than deeper tasks. Perhaps subjects do not have enough time after completing the shallow task to engage in deeper processing. To test this possibility, they presented words at a slow rate: 1-second word presentations with 5-second intervals between words for a 6-second presentation rate. Yet they still found a typical levels-of-processing effect (Experiment 9). Chow et al. (1978) reported an effect with an even slower (12-second) presentation rate. Finally, Craik and Tulving (1975, Experiment 5) rejected the processing time hypothesis in another way: they found a typical levels-of-processing effect even when the shallow task was such that it took longer to complete than the deep task.

Perhaps subjects are just not sufficiently motivated to give items subjected to a shallow task additional and more effective processing. Craik and Tulving (1975, Experiment 10) rejected this possibility when they found a typical levels-of-processing effect for subjects given more financial incentives to remember shallow words than deep words.

Another possibility is that subjects are unaware of how best to prepare for a memory test and therefore fail to engage in deeper processing of words in the shallow
condition. Chow et al. (1978) explored this possibility by giving their subjects a brief lecture on levels-of-processing and instructions to generate an adjective for every word (a semantic task) to assist their performance in an subsequent recognition test. But the levels-of-processing effect remained.

The levels-of-processing effect survives even a reversal of the standard orienting task procedure. Normally, the orienting question (e.g., “Is the word in uppercase?”) precedes target word presentation (FLOWER). With a reverse procedure, the target word is given first (FLOWER), and only after a delay is the orienting question given (“Is the word in uppercase?”). Under these conditions the target words could be deeply processed prior to the imposition of the orienting task, and yet the levels-of-processing effect still occurs (Craik, 1977).

A levels-of-processing effect with a reverse orienting task procedure poses much the same problem as it does under intentional learning conditions. As Roediger and Gallo (2001, p. 35) observed, both findings raise the question of why, given “ample time and conditions designed to encourage meaningful processing, the effect of orienting tasks—the ‘levels-of-processing effect’—still occurs.”

Why do orienting tasks exert such power? The question remains baffling. Roediger and Gallo (2001, p.35) asked: “Does the shallow orienting task short-circuit or curtail subjects’ ability to engage voluntarily in meaningful processing [or does it]
...create some inhibitory process that semantic processing cannot overcome?" Either way, the shallow orienting task would block additional and more effective processing.

Craik and Tulving (1975, p. 292) have asked, "to what extent are the encoding operations performed on an event under the person's volitional strategic control?" The research to be reported here explores this question as it applies to the apparent blocking effect of orienting tasks.

**Volitional Control of Memory**

The ability to covertly control the remembering process has been relatively little investigated, even though this skill has obvious adaptive qualities, can be used to avoid excessive cognitive load (Golding, Roper, & Hauselt, 1996), and is otherwise critical for an efficient memory system (Castel, Benjamin, Craik, & Watkins, 2002). Most research on memory has been concerned with the effect of external factors. For example, the probability of recalling an item is measured as a function of number of items presented, its serial position, presentation rate, contextual cues, and the administration of a "distractor" task. Of course, it can be confidently concluded without research that we do have some, but not complete, control over our memory. We can learn vital facts for an exam or remember a person's name, perhaps through the use of rehearsal (Baddeley, Lewis, & Vallar, 1984), mnemonics (Ericsson & Staszewski, 1989), or other control processes; conversely, intentions to remember often fail, as do intentions to forget.
Such empirical evidence as has been reported confirms that recall can, at least to some extent, be selective in response to a reward or a forget instruction. For example, Harley (1965) found that paired-associates learning was faster for paired-associates that carried a 25-cents reward than for those that carried no reward. However, this effect required a relatively long delay between pair presentations: it occurred with a 4-second delay, but not with a 1.3-second delay. Similarly, Weiner and Walker (1966) found incentive conditions (5 cents, or avoiding a shock) affected the recall of consonant-vowel-consonant (CVC) trigrams, but only with a relatively long period of interpolated activity (15 seconds of digit reading). Cuvo (1974) varied the incentive level of stimulus words (10-cents or 1-cents) and found greater recall of 10-cents words. Cuvo also found that, when asked to rehearse words overtly, subjects both rehearsed and recalled more high incentive words than low incentive words.

In the experiments of Harley (1965), Weiner and Walker (1966), and Cuvo (1974), incentive was signaled by the color of the background against which the target items were displayed. Thus, the incentives were given simultaneously with the targets themselves. Tarpy and Glucksberg (1966), on the other hand, compared the effect of signaling incentives simultaneously with the presentation of the target items to that of signaling after the target presentation. Simultaneous cueing was more effective than delayed cueing. Cuvo and Witryol (1971) also compared the effect of simultaneous and delayed incentive cueing. Their subjects, it might be noted, were fifth-grade children,
and other research (Cuvo, 1974) has shown incentive effects to be reduced in such young subjects. Nevertheless, in a word recall task, they found that the advantage of 10-cents words over 1-cent words was eliminated when the incentive cues were delayed until the recall test.

In a comparatively elaborate paired-associates experiment, Wickens and Simpson (1968) compared the effects of three incentives (positive feedback for a correct response, 5 cents for a correct response, and a mild shock for an incorrect response) signaled at one of three phases of the procedure (during presentation of the target items, during a distractor task between presentation and recall, or at recall). The 5-cents reward or the avoidance of a shock resulted in better recall than feedback alone, but only when the cue and target items were presented simultaneously. Wickens and Simpson also observed that performance in the interpolated distractor task (counting backwards) was reduced, which prompted the suggestion that the effect of the incentives was mediated by rehearsal.

Not all studies have failed to find a delayed incentive effect. Using a continuous paired-associate task, Loftus and Wickens (1970) reported better recall of high value trigrams (worth 99 points each) than low value trigrams (worth 22 or 11 points each) regardless of whether the points values were signaled during the study phase or a test phase. The effect was, however, reduced when the signal was delayed until test.
Another technique designed to shed light on volitional control over the remembering process is the directed forgetting paradigm. In a typical experiment, subjects are presented with a list of items, each immediately followed by a signal either to forget the item or to remember it for a later recall test. Contrary to what they had been led to believe, the subjects are asked to recall both the “remember” and the “forget” items. The principal finding is that items followed by a “remember” signal are more likely to be recalled than items followed by a “forget” signal (Davis & Okada, 1971; Woodward & Bjork, 1971; Woodward, Bjork, & Jongeward, 1973).

In a variation of the directed forgetting method, Golding, Roper, and Hauvelt (1996) presented subjects with items that were signaled as having either a 0%, 50%, or 100% probability of inclusion in an upcoming recognition test. As predicted, recognition of items signaled with 100% probability was better than items signaled with 50% probability, which in turn was better than items signaled with 0% probability. From “think-aloud” protocols, produced in overt on-line introspection of study strategies, the authors concluded that subjects used a betting strategy, where only some of the 50% words received additional processing, and those that did were treated like the 100% words.

More relevant to present purposes is a study in which the effect of three orienting tasks on directed forgetting was investigated (Wetzel, 1975). In a between-subjects design, words were followed by a 5-second interval in which subjects performed one of
three orienting tasks: writing the word itself (rehearsal condition), writing a rhyme of the word (rhyme condition), or writing an associated adjective or noun (adjective-noun condition). Subjects then heard either a tone or click, indicating that the word was to be remembered or forgotten. In a subsequent recall test for the words, an effect of orienting task was observed only for the remember words, with recall in the rehearse and rhyme conditions exceeding that in the adjective-noun condition.

Finally, volitional control has been explored in a “selectivity” paradigm developed by Watkins and Bloom (1999). Words were arbitrarily assigned values of from 1 to 12 points, and the subject’s task was to maximize the points value of their recall. A selectivity index was calculated according to a ratio scale ranging from −1.0 to +1.0, where 0.0 indicates an absence of selectivity, and +1.0 indicates maximum selectivity. Overall, recall was moderately selective (.47), but selectivity increased with increasing disparity in the points value, and with simultaneous rather than sequential presentation of the words. Somewhat surprisingly, selectivity did not increase with additional study time, even with simultaneous presentation of words. These results demonstrate that subjects did have some degree of selectivity in, and hence of control over, their recall.

