Semantic effects in lexical access: Evidence from single-word naming

Lee H. Wurm
Wayne State University, Detroit, USA

Douglas A. Vakoch
SETI Institute and University of California, Davis, USA

Joanna Aycock and Robyn R. Childers
Wayne State University, Detroit, USA

Cognitive psychologists have not devoted much attention to semantic and emotional effects early in word recognition, assuming instead that such effects are primarily post-perceptual. Some evidence of such early effects does exist, but it relies exclusively on a less-than-ideal experimental task, the lexical decision task. In the current study, participants heard words over headphones and repeated them into a microphone as quickly as possible (single-word naming). The Danger and Usefulness of word referents were significantly related to naming times, independent of effects such as word length, familiarity, onset characteristics, stress, neighbourhood density, and concreteness. Results are discussed in terms of the adaptive benefit of making quick classifications along these dimensions, and against a backdrop of evidence from several widely divergent areas of research.

Recent research from a variety of areas is leading to the conclusion that many aspects of cognition are influenced by fast, automatic emotional or evaluative processes. Murphy and Zajone (1993), for example, state that a nonconscious mental system makes immediate judgements about whether something is positive or negative (the affect primacy hypothesis). Bargh, Chaiken, Govender, and Pratto (1992) concluded that this early evaluative process is extremely fast, occurring automatically and preconsciously for everyone as a part of normal perception. This notion also fits a model of appraisal recently proposed by Robinson (1998; see also Scherer, 1984; Smith & Lazarus, 1990). In Robinson’s

Correspondence should be addressed to Lee H. Wurm, Department of Psychology, Wayne State University, 71 West Warren Avenue, Detroit, MI 48202, USA; e-mail: lee.wurm@wayne.edu

We would like to thank the Working Group in Human Cognitive Neuroscience at Wayne State University for making valuable comments about an earlier version of this paper.

© 2003 Psychology Press Ltd
http://www.tandf.co.uk/journals/pp/02699931.html DOI:10.1080/02699930244000057
model, the valence and urgency of a stimulus or situation are judged extremely rapidly and pre-attentively. A preattentive “urgency detection” module determines whether a stimulus is personally relevant, whether it is consistent with a person’s goals, and whether it can be effectively dealt with.

Issues such as these (and issues of semantic effects in word recognition more generally) have not been dealt with extensively by cognitive psychologists (for a review, see Balota, Ferraro, & Conner, 1991; see also Moss, McCormick, & Tyler, 1997; Slobiacez, 1994; Strain & Herdman, 1999; Strain, Patterson, & Seidenberg, 1995). Most models of word recognition do not include explicit discussions of semantic considerations, as these factors are generally thought to be relevant only post-perceptually (i.e., after word recognition has taken place). When such effects are discussed, it is often in only general terms, for example, positing that words related to each other in meaning are linked in some way.

The accessing of semantic information is in reality of paramount importance to word recognition. An average speaking rate is about three or four words per second, and speakers can significantly exceed this rate without causing any comprehension difficulty for their listeners, so access of semantic information must proceed quickly. Some theorists are moving toward a more explicit recognition of the importance of these factors. In Marslen-Wilson’s (1987) cohort model, semantic information begins to become available as soon as a word is included in the cohort. Inclusion in the cohort is based on the first 150 ms of the acoustic input, which is almost always before a given word can be uniquely identified by the signal. Other activation and competition-based models [e.g., TRACE (McClelland & Elman, 1986) and the Neighborhood Activation Model (Luce, Pisoni, & Goldinger, 1990)] make similar claims, although these models do not have implemented semantic levels (Moss et al., 1997). However, there has been surprisingly little empirical work on these issues.

Some of the existing work has relied on priming paradigms. Zwitserslhood (1989) reported a series of important experiments in Dutch that used cross-modal priming. She found that auditory presentation of the word fragment [kapt], for example, facilitated recognition for probe words related to both captain and captive. This evidence of multiple activation was found very early (within the first 130 ms of the prime’s duration), and was found to hold even when the sentence context was strongly biased toward one or the other word. However, when the complete word was presented, rather than just the first few phonemes, facilitation was limited to probe words related to the presented prime (see also Marslen-Wilson, 1987).

In a series of visual priming studies, Flores d’Arcais, Schreuder, & Glazemborg (1985; Schreuder, Flores d’Arcais, and Glazemborg, 1984) found that perceptual properties of objects (e.g., red or round) were activated faster than functional properties (e.g., what something is made of, or what it is used for). Schreuder et al. (1984) argued that this is because perceptual properties are more
directly and easily observable and because children seem to pick up on them earlier (but see also Nelson, 1974). Moss et al. (1997) tested this same distinction using the cross-modal priming paradigm, and reached the opposite conclusion. They demonstrated that information about the function and design of man-made objects became available more quickly than information about the objects’ physical form. Whichever type of information becomes available first, studies such as these suggest that one can observe semantic priming at a point well before word recognition can take place (i.e., where more than one candidate is still consistent with the prime input). Furthermore, the exact nature of this information can apparently be fairly specific and detailed.

