Danger and usefulness effects as a function of concept ancientness

Devon Witherell, Lee H. Wurm, Sean R. Seaman, Nikki A. Brugnone, and Evan T. Fulford
Wayne State University

Danger and Usefulness affect word recognition (e.g., Wurm & Vakoch, 2000), and a related construct affects memory (e.g., Nairne, Thompson, & Pandeirada, 2007). We tested hypotheses about differential effects of these dimensions, based on the relevance of concepts relative to the time perceptual systems underwent selection pressures. In two experiments participants made auditory and visual lexical decisions for nouns rated on Danger, Usefulness, and “Ancientness.” Danger interacted with Ancientness in both auditory and visual processing. Increasing Danger led to faster RTs and better accuracy only for words judged to have ancient relevance. Interactions with participant gender were seen in both auditory and visual analyses. Consistent with Wurm, Whitman, Seaman, Hill, and Ulstad (2007), men’s but not women’s auditory performance improved with increasing Danger. The beneficial effect of Usefulness on accuracy was greater for women than for men in both experiments. Danger and Usefulness effects seem to reflect a general principle underlying human cognition.

Keywords: fitness relevance, survival, perception, word recognition

Individuals are bombarded with environmental stimuli, and how one perceives these stimuli allows for their categorization. This leads to the maintenance of an organized and coherent view of the world. While many of the types of stimuli have changed, this task is the same as that our early ancestors faced, although the consequences then were frequently much more severe for any errors in categorization. It seems reasonable to surmise that those who survived to live another day were able to do so in part because of their ability to quickly and correctly categorize stimuli that were relevant to their continued existence. In fact, many of our species’
cognitive abilities were shaped during the Pleistocene epoch (e.g., Symons, 1992), which began around 2 million years ago and ended about 10,000 years ago. It has been suggested that the “footprint” of survival processing during this time should be evident in current cognitive processing (e.g., Nairne, Pandeirada, Gregory, & Van Arsdall, 2009; Silverman, Choi, & Peters, 2007). Below we discuss evidence for this view from research paradigms in language and memory.

Osgood (1969) attributed the early navigational successes of our species to emotional reactions, and argued that there are three dimensions that encapsulate how affective survival-relevant stimuli are perceived. He labeled these Evaluation, Potency, and Activity. Evaluation is how good or bad something is, Potency is how strong or weak something is, and Activity is how quick or slow something is. In Osgood’s view, these dimensions act as the affective components of the total meaning of a word, allowing for quick and accurate gut reactions that aid in survival. Schlosberg (1954) had previously discussed similar dimensions with regard to communication and facial expression. More recent appraisal models focus on similar dimensions of Valence and Arousal, with Osgood’s quick and accurate gut reactions being facilitated by preattentive processes evaluating whether a stimulus or event is personally relevant, whether it is consistent with or counter to a person’s goals, and whether it can be dealt with effectively (Robinson, Storbeck, Meier, & Kirkeby, 2004; see also Robinson, 1998).

Research showing that affective stimuli elicit either defensive/aversive reactions or appetitive ones (Bradley, Codispoti, Cuthbert, & Lang, 2001) fits nicely with this line of thinking. It also supports the idea of a link between perception and action, offering an adaptive explanation for emotion effects in word recognition (see below). Data showing that information on the function of man-made objects is activated quicker than information on their physical forms (Moss, McCormick, & Tyler, 1997) supports this perception-action link as well, as do studies demonstrating automatic processing on Evaluation. In a classic study Chen and Bargh (1999) had people push or pull a lever in order to classify words as good or bad. In one condition the lever was pushed for bad words and pulled for good words, and in the other the opposite response pattern was used. Reaction times were faster for the “good-pull, bad-push” condition. To show the implicit nature of this reaction, the experiment was repeated with a task that was not associated with categorizing the words as good or bad and the same effect was found. A similar result was found in another study (Moors & De Houwer, 2001) where the task was to judge whether a word was a noun or adjective by responding in a toward-or-away manner. These results suggest that appetitive or aversive reactions to objects are activated quickly and automatically, and often implicitly.

Effects such as these fit well with Schneirla’s (1965) characterization of the evolutionary goals of an organism, which he explained in a theory of biphasic
approach/withdrawal processes. This theory posits that an organism has two goals that elicit very different behavioral responses: approach (stemming from a desire to obtain beneficial resources) and withdrawal (stemming from a desire to avoid dangerous and intense situations). These behaviors are goal-oriented in that they promote the survival of the organism, and these actions are linked with perception (one must perceive something to be useful or dangerous to elicit the relevant reaction). Such perception-action links lend themselves to the idea of cognition being embodied or grounded (e.g. Barsalou, 2008), a point to which we will return below.