Recall Roediger and Gallo’s (2001) hypothesis that the orienting tasks of a typical levels-of-processing procedure block additional processing, even under conditions of abundant time (Craik & Tulving, 1975), intentional learning (Hyde & Jenkins, 1969),
knowledge of the levels-of-processing effect, and encouragement to engage in deeper processing (Chow et al., 1978). They conclude that orienting tasks exert full control over processing in memory, to the exclusion of volitional control. If this conclusion is correct, then the imposition of an orienting task in Watkins and Bloom’s (1999) procedure should eliminate selective recall. The orienting task should trump selectivity. This research project was designed to test this prediction.

**The Selectivity Index**

As mentioned earlier, Watkins and Bloom (1999) developed a method to assess selectivity in, and thus control over, the memory process. They devised a selectivity index (SI) ranging from −1.0 to +1.0, where 0.0 indicates an absence of selectivity, and +1.0 indicates maximum selectivity. An SI of +1.0 is obtained when each recalled item has equal or greater value than each unrecalled item. An SI of −1.0 is obtained when each recalled item has equal or lesser value than each unrecalled item. SI can be calculated whenever items to be recalled have quantifiable positive values that are not all equal.

A selectivity index is based on the subject’s actual score, relative to chance and ideal performance given the number of items recalled. It is derived as follows:
\[ SI = \frac{\text{actual score} - \text{chance score}}{\text{ideal score} - \text{chance score}} \]  

[Equation 1]

Where: *actual score* is the total value of the recalled items

*chance score* is the total of the recalled items expected in the absence of selectivity (average value of presented items x number of items recalled)

*ideal score* is the maximum possible value given the number of items recalled

(obtained by summing the highest n values of the presented items, where n is the number of items recalled).

An SI of +1.0 (maximum selectivity) occurs when the actual and ideal scores are equal;
an SI of 0.0 (absence of selectivity) occurs when the actual and chance scores are equal.

For example, suppose that subjects are presented with lists of 20 words, 10 of which are assigned a value of 10 points each and 10 a value of 1 point each. Suppose that for one of the lists, a given subject recalled six 10-point words and two 1-point words.

The actual value of their recall is \(10 + 10 + 10 + 10 + 10 + 1 + 1 = 62\) points. The average value of the presented words is \((10 + 1) / 2 = 5.5\), and since the subject recalled 8 words, the chance value is \(5.5 \times 8 = 44\). The ideal value, given that 8 words were recalled, would be \(10 + 10 + 10 + 10 + 10 + 10 + 10 = 80\). Therefore, the selectivity index would be \((62 - 44) / (80 - 44) = .5\)

A weighted average option takes account of the number of items recalled in each list, which gives more weight to lists for which roughly half of the items are recalled than
to lists for which either very few or almost all of the items are recalled. Each list is
assigned a weight \( w_i \) such that:

\[
w_i = n_i (N_i - n_i) \quad \text{[Equation 2]}
\]

Where: \( n_i \) is the number of words recalled from list \( i \), and
\( N_i \) is the number of items in list \( i \).

Using the hypothetical example above, \( w = 8 \times (20 - 8) = 96 \).

The weighted selectivity indices may then be combined to form an overall estimate of the
selectivity index for each subject, as follows:

\[
SI = \frac{\sum_{i=1}^{L} w_i S_i}{\sum_{i=1}^{L} w_i} \quad \text{[Equation 3]}
\]

where: \( L \) is the number of lists
\( w_i \) is the weight for list \( i \)
\( S_i \) is the selectivity index for list \( i \)

In the experiments reported here, selectivity was calculated for each subject for each list,
using the weighted average option. Thus, for each subject, an index of selectivity was
computed for each combination of conditions, by taking a weighted average of the
selectivity indices for the recalled lists within that combination of conditions. The resulting weighted averages were then subjected to inferential statistical analyses.

Localization of Selectivity

The levels of processing construct pertains to the encoding phase of memory. Craik and Lockhart (1972) argued that a memory trace should not be regarded as the product of a specialized encoding process, but rather as a “byproduct” of perceptual analysis (Craik & Lockhart, 1972, p. 681) or, more broadly, of “general cognition, rather than just of perception in the narrow sense” (Lockhart, 2001, p. 100). The levels of processing framework says nothing about the recall phase of the remembering process\(^1\).

Studies of selectivity have not yet established the locus of selective processing. For example, Watkins and Bloom (1999) noted that, in their series of six experiments, the temporal locus of selectivity is uncertain, as values were assigned at the study phase of the procedure. Similarly, in their study on the effects of aging on selectivity, Castel et al. (2002, p.1085) observed that, “the question arises as to whether selectivity reflects encoding or retrieval processes,” although they suspected encoding processes were the more likely source.

\(^1\) This exclusive focus on encoding has been regarded as a major deficiency in the levels-of-processing framework by advocates of transfer-appropriate processing (Morris et al., 1977) and the encoding specificity principle (Tulving, 1983). Lockhart and Craik (1990) replied that retrieval is indeed a central focus in memory research, but that the levels-of-processing framework simply does not address questions relating to retrieval, only those relating to encoding.
Other studies of selective recall offer mixed evidence regarding its temporal locus. The majority of incentive-based experiments indicate that differential recall occurs only when the incentives are signaled during target presentation (e.g., Tarpy & Glucksberg, 1966; Wickens & Simpson, 1968), suggesting selectivity has an encoding locus. On the other hand, evidence of selectivity at retrieval has also been reported (Loftus & Wickens, 1970).

Given that orienting tasks have their effect at the time of target study, it would seem prudent to establish that selectivity induced by the procedures used in the research to be reported here is also, at least substantially, achieved at the time of study. With such evidence in hand, these procedures may be used to address the question of primary interest, namely the effect of orienting tasks on selectivity. Experiment 1 was therefore designed to check that selectivity arose, at least substantially, during the study phase of the procedure.

**Experiment 1**

In a typical selectivity experiment, words are displayed along with their values (points) during the study phase of the procedure. To localize selectivity, selectivity with this procedure needs to be compared with selectivity when the values are withheld until the test phase. To enable this manipulation, words in a given list were exemplars of one of two categories, such as birds and fruits, with the exemplars of one category being given one value and those of the other category another value. At issue was whether
recall would be more selective when the category values were indicated during the study phase as well as at test than when they were indicated only during the test phase.

**Method**

*Subjects.* Forty Rice University undergraduates participated for course credit.

*Materials.* To create the stimulus word lists, 10 exemplars were chosen at random from each of the 32 categories (e.g., birds, fruits, sports, house parts) listed in the Toronto Categorized Word Pool (Murdock, 1976; see Appendix A). Categories were paired at random within the constraint that closely related categories (e.g., fish and animals) were not paired. This resulted in 16 20-word lists. The first two lists, unbeknown to the subjects, were designated as practice lists and excluded from the data analysis. The words were presented in Monaco size 60 font.

*Design.* Selectivity was assessed under two conditions: when category values were given prior to the study phase and again at test (study-test condition), and when category values were given only at test (test condition). Thus, in the study-test condition, subjects were told immediately prior to its presentation that the first list would comprise exemplars of birds and fruits categories and that the birds were each worth 10 points and fruits were each worth 1 point (or vice versa), and they were reminded of these values at the beginning of the test phase. In the test condition, subjects were told for the study phase only that they would see exemplars of birds and fruits, and were given the category values only at the time of test.
The conditions were manipulated in a 2 (study-test/test) x 2 (10 points/1 point) within-subjects design. The order of the lists remained constant across subjects, as did the order of the words within each list. List condition changed for each list, such that for half the subjects odd-numbered lists were in the study-test condition and even-numbered lists were in the test condition, and for the other half the lists were in the complementary arrangement. The assignment of points (10 or 1) to categories was also counterbalanced. Thus, across subjects, each category was assigned to each of the four combinations of the two variables equally often.

Procedure. The subjects were given detailed instructions and encouraged to ask questions. They were told repeatedly that their task was to maximize the points value of their recall, rather than the number of words recalled. They proceeded through the experiment by clicking a mouse.