More recently, Lee, Rayner, and Pollatsek (1999) used a technique called fast priming, in which sentences are silently read while eye-movement-contingent changes are made on a specified target region. This paradigm was used to explore the time-course of semantic activation (among other things) in visual word recognition. They found significant semantic priming with a prime duration of 32 ms, which suggests very rapid access of semantic information in reading. Other common strategies used in reading research include monitoring readers’ eye movements, and recording ERPs measurements from their scalps (e.g., Sereno, Rayner, & Posner, 1998). Starr and Rayner (2001) provide a useful discussion of some recent controversies involving the monitoring of eye movements.

There have been many experimental attempts to find semantic effects without using a priming or context manipulation. Most of these have been unsuccessful, or have shown very small effects (Brown & Watson, 1987; de Groot, 1989). With specific reference to visual language processing, Strain et al. (1995) argue that this lack of success is due to researchers’ failure to properly consider the orthography-to-phonology translation process. In one of the few papers that has demonstrated unprimed semantic effects in reading, Strain et al. (1995; Strain & Herdman, 1999) presented words visually to subjects in a series of naming experiments, and found a naming-time advantage for concrete over abstract words. However, the effect held only for low-frequency exception words. The authors argued that semantic activation occurs very early for all words, but it is only observable when the translation from orthography to phonology is slow or noisy, as in the case of low-frequency exception words. According to Strain and Herdman (1999), this finding fits naturally within the connectionist framework of Seidenberg and McClelland (1989; see also Plaut, McClelland, Seidenberg, & Patterson, 1996). The network includes an orthography-to-semantics-to-phonology pathway that shows itself most readily when the orthography-to-phonology pathway is slowed for some reason. This obviously predicts the observed pattern of results.

There are other semantic variables that have attracted some empirical attention over the past three decades. For example, numerous studies have looked at the effects of the number of meanings a word has (e.g., Rubenstein,
Garfield, & Millikan, 1970; Rubenstein, Lewis, & Rubenstein, 1971) on simple reading tasks like lexical decision. In general, it appears that the more meanings a word has, the faster and more accurate performance will be. It is worth noting, however, that some researchers have arrived at the opposite conclusion (e.g., Rayner & Duffy, 1986; Rayner & Frazier, 1989).

Azuma and van Orden (1997) report the results of two lexical decision experiments in a study looking at the effects of “number of meanings” (see also Jastrzembski, 1981). They found an effect on lexical decision times only when the pseudowords used in the experiment were pseudohomophones (e.g., BEAF). Follow-up analyses led the authors to conclude that simple number of meanings is not in reality the key variable; rather, it’s the relatedness of a word’s different meanings. The authors discuss various difficulties that have prevented the separation of these two effects (number of meanings vs. the relatedness of those meanings) in previous studies, and the generally poor state of consensus about the reliability (or even the direction) of the effects (e.g., Gernsbacher, 1984; Rucek, 1995).

There is further cause for caution in interpreting these results. That Azuma and van Orden (1997) only get the effect in question with a certain kind of pseudoword suggests that strategic factors may be operating. This is a particular concern in the Azuma and van Orden (1997) study because only the lexical decision task was used. Many researchers have argued that the lexical decision task is contaminated by post-perceptual decision processes, and by various kinds of biased or strategic responding (e.g., Balota, 1990; Balota & Chumbley, 1984, 1985; Balota & Lorch, 1986; Chumbley & Balota, 1984; Lorch, Balota, & Stamm, 1986; Neely, 1991; Neely & Keele, 1989). To the extent that this is true, the lexical decision task is of limited use in studies of perceptual processing.

Wurm and Vakoch (1996, 2000; Vakoch & Wurm, 1997) have taken a different approach to the role of semantics in word recognition, attempting to be very specific about what the exact nature of early-acting semantic information would be. They have argued (see also Anderson, 1991) that it would be efficient to code meaning according to a small number of dimensions. In a series of studies they asked whether the very time-course of perceiving a spoken isolated word, without prior presentation of a prime and indeed without any kind of semantic or sentential context, can be determined in part by that word’s “weights” on three common semantic dimensions: Evaluation, Potency, and Activity (Osgood, 1969; Osgood, May, & Miron, 1975; Osgood, Suci, & Tannenbaum, 1957). This was indeed found to be the case (Vakoch & Wurm, 1997; Wurm & Vakoch, 1996).

In a subsequent study Wurm and Vakoch (2000) obtained subject ratings of the Danger and Usefulness of the referents of 100 common nouns. These ratings were found to be significantly related to lexical decision times for those nouns, after controlling for factors such as word frequency, word length, subjective familiarity, concreteness, imageability, animate vs. inanimate referent, word-
onset characteristics, bigram frequency, stress, neighbourhood density, and meaningfulness. Independent of these factors, it was found that nouns with more Dangerous or more Useful referents had faster lexical decision times. There was also an interaction between the two dimensions. The results of all three of these studies (Vakoch & Wurm, 1997; Wurm & Vakoch, 1996, 2000) were interpreted within the framework of adaptiveness. It makes good adaptive sense to give priority to things that are Dangerous, for example. Under the alternative set of dimensions, these objects would tend to be coded as low on Evaluation and high on Potency (i.e., Bad and Strong).

The results of these studies suggest that semantic effects are present very early in auditory word recognition, and that they are more detailed and pervasive than has generally been assumed. However, all of the evidence just discussed has been found using the lexical decision task, an experimental paradigm that is frequently criticised, as we noted above. The most frequent and serious criticism of the task is that it is open to a range of post-perceptual biases and decision processes that have little to do with perception (e.g., Balota, 1990; Balota & Chumbley, 1984, 1985; Balota & Lorch, 1986; Chumbley & Balota, 1984; Lorch et al., 1986; Neely, 1991; Neely & Keefe, 1989). Although no experimental task is perfect, the naming task is widely believed to be less susceptible to such biases than the lexical decision task, and is taken to be a superior measure of perceptual processing.