There is evidence that Osgood’s affective dimensions (i.e. Evaluation, Potency, and Activity) affect word recognition as well (e.g. Vakoch & Wurm, 1997; Wurm, 2011; Wurm & Vakoch, 1996; Wurm, Vakoch, & Seaman, 2004a). However, there is still the question of how such effects work, as well as the broader theoretical question of what function they may serve. It could be that emotion constructs tap into perceptual systems originally designed for enhanced processing of stimuli that bear relevance to survival, perhaps through their influence on the allocation of processing resources. The influence of affective dimensions on word recognition could then be a reflection of the evolutionary footprint left behind by the survival processing of our ancestral past.

Interpreting the evolutionary significance of stimuli using Osgood’s dimensions is not straightforward. Wurm and Vakoch (2000) pointed out that simply knowing that something is rated low on Evaluation (i.e. it has negative Valence) does little to inform about how harmful or beneficial something might be. As they noted,

…the low end of the Evaluation dimension can describe a range of constructs that are each evolutionarily important. Something can be bad and dangerous (e.g. lion or cancer), or bad and harmless (e.g. skunk or mosquito); something can be bad and useful (e.g. knife or syringe), or bad and useless (e.g. scorpion or quicksand). The evolutionary significance of words, defined using Osgood’s dimensions, is always somewhat vague (p. 179).

Because of this, Wurm and Vakoch (2000) reconceptualized Osgood’s dimensions in a way that integrated some of Schneirla’s (1965) ideas. They argued that these reconceptualized dimensions, Danger and Usefulness, are more directly connected to human survival than Osgood’s dimensions, and they tested how ratings on these dimensions affected lexical decision times. They found that Usefulness was most important, and that Usefulness also moderated the effect of the Danger ratings. For words low on Usefulness, increasing Danger was associated with faster RTs. For words high on Usefulness, though, as Danger increased, RTs did as well. This unexpected increase in RT was attributed to Schneirla’s (1965) ideas of approach and withdrawal and the conflicting behavioral responses (or motivational states) activated by the words (both approach and withdrawal; Wurm & Seaman, 2008).
As mentioned previously, such perception-action links support the idea that cognition is based on perception, a theoretical perspective called grounded cognition (Barsalou, 2008). This perspective holds that the cognitive system likely evolved to support action in specific situations, resulting in cognition being “grounded” in a number of ways (e.g., situated action, bodily states, simulation). Instead of knowledge being represented by amodal symbols, multimodal representations exist that incorporate experienced perceptual, motor, and introspective states. With each experience involving an instance of a concept, these modal states are captured and integrated with that concept’s representation. When that knowledge is called upon later this state information becomes reactivated and allows for the simulation of how the brain represented those experiences perceptually, motorically, and introspectively. Such simulations have the ability to integrate typical behaviors associated with the experience of a concept, linking perception and action together.

Building on the logic of grounded cognition, Wurm and Seaman (2007) argued that the inhibitory effect of Danger for items relatively high on Usefulness might be due to behavioral uncertainty. They argued that already during lexical processing, the cognitive system spends attentional resources readying the appropriate real-world response that a particular item calls for. This is possible because of how easy and automated word recognition is. Items high on both Danger (withdraw) and Usefulness (approach) were argued to create behavioral uncertainty for listeners, leading to the observed slow-down in RTs.