The two included categories were identified in advance of each list. For half of the lists (study-test condition), the category values were also specified in advance of the list. For the remaining half of the lists (test condition), the category values were specified only after presentation of the list. Thus, the test phase was identical for the study-test and test conditions. The 20 words of a list were presented on a computer monitor, one at a time at a rate of one word per second. The last word of each list was followed by a row of 8 randomly ordered digits, shown for 3 seconds. The subjects then attempted to recall these digits in order. Regardless of list condition, subjects were then informed which
category's exemplars were worth 10 points and which were worth 1 point. They were then asked to write the exemplars on a recall sheet, and were reminded that their task was to maximize their recall points. The recall phase was untimed; the subjects clicked the mouse when they considered themselves to have finished.

This brought up an alphabetically sorted display of the words from each category, which the subject used to calculate their score for the list; they wrote their score on the recall sheet. The purpose of this self-scoring was to reinforce the subject's focus on points, rather than on number of items recalled. The subjects then proceeded to the next list, and so on until all 16 lists had been studied and tested.

**Results**

Table 1 shows the mean selectivity index and mean percentage of words recalled for the study-test and test conditions. Clearly, withholding values until the test sharply reduced selectivity while having little effect on recall.

\[\text{Table 1: Mean Selectivity Indices (SI) and Mean Percentage of Words Recalled (Experiment 1)}\]

<table>
<thead>
<tr>
<th></th>
<th>Study-test</th>
<th>Test</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI</td>
<td>.37</td>
<td>.10</td>
<td>.24</td>
</tr>
<tr>
<td>% recall</td>
<td>52.47</td>
<td>51.02</td>
<td>51.25</td>
</tr>
</tbody>
</table>
**Selectivity.** Both mean selectivity indices were significantly greater than the chance value of zero: for the study-test condition, \( t(39) = 12.27, p < .0001 \); and for the test condition, \( t(39) = 5.50, p < .0001 \). The difference between the two selectivity indices was also statistically significant, \( t(39) = 9.18, p < .0001 \) (see Figure 1\(^1\)).

![Boxplots of selectivity indices for study-test and test conditions (Experiment 1)](image)

*Figure 1.* Boxplots of selectivity indices for study-test and test conditions (Experiment 1)

**Recall.** Overall, subjects recalled approximately half of the words. The proportion was slightly greater in the study-test condition than in the test condition, the difference being of equivocal significance, \( t(39) = 1.99, p = .05 \).

---

\(^1\) The variance shown in all boxplots is within-subjects. Specifically, each subject’s mean across conditions (here, the mean of study-test and test selectivity indices) was subtracted from their score on each condition. Then the overall mean for all conditions and across all subjects was added to each score. See Appendix B for details of boxplots.
Discussion

Selectivity was much greater in the study-test condition than in the test-only condition, which means that the selectivity in the study-test condition was substantially localized at the study phase of the procedure. This finding validates the use of selectivity indices obtained with a study-test procedure to test the hypothesis that orienting tasks—which are imposed in the study phase—block additional processing and preclude selective remembering.

Thus, in the remaining experiments the selectivity task was superimposed on an orienting task procedure, in which the target items were subject to deep, shallow, or no orienting task. If, as has been suggested on the basis of previous levels-of-processing experiments (Roediger & Gallo, 2001), orienting tasks totally control, or fix processing, then selectivity should occur only for the targets not subject to an orienting task.

Experiment 2

Method

Subjects. Twenty-four college-educated people, aged from 18 to 28 years old, participated. They included 19 Rice University undergraduates who participated for course credit.

Materials. The lists were expanded versions of 11 of the lists used in Experiment 1, with the number of words of each category increased to 12 (see Appendix C). For any
given subject, 4 of these 12 were assigned to the deep condition, 4 to the shallow condition, and 4 to the control condition. The first list served for practice.

Target words were displayed in Verdana size 30 font, one at a time, at the top of a computer screen, sometimes in upper and sometimes in lower case. The words of each list were numbered 1 to 24. For the control condition, only the target word and its serial number were shown. For the shallow condition, *is in upper case* was added underneath. For the deep condition, an incomplete sentence and *makes sense* were added; when completed with the target word, the sentence either made sense or did not. These presentation conditions are illustrated in Table 2.

*Table 2: An example of the material presented in the control, shallow, and deep encoding conditions*

<table>
<thead>
<tr>
<th>Condition</th>
<th>1. PEACOCK</th>
<th>1. PEACOCK</th>
<th>1. PEACOCK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td>Is in uppercase</td>
<td></td>
</tr>
<tr>
<td>Shallow</td>
<td>1. PEACOCK</td>
<td></td>
<td>The --- was proud and colorful</td>
</tr>
<tr>
<td>Deep</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The response sheet for the orienting tasks consisted of 11 tables of three columns: the first column showed the serial position of the words (1 to 24), and the second and third were blank spaces headed *Yes* and *No*. The recall sheet consisted of 11 columns of
24 blank lines. Each column was headed by the list number and a row of 8 blank spaces for digit recall.

*Design.* The experiment conformed to a 3 (deep processing/shallow processing/control) x 2 (10 points/1 point) x 2 (yes/no response) within-subjects design. All subjects received the same randomized order of the lists, and the same randomized order of the words within each list. The value of each category (10 points or 1 point per exemplar), however, was counterbalanced across subjects. For example, for half of the subjects the birds exemplars were each worth 10 points and the fruits exemplars were each worth 1 point, whereas for the other half of the subjects these values were reversed. Within each of these subject groups, the encoding (deep, shallow, control) and, for the deep and shallow conditions, response (yes/no) condition of each target word were rotated between subjects such that over all subjects each condition occurred equally often.

*Procedure.* The subjects were given instructions and encouraged to ask questions. It was stressed that their main task was to maximize the sum of the points of the recalled words rather than the number of words recalled. This instruction was repeated on the computer screen frequently during the experiment. After a practice trial the subjects were again invited to ask questions.

Before each list, the two categories comprising the list were identified, as were the respective values of their exemplars. The 24 words of the list then appeared on the
computer screen at a rate of 1 every 5 seconds. If only the word was shown, the subjects put a line through its number on their response sheet; if the word was shown with *is in upper case* or an incomplete sentence and *makes sense*, the subjects checked the “yes” or “no” box as appropriate.

Each list was followed immediately by a row of 8 digits, displayed on the screen for 3 seconds. The subjects attempted to recall the digits in left-to-right order on their recall sheets. The intent here was to eliminate recall of the list words on the basis of primary memory. The subjects were reminded of the two categories in the list and the points value of their exemplars, and allowed 90 seconds to recall the target words, in any order, on the recall sheet. The computer screen then flashed and displayed the target words in alphabetic order along with their point values, and the subjects changed from pencil to pen (reducing any temptation to add to their recall) and wrote the sum of their points at the bottom of their recall sheet. The purpose of asking the subjects to calculate their own total score was to help maintain their focus on the primary task—maximizing their points.

On completion of the scoring task, subjects proceeded to the next list, and repeated the procedure though the 10 lists of the experimental proper. They were then thanked, debriefed, and dismissed.
Results

Selectivity and word recall were examined as a function of encoding condition: deep, shallow, and control. The selectivity index for each condition was calculated for each subject within each list and then combined across lists, according to Equation 3. Table 3 summarizes the mean selectivity indices and percentage of words recalled.

*Table 3: Mean Selectivity Indices (SI) and Percentage of Words Recalled (Experiment 2)*

<table>
<thead>
<tr>
<th></th>
<th>Deep</th>
<th>Shallow</th>
<th>Control</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI</td>
<td>.32</td>
<td>.42</td>
<td>.37</td>
<td>.37</td>
</tr>
<tr>
<td>% recall</td>
<td>52.88</td>
<td>45.50</td>
<td>50.00</td>
<td>49.50</td>
</tr>
</tbody>
</table>

*Selectivity.* The mean selectivity index exceeded zero to a statistically significant extent both overall and in each individual condition: overall, $t(23) = 13.11, p < .0001$; deep, $t(23) = 10.89, p < .0001$; shallow, $t(23) = 11.24, p < .0001$; control, $t(23) = 7.94, p < .0001$.