The current study uses the naming task but without any priming manipulation. On each trial of the main experiment, a participant heard a word presented over headphones, and repeated the word into a microphone as quickly as possible. The use of auditory stimulus presentation is worth noting for several reasons. First, nearly all of the studies discussed prior to this point have used visual stimulus presentation. Speech is a signal that unfolds over time, so auditorily presented stimuli might allow a better look at the effects in question. Second, in speech there is no orthography-to-phonology translation process, so the arguments of Strain & Herdman (1999, Strain et al., 1995) would seem not to apply. Finally, speech is a more basic and fundamental human cognitive process than reading is, and predates it by at least 25,000 years (Pinker, 1994). We might therefore reason that if a low-level link exists between semantics and the early stages of word recognition, it would be most evident with auditorily presented language.

If previous demonstrations of semantic effects in auditory word recognition are due to problems of post-access contamination stemming from the lexical decision task, then these semantic effects should disappear. Similarly, if semantic effects are only observable for low-frequency exception words, then we should not see such effects more generally. If, however, the results of those studies reflect something real about lexical access processes, and if these effects are general rather than being limited to special cases, then we would expect to see similar results with the naming paradigm.
Danger and Usefulness ratings

The stimuli used in this study were the same words as those used by Wurm and Vakoch (2000). A total of 200 nouns were generated by the authors and two undergraduate research assistants. The major constraint in selecting these nouns was “likelihood of recognisability” by the research participants, so these were common, familiar words. Half of the words were selected at random to be changed into pseudowords. These pseudowords were not used in the current study. Only genuine words were presented to participants.

We chose to have the stimulus items rated on Danger and Usefulness, rather than other potential semantic dimensions (valence, dominance, and arousal, for example, or any of a number of other systems). We had three main reasons for choosing these dimensions. First, as most other frameworks have at least three dimensions, parsimony argues for this framework. Second, we expect the reaction-time data to be more easily interpretable under the chosen semantic framework than under some alternative frameworks, because of how we conceptualise the effects in question. Finally, there is support for the idea that organisms have to find some way to balance competing “approach” and “avoid” goals in order to survive. This way of conceptualising motivation maps nicely onto the two dimensions, a point to which we will return below.

The Danger and Usefulness of the referents of these words were rated on 8-point scales by 32 undergraduate research participants in a preliminary rating study. A complete list of the stimuli (along with mean ratings) can be found in the Appendix. The mean rating on Danger for the 100 words was 3.47 ($SEM = 0.22$), and on Usefulness it was 3.52 ($SEM = .18$). The two dimensions were moderately but significantly correlated with each other ($r = -.25, p < .05$).

Calculation of potential control variables

Many variables are known to affect word recognition performance. Given the present research questions, attempting to equate stimuli on all of these variables would have left too few stimuli to work with. Instead, they were included as factors in the regression model in the naming experiment, before the effects of Danger and Usefulness were assessed.

Eight classes of such variables were identified. The following paragraphs list these classes of variables and describe how each was computed.

1. Frequently occurring or familiar words are generally believed to be more easily processed. *Word frequency* was taken from Francis and Kučera (1982) and *familiarity* ratings were taken from the *Oxford psycholinguistic database* (Quinlan, 1987).

2. *Word length* was computed two ways for each word, in milliseconds and in number of phonemes. Longer words require more time to recognise.
3. The initial phoneme can influence the ease (and thus the speed) of lexical processing, and affects naming times (e.g., Bates, Devescovi, Pizzamiglio, & D’Amico, 1995). Typically, researchers using the naming paradigm try to match different groups of stimuli on the initial phoneme. This study does not have groups of stimuli—each word is essentially its own condition. We therefore included three different characteristics of the onset phoneme: vowel vs. consonant; place-of-articulation; and voicing. Two of these characteristics (vowel versus consonant onset, and voicing) are binary, and so each was simply coded as 0 or 1. We coded three places of articulation: front (i.e., bilabial and labiodental), middle (i.e., interdental and alveolar), and back (i.e., palatal and velar). This phoneme characteristic therefore required two (i.e., \( N-1 \)) dummy-coded variables. For example, “front” was coded as 0 0, “middle” was coded as 1 0, and “back” was coded as 0 1 (see Cohen & Cohen, 1983).

4. Bigram frequency is a measure of how frequently pairs of letters occur, as a function of letter position and word length. This was calculated for each word using the tables of Mayzner and Tresselt (1965). Pairs of letters that occur together very frequency tend to be processed more quickly than unusual pairs.

5. First-syllable stress was coded as either strong (1) or weak (0). Some models of speech perception make a distinction between items with strong first syllables and those with weak first syllables (e.g., Cutler & Norris, 1988; Grosjean & Gee, 1987).

6. Neighbourhood density refers to the number of existing words that are orthographically or phonetically similar. There is abundant evidence that processing difficulty for a given word depends on the number of neighbours it has (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977; Goldinger, Luce, & Pisoni, 1989; Luce et al., 1990).