Several studies have shown such effects in auditory lexical decision (Wurm, 2007; Wurm et al., 2007), and they have also been found in other word recognition paradigms like naming and perceptual identification (Wurm & Seaman, 2008; Wurm, Vakoch, Aycock, & Childers, 2003; Wurm, Vakoch, Seaman, & Buchanan, 2004b). Our work with these dimensions has been aimed at explicating the organizing principle, at grounding the why of the interrelationships between emotion and cognition (cf. Nairne & Pandeirada, 2010). We have argued that existing frameworks are not well positioned to deal with such questions (e.g. Wurm, 2007, 2012). For example, in the context of models built around Evaluation/Valence, “…what does good mean? Good for whom, and to what end?” (Wurm, 2007, p. 1218). Why do people respond quickly or slowly to things with positive or negative Valence? Not only do Danger and Usefulness allow researchers to ask a different kind of question than dimensions like Valence, Wurm (2007) showed that Danger and Usefulness effects, including their interaction, remained significant even when Valence and Arousal values were included in the statistical model. Danger and Usefulness explained more variance than Valence and Arousal in auditory lexical decision times, too. Wurm and Seaman (2008) demonstrated the same thing for auditory naming.
Although Danger and Valence are obviously related constructs, it is important to note that they are not the same thing. There is evidence to suggest that negative Valence inhibits performance on simple cognitive tasks (e.g. Pratto & John, 1991; Schimmack, 2005). This is thought to stem from cognitive resources being automatically allocated to the processing of affective components of a stimulus, leaving less available for the primary task itself. In the context of word recognition tasks such inhibitory effects have been demonstrated (Vakoch & Wurm, 1997; Wurm, 2011; Wurm et al., 2004a), and so one might be inclined to expect similar inhibitory effects of Danger. However, it is important to point out that Danger is cut from a different conceptual cloth. Danger is not based on how positive or negative a stimulus is perceived to be, but instead on the survival-relevant risk a stimulus is perceived to have. The difference between Danger and Valence can be illustrated by words rated quite negatively that are harmless (e.g. dust, skunk, tomb) or rated quite positively even though they are dangerous (e.g. canyon, electricity, ocean). Supporting the idea that Danger and Valence are not simply two names for the same underlying construct, in our previous work on word recognition we have found consistent facilitative effects of Danger ratings on lexical processing (Wurm, 2007; Wurm & Seaman, 2008; Wurm & Vakoch, 2000; Wurm et al., 2003; Wurm et al., 2004b; Wurm et al., 2007).

As reinterpretations of Osgood’s dimensions, Danger and Usefulness still maintain a relationship with emotion. Fear is a basic emotion in every model that has basic emotions, and its relation to Danger is straightforward. Öhman, Flykt, and Esteves (2001) hold that there is an evolutionary bias to the allocation of processing resources, such that they are allocated not on the basis of negativity per se, but more specifically survival-related threats that have been faced during human evolution (snakes, spiders, facial expressions of anger, etc.). This view is supported by studies showing that it is easier to detect such threats compared to non-threatening distracters, and that this detection (shown through a visual search task) is automatic (Lipp, Price, & Tellegen, 2009; Öhman et al., 2001; Öhman & Mineka, 2001; but see Fox, Griggs, & Mouchlianitis, 2007, for a different view).

Recent work lends support to the utility of Usefulness in an affective framework, too. Izard’s model (2007) gives a prominent role to an emotion he calls Interest. It is described as “the principle force in organizing consciousness” and “most likely to be the emotion in the human mind that continually influences mental processes” (p.271). Other researchers have argued that Interest (or something similar) fosters exploration and learning, and directs attention (e.g. Silvia, 2008). This is similar to Panksepp’s (2007) basic emotion of Seeking, identified in a review of the animal literature and supported by dopamine circuits. Biederman and Vessel (2006) describe humans as infovores and posit that they are designed in such a way that seeking new information is an integral part of the cognitive
system. Their theory relates information-seeking to the releasing of opioids in association cortex. The Usefulness dimension maps well onto the basic emotion of Interest (or Seeking) because of its fostering of exploration and learning.

Language is not the only cognitive process affected by survival-relevant variables. Nairne et al. (2007) had participants make “survivability” ratings of items, and then later gave them a surprise memory test. The rating instruction was:

In this task, we would like you to imagine that you are stranded in the grasslands of a foreign land, without any basic survival materials. Over the next few months you’ll need to find steady supplies of food and water and protect yourself from predators. We are going to show you a list of words, and we would like you to rate how relevant each of these words would be for you in this survival situation. Some of the words may be relevant and others may not — it’s up to you to decide (p. 264).

Memory following survivability ratings was better than that for other kinds of ratings, including a who’s who of encoding variables (and several other control scenarios).

This memory enhancement has been shown to depend on ancientness, too: Weinstein, Bugg, and Roediger (2008) found that the enhancement effect was bigger when items were rated on an ancient grassland scenario than on a modern city scenario, and Nairne et al. (2009) have shown better memory effects in the context of hunting for survival, as opposed to hunting as part of a contest. The encoding scenarios used in these memory studies are very different things compared to a word’s relevance to our ancestral past, but the fact that similar survival-relevant variables have effects in both memory and word recognition paradigms suggests to us that they may reflect a general organizing principle of human cognition.