When the deep and shallow conditions were collapsed, creating an ‘orienting task’ condition, selectivity was equal to that found in the control condition (see Figure 2).
Figure 2. Selectivity indices for control and orienting task conditions (Experiment 2)

Although an ANOVA failed to show a significant omnibus effect of encoding condition, $F(1.89, 43.45)^1 = 2.72, p = .08$, a pairwise comparison indicated that selectivity was significantly greater in the shallow condition than in the deep condition: $t(23) = 2.39, p = .03$. No other pairwise comparisons achieved statistical significance: control versus shallow, $t(23) = 1.31, p = .20$; control versus deep, $t(23) = 1.06, p = .30$ (see Figure 3).

---

$^1$ For $F$ values that substantially violate the assumption of sphericity, the degrees of freedom and $p$ value are calculated in accordance with the Huynh-Feldt univariate test (Huynh & Feldt, 1976), rather than the standard $F$ test. The Huynh-Feldt test is more appropriate for tests of within-subjects variables with three or more levels, because the ANOVA’s sphericity assumption is typically violated (Huynh & Mandeville, 1979). The Huynh-Feldt test takes this violation into account by adjusting the degrees of freedom.
Recall. Overall, the average number of words recalled per 24-word list was 11.87 (49.5%). Figure 4 shows the mean level and variability of recall separately for each of the three conditions. The difference among the means was statistically significant, $F(2, 46) = 6.95, p = .002$. More words were correctly recalled in the deep condition than in the shallow condition, $t(23) = 3.48, p = .002$; and more were recalled in the control condition than in the shallow condition, $t(23) = 2.51, p = .02$. The difference between deep and control conditions was not statistically significant, $t(23) = 1.42, p = .17$. 

Figure 3. Selectivity indices for control, deep, and shallow conditions (Experiment 2)
Figure 4. Recall in control, deep, and shallow conditions (Experiment 2)

A separate analysis of the 10-point words showed no statistically significant levels-of-processing effect, $F(2, 46) = 1.11, p = .34$ (see Figure 5). Nor were any pairwise comparisons significant: shallow words (60.25%) versus deep words (63.75%), $t(23) = 1.49, p = .15$; deep words versus control words (63.00%), $t(23) = .29, p = .77$; shallow words versus control words, $t(23) = 1.17, p = .25$.

Recall of the 1-point words is summarized in Figure 6. The three means differed to a statistically significant extent, $F(2, 46) = 9.35, p = .0004$. Pairwise comparisons indicate a strong levels-of-processing effect, with recall in the deep condition (42.50%) being higher than that in the shallow condition (31.00%), the difference was significant, $t(23) = 4.11, p = .0004$. Recall was higher in the control condition (36.75%) than in the
shallow condition, $t(23) = 2.53, p = .02$. The difference between recall in the deep and control conditions was of equivocal significance, $t(23) = 2.00, p = .06$.

*Figure 5.* Mean recall of 10-point words (Experiment 2)
Although not of direct relevance to present purposes, the finding of a congruency effect (Bryant, 1990) should be noted. That is, words requiring a "yes" response were more likely to be recalled than words requiring a "no" response (51.38% versus 44.75%). The difference was statistically significant, $t(23) = 5.51, p < .0001$.

**Discussion**

Of primary interest, selectivity did not differ between the control condition, in which subjects were free to direct their own processing, and the orienting task condition (i.e., the combined deep and shallow conditions), in which it was hypothesized that they were not. Previous research has consistently suggested that orienting tasks firmly control cognitive processing, but the present findings demonstrate that the control is not without limits. On average orienting tasks failed, not only to eliminate, but even to attenuate selectivity.

Just as the imposition of an orienting task did not, overall, impair selectivity, so the selectivity procedure did not prevent a levels-of-processing effect. To be sure, the effect was smaller than typical, but recall was significantly more probable in the deep condition than in the shallow condition. The levels-of-processing effect was, at least largely, confined to the 1-point words. This finding is not surprising, given that 1-point words carried little incentive for processing beyond that required by the orienting task. The 10-point words carried a much stronger incentive for beyond-task processing, which
presumably would tend to mask the effect of orienting task. Hence, selectivity was
greater following a shallow orienting task than a deep orienting task.

**Experiment 3**

A potential drawback of Experiment 2 is that the 10-point and 1-point words were
distinguished on a semantic basis. According to a strict interpretation of the levels-of-
processing framework, this detail should eliminate the effect, because all words,
including those in the shallow condition, should have benefited from the deep processing
used to determine their category. This argument is, perhaps, more theoretical than
practical, since the lifelong habit of semantic processing will manifest in awareness of
meaning in any, even nominally shallow, orienting task (Postman & Kruesi, 1977). More
importantly, even if the semantic determination of value had precluded a levels-of-
processing effect, the principal conclusion of the experiment, namely that the imposition
of an orienting task does not preclude selectivity, would still hold.

Notwithstanding these arguments, it was considered worth investigating whether
the surprising findings of Experiment 2 would generalize to a procedure in which value
was signaled without recourse to the meaning of the target words. Thus, in Experiment 3,
the target words were homogenous as opposed to being drawn from distinct semantic
categories, and their values were signaled directly by an appended numeral.
Method

The method was, in most respects, identical to that of Experiment 2. The only differences were that the stimuli were drawn from a non-categorized word pool and each word's points value was displayed alongside the word. Thus, the words did not have to be categorized to ascertain their values.

Subjects. Forty-eight Rice University undergraduates participated in return for course credit.

Materials. Some 264 words were selected from the Toronto word pool and allocated at random to 11 24-word lists, the first of which served as practice. Of the 24 words in each list, 8 were assigned to the deep condition, 8 to the shallow condition, and 8 to the control condition. Within each of these conditions, 4 words were assigned a value of 10 points, and 4 words were assigned a value of 1 point. Details regarding presentation of target words, the nature of the orienting tasks, and the design of the response and recall sheets were as in Experiment 2.

Design. The experiment conformed to a 3 (deep processing/shallow processing/control) x 2 (10 points/1 point) x 2 (yes/no response) within-subjects design. Both the lists and the words within each list were shown in the same random order to all subjects. Counterbalancing measures ensured that, across subjects, each target word occurred in each encoding (deep, shallow, and control), response (yes and no), and value (10-points and 1-point) condition equally often.
Procedure. The procedural details were as in Experiment 2 except that values were shown directly alongside of, and in the same font as, their respective words. Thus, whereas the subjects in Experiment 2 saw, for example, “1. PEACOCK”, those in Experiment 3 saw “1. GALLANT 10”.

Results

Selectivity and word recall were computed for each of the three encoding conditions: deep, shallow, and control. For each subject, selectivity indices were computed for each list and averaged across lists within each of the three encoding conditions according to Equation 3. Table 4 shows these means, averaged over subjects, along with the corresponding mean percentages of words recalled.

<table>
<thead>
<tr>
<th></th>
<th>Deep</th>
<th>Shallow</th>
<th>Control</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI</td>
<td>.51</td>
<td>.64</td>
<td>.65</td>
<td>.60</td>
</tr>
<tr>
<td>% recall</td>
<td>34.63</td>
<td>28.38</td>
<td>30.75</td>
<td>31.25</td>
</tr>
</tbody>
</table>

Selectivity. The mean selectivity index exceeded zero to a statistically significant extent both overall and in each condition: overall, $t(47) = 15.77, p < .0001$; deep, $t(47) = 10.62, p < .0001$; shallow, $t(47) = 15.90, p < .0001$; control, $t(47) = 15.16, p < .0001$. 
As shown in Figure 7, selectivity differed significantly among the three encoding conditions, $F(1,91, 89.80) = 7.88, p = .0008$. The index for a combined shallow and deep orienting task condition (.58) differed from that in the control condition to a marginally significant extent, $t(47) = 2.14, p = .04$. Between the shallow and deep conditions it differed significantly, $t(47) = 4.09, p = .0002$, being greater in the shallow condition. It also differed significantly between the control and deep conditions, $t(47) = 2.76, p = .008$, being greater in the control condition. It did not differ between the shallow and control conditions, $t(47) = 0.64, p = .52$.

Figure 7. Selectivity indices for control, deep, and shallow conditions (Experiment 3)
Recall. Figure 8 shows the mean levels and variability of recall separately for each of the three encoding conditions. The difference among the means was statistically significant, $F(1.97, 92.36) = 14.50, p < .0001$. More words were recalled in the deep condition than in either the shallow condition, $t(47) = 5.24, p = < .0001$, or the control condition, $t(47) = 3.11, p = .003$, and more were recalled in the control condition than in the shallow condition, $t(47) = 2.70, p = .01$.