Four different measures of neighbourhood density were computed. One of the measures used in the present study was the auditory analogue of Coltheart’s N (Coltheart et al., 1977), which is the number of words that can be made from a given target word by the substitution of one letter, preserving letter position. For example, if the target word is pin, then pen, win, and pit (along with many others) would be included in the list of neighbours.

The second approach is like Coltheart’s N, but also allows for the addition or deletion of one phoneme at the beginning or end of a word. Using the above example, we would also add to the list of neighbours words such as spin and pins.

For both of these measures, the summed frequency of all of a word’s neighbours was also computed. This gives a second, frequency-weighted version of each density measure.

7. Concreteness, imagery, and meaningfulness were taken from the Oxford psycholinguistic database (Quinlan, 1987). This database includes two different measures of meaningfulness, both of which were used in the preliminary analyses. Words that are rated as more concrete, more imageable, or more meaningful tend to be recognised more quickly.
8. *Animate or inanimate referent* was also included. It is possible that words referring to animate objects would require less processing time than words referring to inanimate objects.

These eight classes of variables contain a total of 18 variables. These variables were included as factors in the regression model, except in the cases where there were multiple measures of the same construct. In such cases (e.g., for neighbourhood density), the measure that explained the most variance was retained in the analysis. Possible effects of Danger and Usefulness were assessed only after statistically controlling for the effects of these variables.

**NAMING EXPERIMENT**

In this experiment, an attempt is made to observe semantic effects in auditory word recognition using a task less susceptible to post-perceptual biases, the naming task. On each trial, a participant heard a word presented over headphones, and repeated the word into a microphone as quickly as possible. If previous reports that Danger and Usefulness facilitate word recognition are due to shortcomings of the lexical decision task, or are observable only with infrequent or unusual words, then these effects should disappear. If, however, the previously reported semantic effects reflect something real and general about word recognition processes, then this experiment should produce similar results.

**Method**

*Participants.* A total of 45 undergraduate students from the Wayne State University psychology subject pool participated. All were native speakers of English with normal hearing. Participants received extra credit in a psychology course for their participation.

*Materials.* The 100 words described in the previous section were used, along with 115 filler items. The fillers were highly recognisable English words, of various parts of speech, and with differing morphological structures. The filler items were from one to five syllables in length, as were the critical items. Each stimulus was read by a male native speaker of English who was unfamiliar with the purpose of the study, digitised at a sampling rate of 20 kHz (low-pass filtered at 9.8 kHz) and stored in a disk file.

*Procedure.* Participants were tested individually a sound-attenuating booth. They listened to the digitised speech files, played over headphones at a comfortable listening level, with 1500 ms between trials. Participants were instructed to repeat each word they heard, as quickly and accurately as possible, into a microphone that was positioned approximately 10 cm away. If no response was detected within 2000 ms, the program moved on to the next trial. A different random stimulus order was used for each participant. Accuracy was
scored trial-by-trial by the experimental software (Inquisit, version 1.28, 2000). Responses were also tape-recorded and subsequently checked for accuracy. A practice set of 20 words was used prior to the main experiment to familiarise participants with the procedure.

Results and discussion

Naming times were measured from the uniqueness point (UP) of each word (see Marslen-Wilson, 1984; Marslen-Wilson & Welsh, 1978). The first author and a research assistant made independent measurements of each UP, which was defined as the middle of the prototypical segment of the particular phone in question (following Radeau, Mousty, & Bertelson, 1989; see also Wurm & Ross, 2001). This point was located using both visual and auditory criteria, with the help of a commercial waveform editor. Measurements for a given word were generally within a few ms of each other: the mean difference for the 100 critical items was 4.3 ms (SEM = 3.1 msec). The value used for the UP of each stimulus was the mean of these two independent measurements.

Naming times for trials on which the participant pronounced the wrong word were not included (2.3% of the trials). RTs were also excluded from the analyses if they were due to a noise other than initiation of the verbal response, such as coughing, or shuffling feet (< 1% of the trials). The grand mean for the 4395 valid naming times was 550 ms (SD = 350.8 ms).

Main regression analysis. The analytic strategy used in the current study parallels that of our earlier studies (Vakoch & Wurm, 1997; Wurm & Vakoch, 1996, 2000). Lorch and Myers (1990) provide a discussion of repeated-measures regression analyses in cognitive research that is extremely valuable in the present context (see also Cohen & Cohen, 1983). A hierarchical multiple regression analysis was conducted, in which the main effects of Danger and Usefulness were assessed, along with their interaction. The results of this analysis are shown in Table 1. Variables were entered in four steps, with simultaneous entry for all variables within a given step. Step 1 of the analysis was to partition the between-subjects variance. In repeated-measures regression analyses this is done by entering a block of N−1 (i.e., 44) dummy variables that represent the participants. At Step 2 of the analysis, the control variables were entered.