The only data that exist on the question of ancientness effects in word recognition come from a reanalysis of auditory lexical decision data from Wurm (2007) performed by Wurm, Witherell, Seaman, Brugnone, and Fulford (2009). Items judged by the authors to be Ancient showed stronger Danger and Usefulness effects than did Modern items. The main purpose of the current study was to examine the possible role of ancientness in lexical processing with a study specifically designed to do so.

Based on previous studies (e.g. Wurm, 2007; Wurm & Seaman, 2008; Wurm & Vakoch, 2000; Wurm et al., 2007), we predict facilitative main effects of Danger and Usefulness and an interaction between them. Our account makes no prediction about the main effect of Ancientness. Ancientness of a construct per se does not dictate that it should have fast recognition (e.g., there is no reason to think that an Ancient concept that is also useless and harmless would enjoy a processing advantage). However, we do predict interactions. Specifically, if processing reflects
Ancestral cognitive footprint of survival processing, Danger and Usefulness effects are predicted to be stronger for words with ancient relevance.

We also tested hypotheses about participant gender. There is research demonstrating gender differences in both verbal memory and lexical processing (e.g. Kimura, 2000; Ullman, 2007; Ullman et al., 2002), including research on Danger and Usefulness (Wurm et al., 2007). When combined with research on evolution and gender role differences thought to exist during the Pleistocene epoch (Eagly & Wood, 1999) and their lasting cognitive effects (Silverman et al., 2007), this predicts that men will show larger Danger effects and women will show larger Usefulness effects. These predictions can also be derived from an influential model in which men and women have differing biobehavioral responses to stress (Taylor et al., 2000). One of these predicted interactions has already been observed in the lexical processing literature: For men but not women, increasing Danger was associated with faster lexical decision times (Wurm et al., 2007).

Preliminary rating study 1: Ancientness

Method

Participants. 495 Wayne State University students participated for course credit. All were native speakers of English.

Materials. From the English Lexicon Project (Balota et al., 2007), 109 items were chosen that are not homophones or homographs, and have values listed in the MRC psycholinguistic database (Wilson, 1988) on meaningfulness, concreteness, familiarity, and age of acquisition.

Procedure. Participants were told to judge each word as having ancient relevance (“nomads and cavepeople would have known about it”) or modern relevance (“nomads and cavepeople would not have known about it”). This two-point scale reflected our conceptualization of Ancientness as a dichotomous construct: a concept has either existed since some evolutionarily-important point in human history, or it hasn’t. It should also be noted that this rating procedure defines ancientness solely in terms of age. No distinctions were made as to potential importance or significance of the concepts being rated. We will return to these points below. Stimuli were presented visually in a different random order for each participant.

Results and discussion

We calculated a mean Ancientness rating for each item, hoping for a bimodal distribution of ancient and modern items. However, because the distribution was flat, we
decided to use mean Ancientness as a continuous predictor variable rather than as a two-level factor. Mean ratings ranged from 1.07 (mother) to 1.95 (magazine, hotel).

To make sure that this methodology worked, we had the mean Ancientness ratings examined by experts in evolutionary psychology, historical linguistics, and borrowing (borrowing is relevant because languages are unlikely to borrow words for ancient concepts and objects). While there was an occasional odd item that did not get the kind of rating expected (e.g., participants rated joy as ancient, as expected, but sadness as more modern), the ratings largely corresponded with the experts’ judgments (Martha Ratliff and Glenn Weisfeld, personal communication, Nov. 9, 2009). Thus, while it is debatable whether undergraduates know what a caveperson might or might not have known about, the rating method seems adequate.

Preliminary rating study 2: Danger and usefulness

Method

Participants. 265 (Danger) and 262 (Usefulness) Wayne State University students participated for course credit. All were native speakers of English.

Materials. Stimuli were described above.

Procedure. Participants were instructed to rate either the Usefulness or Danger of each stimulus, using an 8-point scale. Anchor points on the scale were labeled “not at all Useful (or Dangerous) for human survival” (1) and “extremely Useful (or Dangerous) for human survival” (8). Stimuli were presented visually in a different random order for each participant.

Our belief is that these dimensions tap into whatever aspects of Danger or Usefulness allow someone to avoid premature death. These might include things that aid in hunting or gathering, things that can be used for shelter, things that can be eaten or drunk, and so on. No additional instructions or guidance were given to participants on how to make their judgments. The dimension being rated meant whatever each participant decided it meant.