![Box plot showing recall in control, deep, and shallow conditions](image)

*Figure 8. Recall in control, deep, and shallow conditions (Experiment 3)*

A separate analysis of the 10-point words revealed a significant difference among encoding conditions, $F(2, 94) = 4.56, p = .01$. As shown in Figure 9, more words were recalled in the deep condition (50.88%) than in the shallow condition (45.95%), $t(47) = 3.23, p = .002$; neither pairwise comparison involving the control condition (48.70%) was
statistically significant: deep versus control, $t(47) = 1.31, p = .20$; shallow versus control, $t(47) = 1.61, p = .11$.

![Box plots showing percentage recall for control, deep, and shallow conditions](image)

*Figure 9. Recall of 10-point words in control, deep, and shallow conditions (Experiment 3)*

Figure 10 summarizes the recall of the 1-point words. The three means differed to a statistically significant extent, $F(1.76, 82.79) = 11.32, p < .0001$. More words were recalled in the deep condition (17.93%) than in either the shallow condition (11.40%), $t(47) = 4.21, p = .0001$, or the control condition (12.65%), $t(47) = 3.21, p = .002$. The difference between the shallow and control conditions was not statistically significant, $t(47) = 1.12, p = .27$. 
Figure 10. Recall of 1-point words in control, deep, and shallow conditions (Experiment 3)

Finally, recall was somewhat more likely for words requiring a “yes” response (33.99%) than for words requiring a “no” response (29.09%). The difference was significant, t(47) = 4.42, p < .0001.

Discussion

Recall was selective regardless of whether an orienting task was imposed. The conclusion drawn from Experiment 2 that orienting tasks do not invariably dominate the encoding process is thus reinforced.

Given that, on the one hand, selectivity survived the imposition of an orienting task and that, on the other hand, the orienting tasks did produce a levels-of-processing effect, it might be expected that recall in the shallow condition would be more selective
than recall in the deep condition, for a deep processing task is likely to boost recall more for 1-point words than for 10-point words. This prediction was borne out. Similarly, the levels-of-processing effect might be expected to be more pronounced for the 1-point words than for the 10-point words, for the higher value should encourage effective processing of all words, regardless of task. This prediction was also borne out.

It should be noted that, overall, level of recall was appreciably lower than in Experiment 2 (31% versus 50%). This discrepancy is presumably a consequence of switching to uncategorized lists. In Experiment 2, the inclusion of categories within a list (e.g., *birds* and *fruits*) presumably reduced the problem of cue overload (Watkins, 1979). Also, the vocabularies from which these items are drawn are much smaller than for the uncategorized lists of the present experiment, making recall easier (Crannell & Parrish, 1957).

**Experiment 4**

That the imposition of the orienting task did not eliminate selectivity is contrary to the notion that orienting tasks exert total control over the encoding process; that they had little (Experiment 3) or no (Experiment 2) effect on selectivity is contrary to all expectation. Could it be that something of the processing induced by orienting tasks carried over to the control condition? If so, the within-subjects design of Experiments 2 and 3 could have failed to capture the real effect of orienting tasks on selectivity. This possibility was addressed in Experiment 4. Specifically, in addition to the three
conditions of the previous two experiments, a separate group of subjects was tested in a
selectivity-only condition, in which no orienting tasks were given.

It is also possible that the selectivity task attenuated the levels-of-processing
effect. The levels-of-processing effect is often larger than found in Experiments 2 and 3.
For example, in their seminal study, Hyde and Jenkins (1969) found an effect size, in
terms of Cohen’s d, of 1.70. Although statistically significant in Experiment 2 and
Experiment 3, the levels-of-processing effect was by comparison rather muted, with d =
.71 in both experiments. Given that selectivity did in fact survive the imposition of
orienting tasks, it is plausible to assume that it masked, or at least attenuated, the relative
effects of the shallow and deep orienting tasks in the high value (10-point) condition.
This possibility was also addressed in Experiment 4. Specifically, a levels-of-processing-
only group was added to the design.

In all, then, Experiment 4 included three groups of subjects: an orienting-task-
plus-selectivity (LOP&SEL) group, just as in Experiments 2 and 3; a selectivity-only
(SEL) group; and an orienting-task-only (LOP) group.

Method

Subjects. A total of 72 Rice University undergraduates participated in return for
course credit. They were randomly assigned to one of the three groups: LOP&SEL,
LOP, or SEL.
Materials and design. Some 264 words were selected at random from the Toronto word pool and randomly assigned to 11 24-word lists. These lists and their constituent words were presented in the same order to all three groups. The first list served as practice. For any given subject in the LOP and LOP&SEL groups, 8 words in each list were assigned to the deep condition, 8 to the shallow condition, and 8 to the control condition; for each of these 8-word sets, 4 words required a “yes” response and 4 a “no” response. For any given subject in the LOP&SEL and SEL groups, half of the words of each list were designated as worth 10 points and half as worth 1 point; for the LOP&SEL group, there were two of each value within each combination of the encoding and response conditions. For the LOP group, all words were assigned a 10-point value for half of the subjects and a 1-point value for the other half.

Counterbalancing measures ensured that, across subjects within each applicable group, each word was assigned equally often to each encoding condition (deep, shallow, and control), response condition (“yes” or “no”), and value (10-points or 1-point). As in Experiments 2 and 3, values were shown to the right of the words.

For the LOP&SEL and LOP groups, the orienting task response sheets and recall sheets were as in Experiments 2 and 3. For the SEL group, the orienting task response sheet was modified, since there were no orienting questions. Specifically, subjects checked a Viewed column instead of Yes and No columns.
Broadly, the experiment can be conceptualized as conforming to a 3 (deep processing/shallow processing/control) x 2 (10 points/1 point) x 2 (yes/no response) x 3 (LOP/SEL/LOP&SEL groups) factorial design, with repeated measures on the first three factors. Note, however, that the levels of some variables were sometimes collapsed into a dummy variable. Thus, no encoding (deep/shallow/control) or response ("yes" or "no") variables were apparent to the SEL group; there was no response variable for the control encoding condition for the LOP and LOP&SEL groups; and there was no values variable for the LOP group.

Procedure. For the LOP&SEL group the procedure was as in Experiment 3. For the LOP group the procedure differed only in that the instruction called for recall of as many words as possible, rather than the highest possible score. Words were presented without points values, and subjects completed the orienting task just as did those in the LOP&SEL group. For the SEL group, no orienting tasks were given. Words were presented along with their points values, and subjects checked the serial number of each word in the Viewed column of their response sheet.

After presentation of each list, subjects attempted to recall the 8 digits of the distractor task, and then recalled the target words. After 90 seconds, they were shown the target words to facilitate calculation of their recall scores. For the LOP&SEL and SEL groups, the scoring procedure was exactly as in Experiment 3. For the LOP group, the target words were shown in alphabetized order and half of the subjects were reminded
that each word was worth 10 points and the other half were reminded that each word was worth 1 point.

**Results**

The principle selectivity and recall findings are summarized in Table 5.

*Table 5: Mean Selectivity Indices (SI) and Percentage of Words Recalled, as appropriate, for the orienting-tasks-plus-selectivity (LOP&SEL) group, the orienting-tasks-only (LOP) group, and the selectivity-only (SEL) group (Experiment 4).*

<table>
<thead>
<tr>
<th></th>
<th>LOP&amp;SEL</th>
<th></th>
<th></th>
<th></th>
<th>LOP</th>
<th></th>
<th></th>
<th></th>
<th>SEL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deep</td>
<td>Shallow</td>
<td>Control</td>
<td>Overall</td>
<td>Deep</td>
<td>Shallow</td>
<td>Control</td>
<td>Overall</td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>.57</td>
<td>.70</td>
<td>.67</td>
<td>.65</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.57</td>
</tr>
<tr>
<td>% recall</td>
<td>38.54</td>
<td>32.55</td>
<td>34.13</td>
<td>35.10</td>
<td>34.38</td>
<td>28.18</td>
<td>30.99</td>
<td>31.18</td>
<td>39.50</td>
</tr>
</tbody>
</table>

*Selectivity.* The mean selectivity index exceeded zero to a statistically significant extent both overall and in all relevant conditions. For the LOP&SEL group: overall, \( t(23) = 13.29, p < .0001 \); deep, \( t(23) = 10.50, p < .0001 \); shallow, \( t(23) = 11.13, p < .0001 \); control, \( t(23) = 13.37, p < .0001 \). For the SEL group: \( t(23) = 11.01, p < .0001 \).