Steps 3 and 4 of the analysis assessed the semantic effects of interest. On Step 3, the main effects of Danger and Usefulness were assessed. Higher ratings on both Danger and Usefulness were significantly associated with faster RTs (B =

1An alternative to measuring RTs for auditorily presented materials is to time from word onset, and then statistically partial out item durations and UPS. The effects reported are significant regardless of the method chosen, but we will focus on the UP measurements so that direct comparisons can be made to the study by Wurm and Vakoch (2000).
TABLE 1
Summary of hierarchical regression analysis for variables predicting reaction time

<table>
<thead>
<tr>
<th>Variable(s)</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td>44</td>
<td>5,922,197</td>
<td></td>
</tr>
<tr>
<td>Control variables</td>
<td>10</td>
<td>3,876,666</td>
<td></td>
</tr>
<tr>
<td>Main effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danger</td>
<td>1</td>
<td>277,276</td>
<td>5.56*</td>
</tr>
<tr>
<td>Usefulness</td>
<td>1</td>
<td>655,136</td>
<td>15.36***</td>
</tr>
<tr>
<td>Danger × Usefulness</td>
<td>1</td>
<td>3,061</td>
<td>&lt;1.0</td>
</tr>
<tr>
<td>Error terms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects × Danger</td>
<td>44</td>
<td>49,893</td>
<td></td>
</tr>
<tr>
<td>Subjects × Usefulness</td>
<td>44</td>
<td>42,659</td>
<td></td>
</tr>
<tr>
<td>Subjects × Danger × Usefulness</td>
<td>44</td>
<td>51,494</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; ***p < .001.

−4.01, \( SE \, B = 1.79, \beta = −.02 \); and \( B = −7.56, \, SE \, B = 2.20, \beta = −.04 \), respectively). Step 4 of the analysis assessed the Danger × Usefulness interaction, which was not significant (\( B = −0.23, \, SE \, B = 0.98, \beta = −.01 \)). The remaining values in Table 1 provide the appropriate error terms for the three effects of interest (see Lorch & Myers, 1990).

Each step up the 8-point Danger scale predicts a 4.01 ms naming time advantage, and each step up the Usefulness scale translates to a 7.56 ms advantage. Thus, Danger could potentially be thought of as a 32-ms effect, if we were to contrast very low and very high; Usefulness would be a 60-ms effect by the same reasoning. These are substantial effects when conceptualised in this way.

It is conceivable that the selection of stimuli resulted in unintentional semantic priming across trials, because even though the stimulus list included 115 words unrelated to any of the critical items, many of the critical items themselves were at least weakly related to one another. If such priming occurred, there should be a significant negative relationship between naming times and trial number. That is, the later in the experiment a particular word was heard, the faster the naming time should be. A check of the data showed a very weak relationship in the opposite direction (\( r = +.08 \), n.s.). Another way to assess for semantic priming effects is to compare performance for words that belong to thematically related subsets (e.g., the food-related items, the weapon-related items, and so on) to performance for words that do not belong to such a subset. These comparisons showed no hint of a difference (548 ms vs. 550 ms, with error rates of .02 for each group). Therefore, it seems reasonable to conclude that there was no unintentional semantic priming across trials.
GENERAL DISCUSSION

This study replicates the Wurm and Vakoch (2000) lexical decision study, with the exception of the significant Danger × Usefulness interaction reported in that study. The significant effect of Danger found with the naming paradigm is almost exactly the same size as the one reported with lexical decision; the Usefulness effect is about one-third smaller with naming than it was with lexical decision. The interpretation offered in the earlier study (based on the relative sizes of the regression coefficients—see Pedazur, 1982) was that the classification of something along the Usefulness dimension takes precedence; classification along the Danger dimension, although certainly important, was not as important. The current results lend additional support to this interpretation, with the Usefulness coefficient being nearly twice as large as the Danger coefficient. However, one of the primary conclusions of the current study warrants emphasizing here: although the relative sizes of these effects may change a bit across different experimental paradigms, in both cases each effect is reliable.

A primary goal of the current study was to demonstrate that one need not rely on the potentially flawed lexical decision task in order to observe early semantic effects in lexical access. This was demonstrated quite unambiguously. One may question, however, whether these semantic effects are limited to this particular set of words, because the stimulus items used in this study were also used by Wurm and Vakoch (2000). This seems very unlikely, as we have demonstrated semantic effects with other stimulus sets and under other test conditions. Specifically, Wurm and Vakoch (1996) found semantic effects in the processing of emotion words, whereas Vakoch and Wurm (1997) found such effects in the processing of words drawn at random from a large dictionary. We do not believe that there is anything unusual about the stimulus set of the current study that would provide an alternative explanation for the results.

The idea that the process of natural selection has had a role in shaping language has been argued for more than a century (e.g., Darwin, 1871), and both of the main effects found in the current study have natural explanations in terms of adaptiveness and survival. Dangerous things call for quick decisions (the familiar ‘fight or flight’ dilemma), and indeed, it was found that words with more dangerous referents were named faster. The other main effect reflects a motive less often mentioned but no less important: The need to acquire valuable resources. Words with more useful referents were named faster (and it bears repeating that ‘useful’ was defined in the preliminary rating study with explicit reference to human survival). Information exchange is beneficial to the extent that it serves one or both of these purposes (Davidson, 1992; Darwin, 1859/1968; Schneirla, 1965).

There is experimental support from physiological and brain research for a biphasic view of emotion. Specifically, researchers have found evidence that there are two different kinds of reactions to affective stimuli: appetitive
reactions and defensive reactions (Bradley, Codispoti, Cuthbert, & Lang, 2001; Lang, Bradley, & Cuthbert, 1998), which map very nicely to the semantic dimensions used in the current study (Usefulness and Danger, respectively). Their research has examined physiological responses (e.g., skin conductance, heart rate, and modulation of the startle reflex), as well as subjective participant ratings, to visually presented materials with different kinds of affective content. It is argued by the authors that the appetitive and defensive responses evolved to promote the survival of individuals and species.