Results and discussion

For each item, the mean Danger and Usefulness ratings were calculated and used as continuous predictors in the main experiment. Mean Danger ratings ranged from 1.64 (circle) to 6.00 (crime). Mean Usefulness ratings ranged from 2.16 (bagpipe) to 7.19 (health).
Experiment 1: Auditory lexical decision

Method

Participants. Ninety-eight Wayne State University students (53 female) participated for course credit. All were native speakers of English.

Materials. Real-word stimuli were described above. To create an equal number of pseudowords, words of the same length were randomly chosen from CELEX and had one phoneme at random changed to another phoneme from the same class. Each stimulus was read by a male native speaker of English who was unfamiliar with the purpose of the study, digitized at a sampling rate of 22.05 kHz and stored in a disc file. Word and pseudoword durations were well matched ($M = 656$ ms, $SEM = 11.3$ for words; $M = 679$ ms, $SEM = 11.6$ for pseudowords; $t(216) = 1.42$, n.s.).

Procedure. Stimuli were played over headphones. Participants were instructed to make speeded lexical decisions by pressing one button for words and another button for pseudowords. A practice set of 20 items allowed participants to get used to the task. Each stimulus was heard once by each participant, in a different random order for each participant. RTs were measured from the onset of each stimulus.

Data analysis. We analyzed the log RT on correct trials with a multilevel linear mixed-effects analysis of covariance, with participant and item as crossed random effects (Baayen, 2008a; Baayen, Davidson, & Bates, 2008; Bates & Sarkar, 2005; see also Crawley, 2002; de Vaan, Schreuder, & Baayen, 2007). In addition to our variables of interest, several control variables were included that capture various characteristics of the stimuli: item duration and uniqueness point (UP) location, in ms; onset voicing (e.g. Baayen, Wurm, & Aycock, 2007); phonological neighborhood density; age of acquisition; meaningfulness; concreteness; and familiarity (final four from the MRC database). Familiarity is a subjective measure of word frequency. It is not wise to have two measures of the same construct in a statistical model, so we opted for familiarity because it had a stronger bivariate correlation with by-item mean RTs than word frequency did. Finally, we also included trial number (to assess practice or fatigue effects). None of these are central to the hypotheses of the current study. They are included either for variance reduction or to eliminate rival hypotheses about the data patterns.

Several variables were log transformed in order to remove the skewness in their distributions and minimize the effect of atypical outliers: word frequency, UP location, item duration, Danger and Usefulness ratings, phonological neighborhood size, and age of acquisition. In addition, the distributions of familiarity and meaningfulness had severe negative skew. These values were inverse, reflect, and square root transformed (Tabachnick & Fidell, 2001; we then multiplied the
transformed values by -1 so that the regression coefficients could be interpreted directly). We also included participant gender as a predictor variable.

We analyzed accuracy with a separate analysis analogous to that above, except with a dichotomous DV (correct vs. incorrect response).

Results and discussion

Three items were discarded from all further analyses due to accuracies at or below chance performance (offshoot, pelt, abode). RTs were log transformed to improve normality. The results of the RT analysis can be seen in the left portion of Table 1. Words with longer durations, later uniqueness points, and later age of acquisition values were associated with slower RTs. Words higher on meaningfulness were associated with faster RTs. Danger, Usefulness, and Ancientness all had facilitative main effects, with the Ancientness effect indicating more Modern words being associated with faster RTs. With the exception of Ancientness, about which we did not make a prediction, all effects were in the expected direction. Figure 1 shows the partial effects of the significant predictors, and as all are plotted on the same y-axis scale, one can see the relative effect size of each.

The Danger x Usefulness interaction was significant and had its expected shape (Figure 2). This plot was created by setting all variables equal to their means except for those involved in the interaction. Danger was represented by a vector running from one to eight. High and low Usefulness were defined (for plotting purposes only) as the mean +/- one standard deviation. Readers should note that our main variables of interest are continuous and were dichotomized only to create interaction plots; thus error bars or confidence intervals would be meaningless. Other plots were created in an analogous fashion. Higher rated Danger was associated with faster RTs for words lower on Usefulness, and with slower RTs for words higher on Usefulness. This is the same pattern seen in previous lexical decision studies (e.g. Wurm, 2007; Wurm & Vakoch, 2000; Wurm et al., 2007). We have interpreted the paradoxical slowing of RTs for words high on both Danger and Usefulness as reflecting behavioral uncertainty associated with an approach/withdraw response conflict (Wurm & Seaman, 2007, 2008).