In the LOP&SEL group, selectivity differed significantly among the three encoding conditions: \( F(1.91, 43.88) = 3.94, p = .03 \). More specifically, it differed between the shallow and deep conditions: \( t(23) = 2.44, p = .02 \), being greater in the
shallow condition (see Figure 11). Selectivity did not differ significantly between the shallow and control conditions, $t(23) = .86, p = .40$, or between the deep and control conditions, $t(23) = 1.89, p = .07$.

Selectivity did not differ between the SEL group and the control condition of the LOP&SEL group, $t(46) = 1.33, p = .19$. In addition, there was no significant difference between overall selectivity in the LOP&SEL group and the SEL group, $t(46) = 1.05, p = .30$; this finding is reinforced by a selectivity effect size analysis: for the LOP&SEL group, Cohen’s d = 2.04; for the SEL group, Cohen’s d = 1.91. These effect sizes did not differ significantly, $t(23) = 0.03, p = .98$.

*Figure 11.* Selectivity indices for the control, deep, and shallow conditions of the orienting-tasks-plus selectivity (LOP&SEL) group and for the selectivity-only (SEL) group (Experiment 4).
Recall. Over all subjects, the average rate of recall was 35.20%. As is clear from Table 5, the rate differed among the three groups, $F(2, 69) = 4.51, p = .01$. The rate differed significantly between the SEL and LOP groups, $t(46) = 3.61, p = .0007$. The other two pairwise comparisons were not statistically significant: for LOP&SEL versus LOP, $t(46) = 1.29, p = .20$; and for LOP&SEL versus SEL, $t(46) = 1.51, p = .14$.

In the LOP&SEL group, recall differed significantly among encoding conditions, $F(2, 46) = 7.03, p = .002$. More specifically, recall differed between the deep and shallow conditions, $t(23) = 3.23, p = .004$, being greater in the deep condition. The size of this levels-of-processing effect, as indexed by Cohen’s d, was .66. Recall also differed significantly between the deep and the control conditions, $t(23) = 2.71, p = .01$, being greater in the deep condition. Recall did not differ significantly between the shallow and control conditions, $t(23) = 1.04, p = .31$ (see Figure 12).

In the LOP group, recall differed significantly among encoding conditions: $F(2, 46) = 5.49, p = .007$. It differed significantly between the deep and shallow conditions, $t(23) = 3.19, p = .004$, being greater in the deep condition, with an effect size of $d = .65$. There was no significant difference in recall between the deep and the control conditions, $t(23) = 1.87, p = .07$, or between the shallow and the control conditions, $t(23) = 1.51, p = .14$. Also, there was no significant difference in recall between the subjects who were told that all words were worth 10 points each and those who were told that all words were worth 1 point each (31.04% versus 31.32%), $t(22) = 0.08, p = .94$. This result held up for
each of the encoding conditions: for the deep condition (35.52% versus 33.23%), \( t(22) = 0.71, p = .49 \); for the shallow condition (26.35% versus 30.00%), \( t(22) = 0.86, p = .40 \); for the control condition (31.25% versus 30.73%), \( t(22) = 0.12, p = .91 \).

\[\text{Figure 12. Percentage of words recalled in each condition for the orienting-tasks-plus-selectivity (LOP&SEL) group, the orienting-tasks-only (LOP) group, and the selectivity-only (SEL) group (Experiment 4).}\]

For the LOP&SEL group, recall was examined separately for the 10-point and 1-point words. For the 10-point words, there was no significant main effect of encoding condition: \( F(2, 46) = 1.12, p = .34 \) (see Figure 13), but there was for the 1-point words, \( F(1.65, 38.96) = 9.93, p = .0007 \) (see Figure 14). Pairwise comparisons for the 1-point words showed that the levels-of-processing effect (19.90% for the deep condition and 11.35% for the shallow condition) was statistically significant, \( t(23) = 3.71, p = .001 \).
The intermediate level of recall for the control condition (14.27%) differed significantly from both that for the deep condition, \( t(23) = 2.74, p = .01 \), and that for the shallow condition, \( t(23) = 2.13, p = .04 \).

![Box plot showing percentage recall across conditions](image)

*Figure 13. Percentage of 10-point words recalled in control, deep, and shallow conditions (LOP&SEL, Experiment 4).*

For the SEL subjects, who were not given orienting tasks, recall of the 10-point and 1-point words differed sharply (59.10% versus 19.72%), \( t(23) = 9.66, p < .0001 \).

Finally, a congruency effect occurred for both the LOP&SEL and LOP groups. For the LOP&SEL group, the recall rates for words requiring a "yes" and "no" response were 37.03% and 34.06% respectively, \( t(23) = 2.38, p = .04 \); and for the LOP group, they were 33.44% and 29.11% respectively, \( t(23) = 2.45, p = .02 \).
Figure 14. Percentage of 1-point words recalled in control, deep, and shallow conditions (LOP&SEL, Experiment 4)

Discussion

The principal finding of Experiments 2 and 3, namely that recall was selective even for words studied in the context of an orienting task, was replicated. Moreover, selectivity was no less pronounced in the orienting-task conditions of the LOP&SEL group than in, not only the no-orienting-task control condition of the LOP&SEL group, but also the SEL group, which never engaged in orienting tasks. Also replicated from the previous two experiments was the finding that recall was more selective in the shallow condition than in the deep condition.

Just as, overall, imposition of an orienting task did not diminish selectivity, so too the imposition of a selectivity strategy did not diminish the levels-of-processing effect.
Thus, the recall advantage of the deep orienting task over the shallow orienting task was virtually identical for the LOP and LOP&SEL groups (see Table 5). On the other hand, for the LOP&SEL group, the levels-of-processing effect was substantially, if not wholly, confined to the 1-point words.

**General Discussion**

The purpose of the research reported here was to explore the generality of one of the most puzzling phenomena in the memory research literature: why the imposition of a “shallow” orienting task during study of a set of target items impairs subsequent recall even under study conditions that could reasonably be presumed to induce additional “deeper” processing. Thus, for example, items subjected to a shallow orienting task are poorly recalled: even when a recall test is fully expected; even when ample study time is allowed; even when an incentive is offered; and even when subjects are lectured on levels of processing and encouraged to process the items deeply. As Roediger and Gallo (2001) have observed, it is as though orienting tasks preclude additional processing.

The experiments reported here were designed to test this conclusion by determining whether orienting tasks precluded the ability to modulate memory according to incentives. To this end, a “selectivity” procedure, described by Watkins and Bloom (1999), was adopted. Words of a study list were arbitrarily assigned different values, specifically either 1 or 10 points, and the effect on recall was measured in terms of a selectivity index.
Experiment 1 demonstrated that the temporal locus of the selectivity normally observed with this procedure was dominantly at the encoding rather than the recall phase of the remembering process, thereby confirming the selectivity procedure as appropriate for addressing a phenomenon known to have an encoding locus. The remaining three experiments failed to preclude selective recall. Indeed, the effect of incentives was far stronger than the effect of orienting tasks, and moreover was little, if at all, diminished by the imposition of an orienting task.