Any attempts to answer the question of which brain structures might be involved in the influence of Danger and Usefulness on lexical access are of necessity speculative at this point. However, there are a number of findings that might provide useful points of departure in future investigations in this area. Lang et al. (1998) used functional magnetic resonance imaging (fMRI) to examine activity in the visual cortex during viewing of affective pictures. They found that activity was greater in all sampled brain regions for emotional pictures than for neutral pictures. Both emotional and neutral pictures produced activity centred on the calcarine fissure (Area 17), but only emotional pictures led to activity bilaterally in the occipital gyrus, in the right fusiform gyrus, and in the right inferior and superior parietal lobules.

In addition to fMRI, an increasingly common and very useful approach has been to take positron emission tomographic (PET) measurements of regional cerebral blood flow during picture viewing (see Posner & Raichle, 1994, for a useful and thorough discussion of brain imaging techniques in cognitive research). This technology has been used to demonstrated that the amygdala plays a role in human fear conditioning (Furmark, Fischer, Wik, Larsson, & Fredrikson, 1997), and that visually induced anxiety is associated with alterations in limbic, paralimbic, and cortical brain regions that are of relevance for cognition and affect (Fredrikson, Fischer, & Wik, 1997). More to the point of the present discussion, Lane et al. (1997) attempted to explore the neural substrates of the appetitive and defensive motivational system by using this technology. They found that both pleasant and unpleasant pictures led to increased cerebral blood flow in the vicinity of the medial prefrontal cortex (Area 9), thalamus, hypothalamus and midbrain, as compared to neutral pictures. In addition, only unpleasant pictures produced activation of the bilateral occipito-temporal cortex and cerebellum, and left parahippocampal gyrus, hippocampus, and amygdala. Finally, the authors found that pleasant pictures produced more activity than neutral pictures in the head of the left caudate nucleus.

This research lends further support to the idea that there are qualitatively different kinds of responses to affective stimuli. The mapping of Danger and Usefulness to the “pleasant versus unpleasant” distinction that many other studies have used is not entirely straightforward, so we should be careful not to overstate the link between existing research and the current study. Nevertheless, although these imaging studies are quite different from the lexical decision and
naming studies we have conducted, there appears to be a compelling correlation between the supposed biphasic reaction types and differing levels of activity across several regions of the brain. It is also worth noting that acoustic cues appear to activate these two “motivational circuits” (i.e., appetitive and defensive) in ways similar to pictures (e.g., Bradley & Lang, 2000).

It is crucial to distinguish the current findings from other kinds of semantic effects. A great number of studies have demonstrated semantic effects in memory, or with respect to what might be considered special situations: priming, various kinds of contextual manipulations, resolution of ambiguity, and interpretation of idiomatic expressions (e.g., Clark, 1984; Cutting & Bock, 1997; Damian & Martin, 1999; Slowiaczek, 1994). Studies such as these are valuable and interesting, but the focus of the current study is much more basic. This experiment has shown that the very time-course of perceiving an isolated word, without prior presentation of a prime and indeed without any kind of semantic or sentential context, is determined in part by characteristics of that word’s meaning, independent of a host of other factors.

The present study confirms and extends the evidence for early semantic access provided by Strain et al. (1995). As mentioned above, they hypothesised that early semantic activation takes place for all words, but they were only able to observe this activation for low-frequency exception words. This was to be expected because that study used the reading-aloud task in which the orthography-to-phonology mapping is of crucial importance. The current study used a different experimental paradigm that allows for a demonstration of early semantic effects for all word types.

To be specific about the frequencies, stimuli in the current study had a mean frequency of 25 (SD = 53.4), whereas those of Strain et al. (1995) had a mean frequency of 6 (SD approximately 4). In another study Strain and Herdman (1999, experiment 2) wanted to demonstrate a concreteness effect outside the low-frequency range, but even in that experiment, the mean frequency of the stimuli was about 8. There are obviously many other differences between the present study and these earlier ones: presentation modality (visual vs. auditory), stimulus length (mono- and bisyllabic words, vs. a mixture of words from 1 to 5 syllables), and variable of main interest (concreteness, which was treated as a potential confounding variable in the current study). Nevertheless, it can be concluded that the emergence of early semantic effects does not depend on very low word frequency.

Work from a number of different areas is converging on a conclusion similar to the one we make in the current study. For example, recent research shows that semantic factors affect the storage and processing of morphologically complex words (e.g., Libben, 1998; Marslen-Wilson, Tyler, Waksler, & Older, 1994; Schreuder & Baayen, 1995; Wurm, 1997; Wurm & Ross, 2001). There is also intriguing evidence of early semantic effects from studies that measure event-related potentials (ERPs) from the surface of the scalp (van Petten, Coulson,
Rubin, Plante, & Parks, 1999): N400 ERPs were found to be different for contextually appropriate versus contextually inappropriate words 200 ms before the UP of the word. The authors concluded that “…semantic integration can begin to operate with only partial, incomplete information about word identity” (p. 394). Finally, as outlined in the Introduction, Murphy and Zajonc’s (1993) affect primacy hypothesis, the study by Bargh et al. (1992), and Robinson’s model of appraisal (1998) all lead one toward the possibility that widespread, early acting, automatic evaluative/emotional classifications are a part of everyday cognition, underlying a wide range of mental processes.

As outlined in the introduction, models of lexical processing have not generally included explicit discussions of semantic considerations; semantic effects have often been assumed to come into play chiefly at some later, post-perceptual stage. Together with previous work from our lab and others, the current study suggests that this exclusion of semantics as a routine factor in the early stages of perceptual processing is misguided.

A wide range of models could in principle accommodate the current findings. Dimension weights on Danger and Usefulness could be incorporated in such a way as to alter the resting activation levels of words, or to adjust the threshold values needed for word recognition. Words might be ordered in terms of their Danger or Usefulness values, very much akin to the way some models have bins of words ordered by their frequencies of occurrence. Information about Danger and Usefulness could be conceptualised as relatively high-level information that is “fed down” to lower processing levels in a network. Any of these options would be fairly straightforward (it depends chiefly on what the particular model or metaphor calls for), once it is determined more definitively what the relevant semantic dimensions are. We think Danger and Usefulness (or a set like Evaluation, Potency, and Activity) are good candidates, given some of the motivational priorities of the human organism.

Manuscript received 12 February 2001
Revised manuscript received 24 January 2002

REFERENCES


### APPENDIX

**Stimuli and Mean Ratings**

<table>
<thead>
<tr>
<th>Word</th>
<th>Danger</th>
<th>Usefulness</th>
<th>Word</th>
<th>Danger</th>
<th>Usefulness</th>
</tr>
</thead>
<tbody>
<tr>
<td>ant</td>
<td>1.97</td>
<td>2.00</td>
<td>cannon</td>
<td>6.63</td>
<td>2.78</td>
</tr>
<tr>
<td>appendix</td>
<td>2.25</td>
<td>3.22</td>
<td>canteen</td>
<td>1.63</td>
<td>5.22</td>
</tr>
<tr>
<td>apple</td>
<td>1.22</td>
<td>5.97</td>
<td>canyon</td>
<td>3.53</td>
<td>2.09</td>
</tr>
<tr>
<td>arrow</td>
<td>6.09</td>
<td>4.72</td>
<td>card</td>
<td>1.16</td>
<td>1.81</td>
</tr>
<tr>
<td>badger</td>
<td>2.97</td>
<td>2.73</td>
<td>carrot</td>
<td>1.38</td>
<td>5.34</td>
</tr>
<tr>
<td>balloon</td>
<td>1.84</td>
<td>1.81</td>
<td>chicken</td>
<td>1.97</td>
<td>5.88</td>
</tr>
<tr>
<td>banana</td>
<td>1.16</td>
<td>5.97</td>
<td>clothing</td>
<td>1.41</td>
<td>6.88</td>
</tr>
<tr>
<td>basket</td>
<td>1.19</td>
<td>3.66</td>
<td>club</td>
<td>4.53</td>
<td>3.28</td>
</tr>
<tr>
<td>blanket</td>
<td>1.31</td>
<td>6.03</td>
<td>collision</td>
<td>6.34</td>
<td>1.63</td>
</tr>
<tr>
<td>bottle</td>
<td>2.69</td>
<td>4.66</td>
<td>constellation</td>
<td>1.06</td>
<td>2.69</td>
</tr>
<tr>
<td>buffalo</td>
<td>3.44</td>
<td>3.88</td>
<td>cork</td>
<td>1.22</td>
<td>2.69</td>
</tr>
<tr>
<td>burglar</td>
<td>6.50</td>
<td>1.16</td>
<td>corner</td>
<td>1.72</td>
<td>2.34</td>
</tr>
<tr>
<td>butterfly</td>
<td>1.13</td>
<td>2.19</td>
<td>crime</td>
<td>6.81</td>
<td>1.81</td>
</tr>
<tr>
<td>cancer</td>
<td>7.81</td>
<td>1.28</td>
<td>crossbow</td>
<td>5.61</td>
<td>4.81</td>
</tr>
</tbody>
</table>

(Continued)
### APPENDIX
(Continued)