Why should the incentives of the selectivity procedure remain immune to the imposition of an orienting task when other encoding manipulations are rendered impotent? A potential, if ad hoc, partial answer to this question can be derived from the notion that subjects tend to focus on the task at hand. Given an orienting task, the subject’s immediate task is to respond to a question: Is the word in uppercase? Does it make sense to insert the word into this sentence frame? Knowing that the words will have to be recalled does not change the task at hand, namely to respond to the orienting questions. In the selectivity procedure the values are prominent in the study experience, and so set up a task at hand, namely selective remembering. Other procedures designed to break the apparent processing block induced by the orienting task, such as giving extra study time or a lecture on levels of processing, do not change the status of the orienting questions as the task at hand.
One limitation of this account is that the "task at hand" is not always clear. Consider the only two findings, other than those reported herein, of the failure of an orienting task to block additional processing. In the first, additional effective processing after a shallow orienting task occurred in response to a second, deep orienting task (Chow et al., 1978). Here, the task at hand is clear: to respond to the second (deep) orienting question. But in the second finding, the task at hand is not so clear: subjects were lectured on the levels-of-processing construct and instructed to generate, though not report, an adjective appropriate to each target noun, in addition to overtly completing the standard orienting task. Here, it is not clear whether the adjective-generating instruction constitutes a task at hand. Thus, on the one hand, it is well defined, item-specific, and part of the study procedure; but on the other hand, even if it were carried out, the generation of the adjectives would be entirely covert and thus of equivocal task status. As it happens, the release from the normal blocking effect of orienting tasks was also equivocal, for it occurred with a 12-second, but not with a 6-second, presentation rate, and even then only in a recall test and not a recognition test (Chow et al., 1978).

Given that selectivity prevailed in the face of an orienting task, it is perhaps not surprising that with the high value items the levels-of-processing effect was attenuated, or even eliminated. Apparently, items were effectively processed either in response to a deep orienting question or under high incentive. When neither occurred, recall was
especially poor, and hence the levels-of-processing effect was strongest with, if not
confined to, the low-value items.

The present demonstration of a clear limitation of the blocking effect of the
orienting tasks does not, of course, actually resolve the puzzle of why the effect occurs
when it does. For that, more research is required. It might prove instructive, for
example, to explore the importance of making an overt response to the orienting tasks.
Thus, subjects could be encouraged to covertly execute an orienting task under the threat
that they may be asked to give a speeded response. Recall could then be compared for
items given overt and covert responses. But the interpretation of whatever findings such
future research might yield should be constrained by the present demonstration that the
apparent blocking effect of an orienting task can be overcome.
References


Appendix A

Please note that 10 words from each category were used in Experiment 1, whereas 12 words from each category were used in Experiment 2.

**Practice Lists**

<table>
<thead>
<tr>
<th>Category: Ships</th>
<th>Category: Diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATTLESHIP</td>
<td>ARTHRITIS</td>
</tr>
<tr>
<td>DESTROYER</td>
<td>BERIBERI</td>
</tr>
<tr>
<td>DINGHY</td>
<td>CANCER</td>
</tr>
<tr>
<td>FERRY</td>
<td>CHOLERA</td>
</tr>
<tr>
<td>FREIGHTER</td>
<td>DIPHTHERIA</td>
</tr>
<tr>
<td>HYDROFOIL</td>
<td>JAUNDICE</td>
</tr>
<tr>
<td>ROWBOAT</td>
<td>LEUKEMIA</td>
</tr>
<tr>
<td>RUNABOUT</td>
<td>MALARIA</td>
</tr>
<tr>
<td>MOTORBOAT</td>
<td>RABIES</td>
</tr>
<tr>
<td>SLOOP</td>
<td>RICKETS</td>
</tr>
<tr>
<td>SPEEDBOAT</td>
<td>SCURVY</td>
</tr>
<tr>
<td>SUBMARINE</td>
<td>TYPHOID</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category: Men’s names</th>
<th>Category: Crimes</th>
</tr>
</thead>
<tbody>
<tr>
<td>BILL</td>
<td>ADULTERY</td>
</tr>
<tr>
<td>CHARLES</td>
<td>ARSON</td>
</tr>
<tr>
<td>DAVE</td>
<td>BLACKMAIL</td>
</tr>
<tr>
<td>DON</td>
<td>EMBEZZLING</td>
</tr>
<tr>
<td>ELMER</td>
<td>FRAUD</td>
</tr>
<tr>
<td>GEORGE</td>
<td>KIDNAPPING</td>
</tr>
<tr>
<td>HARRY</td>
<td>LARCENY</td>
</tr>
<tr>
<td>HENRY</td>
<td>LIBEL</td>
</tr>
<tr>
<td>JOHN</td>
<td>PERJURY</td>
</tr>
<tr>
<td>MIKE</td>
<td>SLANDER</td>
</tr>
<tr>
<td>PAUL</td>
<td>THEFT</td>
</tr>
<tr>
<td>PETE</td>
<td>TREASON</td>
</tr>
</tbody>
</table>
Experimental Lists

Category: Fish
BLUEGILL
DOLPHIN
FLOUNDER
GUPPY
HALIBUT
HERRING
MACKEREL
MINNOW
PIRANHA
SALMON
STURGEON
WALLEYE

Category: Occupations
ACCOUNTANT
ARCHITECT
BANKER
CHEMIST
ENGINEER
MECHANIC
PHARMACIST
PROFESSOR
SECRETARY
TEACHER
UNDERWRITER
WRITER

Category: Sports
BASEBALL
BOWLING
BOXING
CURLING
FENCING
FOOTBALL
GOLF
JUDO
RACING
SAILING
SNORKELING
SURFING

Category: Trees
BALSAM
BEECH
CEDAR
CHESTNUT
EUCALYPTUS
EVERGREEN
GINKO
HICKORY
REDWOOD
SPRUCE
SYCAMORE
WILLOW
<table>
<thead>
<tr>
<th>Category: Topography</th>
<th>Category: Insects</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAY</td>
<td>APHID</td>
</tr>
<tr>
<td>CAVE</td>
<td>BEETLE</td>
</tr>
<tr>
<td>CRATER</td>
<td>BUTTERFLY</td>
</tr>
<tr>
<td>ISLAND</td>
<td>CUTWORM</td>
</tr>
<tr>
<td>LAKE</td>
<td>HORNET</td>
</tr>
<tr>
<td>PLAIN</td>
<td>HORSEFLY</td>
</tr>
<tr>
<td>PLATEAU</td>
<td>MOTH</td>
</tr>
<tr>
<td>RIDGE</td>
<td>SCORPION</td>
</tr>
<tr>
<td>ROCK</td>
<td>SLUG</td>
</tr>
<tr>
<td>STREAM</td>
<td>TARANTULA</td>
</tr>
<tr>
<td>TUNDRA</td>
<td>WASP</td>
</tr>
<tr>
<td>VOLCANO</td>
<td>WATERBUG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category: House parts</th>
<th>Category: Seasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALCOVE</td>
<td>ALLSPICE</td>
</tr>
<tr>
<td>BATHROOM</td>
<td>ALMOND</td>
</tr>
<tr>
<td>BATHTUB</td>
<td>ANISE</td>
</tr>
<tr>
<td>BEDROOM</td>
<td>BASIL</td>
</tr>
<tr>
<td>CEILING</td>
<td>CARDAMON</td>
</tr>
<tr>
<td>CELLAR</td>
<td>CLOVES</td>
</tr>
<tr>
<td>CUPBOARD</td>
<td>KETCHUP</td>
</tr>
<tr>
<td>DEN</td>
<td>MARJORAM</td>
</tr>
<tr>
<td>FOUNDATION</td>
<td>PAPRIKA</td>
</tr>
<tr>
<td>HALL</td>
<td>PEPPERMINT</td>
</tr>
<tr>
<td>LIVINGROOM</td>
<td>SALT</td>
</tr>
<tr>
<td>ROOF</td>
<td>SAVOURY</td>
</tr>
<tr>
<td>Category: U.S. States</td>
<td>Category: Clothing</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>ARKANSAS</td>
<td>COAT</td>
</tr>
<tr>
<td>CALIFORNIA</td>
<td>GLOVES</td>
</tr>
<tr>
<td>CONNECTICUT</td>
<td>HAT</td>
</tr>
<tr>
<td>DELAWARE</td>
<td>PAJAMAS</td>
</tr>
<tr>
<td>FLORIDA</td>
<td>ROBE</td>
</tr>
<tr>
<td>HAWAII</td>
<td>SCARF</td>
</tr>
<tr>
<td>INDIANA</td>
<td>SHIRT</td>
</tr>
<tr>
<td>IOWA</td>
<td>SHORTS</td>
</tr>
<tr>
<td>MICHIGAN</td>
<td>SKIRT</td>
</tr>
<tr>
<td>MONTANA</td>
<td>SLIP</td>
</tr>
<tr>
<td>OKLAHOMA</td>
<td>SUIT</td>
</tr>
<tr>
<td>OREGON</td>
<td>SWEATER</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category: Body parts</th>
<th>Category: Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMS</td>
<td>BRAZIL</td>
</tr>
<tr>
<td>CHEST</td>
<td>CUBA</td>
</tr>
<tr>
<td>HAND</td>
<td>DENMARK</td>
</tr>
<tr>
<td>HIP</td>
<td>EGYPT</td>
</tr>
<tr>
<td>KNEE</td>
<td>FINLAND</td>
</tr>
<tr>
<td>NOSE</td>
<td>ISRAEL</td>
</tr>
<tr>
<td>SHOULDER</td>
<td>POLAND</td>
</tr>
<tr>
<td>STOMACH</td>
<td>PORTUGAL</td>
</tr>
<tr>
<td>THROAT</td>
<td>SCOTLAND</td>
</tr>
<tr>
<td>TOE</td>
<td>SPAIN</td>
</tr>
<tr>
<td>TOOTH</td>
<td>SWEDEN</td>
</tr>
<tr>
<td>TRUNK</td>
<td>VENEZUELA</td>
</tr>
</tbody>
</table>
Category: Sciences
AGRICULTURE
ANATOMY
ENGINEERING
ENTOMOLOGY
GENETICS
GEOGRAPHY
MEDICINE
METEOROLOGY
OCEANOGRAPHY
PHILOSOPHY
SOCIOLOGY
ZOOLOGY

Category: Vegetables
BEET
CABBAGE
CARROT
CAULIFLOWER
ONION
PEA
PUMPKIN
RHUBARB
SPINACH
STRINGBEAN
YAMS
ZUCCHINI

Category: Automobiles
BUICK
CADILLAC
CORVETTE
FIAT
HONDA
PONTIAC
PORSCHE
RAMBLER
SAAB
THUNDERBIRD
VOLKSWAGEN
VOLVO

Category: Clergy
BROTHER
CANON
CLERIC
DEACON
MINISTER
MISSIONARY
MONK
NUN
PADRE
POPE
PREACHER
RECTOR
<table>
<thead>
<tr>
<th>Category: Elements</th>
<th>Category: Birds</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALUMINUM</td>
<td>BLUEJAY</td>
</tr>
<tr>
<td>ARGON</td>
<td>DOVE</td>
</tr>
<tr>
<td>BORON</td>
<td>DUCK</td>
</tr>
<tr>
<td>BROMINE</td>
<td>EAGLE</td>
</tr>
<tr>
<td>CARBON</td>
<td>ORIOLE</td>
</tr>
<tr>
<td>HELIUM</td>
<td>OSTRICH</td>
</tr>
<tr>
<td>HYDROGEN</td>
<td>OWL</td>
</tr>
<tr>
<td>LEAD</td>
<td>PARAKEET</td>
</tr>
<tr>
<td>MAGNESIUM</td>
<td>PEACOCK</td>
</tr>
<tr>
<td>NITROGEN</td>
<td>PENGUIN</td>
</tr>
<tr>
<td>OXYGEN</td>
<td>ROBIN</td>
</tr>
<tr>
<td>TIN</td>
<td>THRUSH</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Category: Musical instruments</td>
<td>Category: Fruits</td>
</tr>
<tr>
<td>CELLO</td>
<td>BANANA</td>
</tr>
<tr>
<td>CLARINET</td>
<td>DATE</td>
</tr>
<tr>
<td>CYMBALS</td>
<td>FIG</td>
</tr>
<tr>
<td>DRUM</td>
<td>GRAPe</td>
</tr>
<tr>
<td>FLUTE</td>
<td>GRAPEFRUIT</td>
</tr>
<tr>
<td>GUITAR</td>
<td>LIME</td>
</tr>
<tr>
<td>LUTE</td>
<td>PAPAYA</td>
</tr>
<tr>
<td>MANDOLIN</td>
<td>POMEGRANATE</td>
</tr>
<tr>
<td>TAMBOURINE</td>
<td>PRUNES</td>
</tr>
<tr>
<td>TROMBONE</td>
<td>RASPBERRY</td>
</tr>
<tr>
<td>TRUMPET</td>
<td>TANGERINE</td>
</tr>
<tr>
<td>VIOLIN</td>
<td>WATERMELON</td>
</tr>
</tbody>
</table>
### Category: Animals
- ANTELOPE
- COW
- DEER
- DOG
- HIPPOPOTAMUS
- LION
- MONKEY
- MOUSE
- PANTHER
- RACCOON
- SHEEP
- SQUIRREL

### Category: Tools
- CROWBAR
- HACKSAW
- HAMMER
- JIGSAW
- LATHE
- NAILSET
- PLANE
- RIPSAW
- RULER
- SAW
- VISE
- WEDGE

### Category: Colors
- AUBURN
- AZURE
- BURGUNDY
- CHARTREUSE
- EBONY
- FUCHSIA
- INDIGO
- MAGENTA
- MAROON
- RED
- RUSSET
- TURQUOISE

### Category: Weather
- BREEZY
- DEW
- DRIZZLE
- DUSTSTORM
- FROST
- HAIL
- HUMIDITY
- LIGHTNING
- SLEET
- SMOG
- STORM
- TYPHOON
<table>
<thead>
<tr>
<th>Category: Flowers</th>
<th>Category: Cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZALIA</td>
<td>ATLANTA</td>
</tr>
<tr>
<td>CROCUS</td>
<td>BOMBAY</td>
</tr>
<tr>
<td>DANDELION</td>
<td>BOSTON</td>
</tr>
<tr>
<td>GARDENIA</td>
<td>HELSINKI</td>
</tr>
<tr>
<td>GLADIOLA</td>
<td>MEMPHIS</td>
</tr>
<tr>
<td>HYACINTH</td>
<td>MOSCOW</td>
</tr>
<tr>
<td>JONQUIL</td>
<td>PARIS</td>
</tr>
<tr>
<td>MARIGOLD</td>
<td>RENO</td>
</tr>
<tr>
<td>NARCISSUS</td>
<td>SHANGHAI</td>
</tr>
<tr>
<td>PETUNIA</td>
<td>STOCKHOLM</td>
</tr>
<tr>
<td>POINSETTIA</td>
<td>TOKYO</td>
</tr>
<tr>
<td>SUNFLOWER</td>
<td>VENICE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category: Textiles</th>
<th>Category: Kitchenware</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACRILON</td>
<td>BEATER</td>
</tr>
<tr>
<td>CHIFFON</td>
<td>BOWL</td>
</tr>
<tr>
<td>COTTON</td>
<td>BROILER</td>
</tr>
<tr>
<td>CREPE</td>
<td>FRYPAN</td>
</tr>
<tr>
<td>FELT</td>
<td>LADLE</td>
</tr>
<tr>
<td>FLANNEL</td>
<td>PAN</td>
</tr>
<tr>
<td>GINGHAM</td>
<td>POT</td>
</tr>
<tr>
<td>LACE</td>
<td>REFRIDGERATOR</td>
</tr>
<tr>
<td>MUSLIN</td>
<td>SAUCER</td>
</tr>
<tr>
<td>RAYON</td>
<td>SPATULA</td>
</tr>
<tr>
<td>SATIN</td>
<td>SPOON</td>
</tr>
<tr>
<td>VELVET</td>
<td>STRAINER</td>
</tr>
</tbody>
</table>
Appendix B

All boxplots were generated by JMP IN 4.0.2. The annotated figure below describes the components of a boxplot. Data points were jittered to illustrate frequency.

Upper Adjacent Value:
Largest value below upper inner fence (1.5 x H-Spread beyond 75th percentile).

Data Point

75th Percentile

Median

Mean

Lower Adjacent Value:
Smallest value above lower inner fence (1.5 x H-Spread beyond 25th percentile).

25th Percentile

H-Spread

Note: “H-spread” is defined as the difference between the 25th and 75th percentiles, or as the interquartile range.
Appendix C

Category examples are the same as in Appendix A.

**Practice List**
Categories: Birds & Fruits

**Experimental Lists**
Categories: Weather & Tools
Musical Instruments & Elements
Kitchenware & Cities
Fish & Sports
Occupations & Trees
Countries & Clothing
Flowers & Textiles
Animals & Colors
Seasonings & Insects
Body parts & U.S. States