<table>
<thead>
<tr>
<th>Word</th>
<th>Danger</th>
<th>Usefulness</th>
<th>Word</th>
<th>Danger</th>
<th>Usefulness</th>
</tr>
</thead>
<tbody>
<tr>
<td>crow</td>
<td>1.97</td>
<td>2.47</td>
<td>pesticide</td>
<td>5.69</td>
<td>3.34</td>
</tr>
<tr>
<td>desert</td>
<td>3.69</td>
<td>1.91</td>
<td>philosophy</td>
<td>1.63</td>
<td>2.63</td>
</tr>
<tr>
<td>desk</td>
<td>1.56</td>
<td>2.97</td>
<td>pin</td>
<td>2.72</td>
<td>2.66</td>
</tr>
<tr>
<td>dust</td>
<td>1.88</td>
<td>1.41</td>
<td>plague</td>
<td>7.81</td>
<td>1.09</td>
</tr>
<tr>
<td>dynamite</td>
<td>7.28</td>
<td>3.16</td>
<td>poetry</td>
<td>1.13</td>
<td>2.00</td>
</tr>
<tr>
<td>eagle</td>
<td>2.19</td>
<td>3.47</td>
<td>poison</td>
<td>7.59</td>
<td>1.41</td>
</tr>
<tr>
<td>echo</td>
<td>1.25</td>
<td>1.44</td>
<td>pollution</td>
<td>5.94</td>
<td>1.38</td>
</tr>
<tr>
<td>egg</td>
<td>1.69</td>
<td>5.50</td>
<td>potato</td>
<td>1.31</td>
<td>6.03</td>
</tr>
<tr>
<td>electricity</td>
<td>5.00</td>
<td>6.03</td>
<td>quicksand</td>
<td>5.94</td>
<td>1.19</td>
</tr>
<tr>
<td>elephant</td>
<td>3.50</td>
<td>3.19</td>
<td>rabbit</td>
<td>1.72</td>
<td>3.84</td>
</tr>
<tr>
<td>fire</td>
<td>7.03</td>
<td>7.16</td>
<td>razor</td>
<td>5.34</td>
<td>4.00</td>
</tr>
<tr>
<td>fish</td>
<td>2.50</td>
<td>5.47</td>
<td>scorpion</td>
<td>6.44</td>
<td>1.94</td>
</tr>
<tr>
<td>flag</td>
<td>1.91</td>
<td>2.25</td>
<td>shoe</td>
<td>1.66</td>
<td>5.13</td>
</tr>
<tr>
<td>food</td>
<td>1.94</td>
<td>7.97</td>
<td>skunk</td>
<td>1.78</td>
<td>2.31</td>
</tr>
<tr>
<td>fork</td>
<td>2.56</td>
<td>3.91</td>
<td>snake</td>
<td>5.84</td>
<td>2.28</td>
</tr>
<tr>
<td>hammer</td>
<td>3.97</td>
<td>4.56</td>
<td>spear</td>
<td>6.28</td>
<td>5.72</td>
</tr>
<tr>
<td>hook</td>
<td>4.13</td>
<td>4.34</td>
<td>spoon</td>
<td>1.56</td>
<td>3.47</td>
</tr>
<tr>
<td>hurricane</td>
<td>7.06</td>
<td>1.56</td>
<td>stove</td>
<td>3.31</td>
<td>5.69</td>
</tr>
<tr>
<td>joke</td>
<td>1.50</td>
<td>3.97</td>
<td>strawberry</td>
<td>1.34</td>
<td>5.16</td>
</tr>
<tr>
<td>kerosene</td>
<td>4.94</td>
<td>4.78</td>
<td>sunset</td>
<td>1.16</td>
<td>3.50</td>
</tr>
<tr>
<td>knife</td>
<td>6.38</td>
<td>5.75</td>
<td>syringe</td>
<td>5.28</td>
<td>5.75</td>
</tr>
<tr>
<td>ladder</td>
<td>2.44</td>
<td>3.25</td>
<td>tarantula</td>
<td>5.81</td>
<td>1.94</td>
</tr>
<tr>
<td>lamp</td>
<td>1.59</td>
<td>4.41</td>
<td>telescope</td>
<td>1.34</td>
<td>2.34</td>
</tr>
<tr>
<td>lava</td>
<td>6.06</td>
<td>1.50</td>
<td>thorn</td>
<td>2.59</td>
<td>1.69</td>
</tr>
<tr>
<td>leaf</td>
<td>1.56</td>
<td>4.75</td>
<td>tiger</td>
<td>6.13</td>
<td>2.97</td>
</tr>
<tr>
<td>lightning</td>
<td>6.06</td>
<td>2.72</td>
<td>toad</td>
<td>1.97</td>
<td>2.16</td>
</tr>
<tr>
<td>lint</td>
<td>1.34</td>
<td>1.22</td>
<td>tobacco</td>
<td>6.44</td>
<td>1.53</td>
</tr>
<tr>
<td>lion</td>
<td>5.75</td>
<td>3.03</td>
<td>tonsil</td>
<td>2.44</td>
<td>3.44</td>
</tr>
<tr>
<td>machete</td>
<td>6.13</td>
<td>3.91</td>
<td>tornado</td>
<td>7.00</td>
<td>1.41</td>
</tr>
<tr>
<td>moose</td>
<td>3.75</td>
<td>3.16</td>
<td>toxin</td>
<td>7.22</td>
<td>1.38</td>
</tr>
<tr>
<td>mugger</td>
<td>6.75</td>
<td>1.09</td>
<td>tree</td>
<td>2.72</td>
<td>6.88</td>
</tr>
<tr>
<td>nail</td>
<td>3.94</td>
<td>4.78</td>
<td>twig</td>
<td>1.38</td>
<td>2.75</td>
</tr>
<tr>
<td>needle</td>
<td>3.41</td>
<td>5.00</td>
<td>waltz</td>
<td>1.09</td>
<td>1.78</td>
</tr>
<tr>
<td>ointment</td>
<td>1.66</td>
<td>5.25</td>
<td>water</td>
<td>3.19</td>
<td>7.97</td>
</tr>
<tr>
<td>opera</td>
<td>1.25</td>
<td>1.81</td>
<td>wood</td>
<td>1.91</td>
<td>6.91</td>
</tr>
<tr>
<td>parsley</td>
<td>1.13</td>
<td>3.25</td>
<td>wool</td>
<td>1.22</td>
<td>5.41</td>
</tr>
</tbody>
</table>

*Note:* Ratings were made on 1–8 scales.