The hypothesized interaction between Danger and Ancientness was also significant (Figure 3, left panel). Higher rated Danger was associated with faster RTs for more Ancient words as predicted. As word ratings moved toward the Modern end of the dimension, this relationship weakened, and eventually showed a small reversal. Readers are reminded that Ancientness was not dichotomized in these analyses. This is merely a convenient way to show the nature of these continuous interactions.
The right portion of Table 1 shows the results of the error analysis. Words with later age of acquisition or denser phonological neighborhoods had higher error likelihoods. Words higher on meaningfulness, higher on Usefulness, or more Modern had lower error likelihoods. Except for the neighborhood density measure, all of these main effects were also significant in the RT analysis.

The interaction between Danger and Ancientness was significant (Figure 3, right panel). Higher Danger ratings increased accuracy for more Ancient words.
As in the RT interaction shown in the left panel, as word ratings became more Modern, this relationship weakened and then reversed.²

There were also some interactions involving participant gender. Gender interacted significantly with both Danger and Usefulness. As can be seen in Figure 4

**Figure 1.** Partial-effects plot for predicted log RT. Significant effects are shown with 95% credible intervals added (Baayen, 2008b), all on the same y-axis scale. Meaningfulness values were inverse, reflect, and square root transformed, and then multiplied by -1 (see text). The raw RTs represented on this log plot range from approximately 775 to 1050 ms. Data are from Experiment 1 (auditory lexical decision).
Ancient semantic processing

(left panel), with increasing Danger men have a decreasing probability of making an error, while for women the probability of making an error increases. Wurm et al. (2007) found a similar gender x Danger interaction in auditory lexical decision, though the effect was on RT rather than on accuracy.

The right panel of Figure 4 shows the significant gender x Usefulness interaction. For both genders, increasing Usefulness was associated with better accuracy, but the effect was significantly stronger for women. This pattern fits a hypoth-

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure2}
\caption{Predicted log RT as a function of log Danger and Usefulness ratings. Data are from Experiment 1 (auditory lexical decision). "L" means low Usefulness and "H" means high Usefulness (defined as one standard deviation below and above the mean, respectively). Usefulness was not dichotomized for the statistical analysis.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure3}
\caption{Predicted log RT (left panel) and probability of an error (right panel) as a function of Ancientness and log Danger ratings. Data are from Experiment 1 (auditory lexical decision). "A" means more ancient words and "M" means more modern words (defined as one standard deviation below and above the mean, respectively). Ancientness was not dichotomized for the statistical analysis.}
\end{figure}
esized but not observed outcome discussed in Wurm et al. (2007). We will return to this point below.

Finally, the three-way interaction between gender, Danger, and Ancientness was significant. As Figure 5 illustrates, although the patterns were relatively similar, men showed a stronger effect than women on the Ancient words, while women showed a stronger effect than men on the Modern words.

It was important to conduct Experiment 1 in a way that allowed for direct contact with the existing literature on Danger and Usefulness, virtually all of which has used spoken stimuli. In light of the gender effects we found, it is possible that some of the results of Experiment 1 are due to interactions between the gender of

Figure 4. Predicted probability of an error as a function of gender, log Danger ratings (left panel), and log Usefulness ratings (right panel). Data are from Experiment 1 (auditory lexical decision). “M” means male and “F” means female.

Figure 5. Predicted probability of error as a function of gender, Ancientness, and log Danger ratings. Data are from Experiment 1 (auditory lexical decision), and are plotted separately for men (left panel) and women (right panel). “A” means more ancient words and “M” means more modern words (defined as one standard deviation below and above the mean, respectively). Ancientness was not dichotomized for the statistical analysis.
the participants and the gender of the speaker who recorded the stimuli (male). To ensure that our data patterns hold when this factor is removed as a possibility, we ran a second lexical decision study in which the stimuli were presented visually.

Experiment 2: Visual lexical decision

Method

Participants. Eight-eight Wayne State University students (48 female) participated for course credit. All were native speakers of English.

Materials. Stimuli were the same words and pseudowords used in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, except that instead of hearing the stimuli over headphones, participants saw each one printed in uppercase letters in the center of a computer monitor.

Data analysis. We analyzed log RT on correct trials using the same analytic strategy used in Experiment 1. We removed the control variables based on phonology (item duration and uniqueness point, onset voicing, and phonological neighborhood density) and added control variables based on orthography (item length in numbers of letters and orthographic neighborhood density). Item length was log transformed in order to remove skewness and minimize the effect of atypical outliers. Orthographic neighborhood density values were obtained from the English Lexicon Project (Balota et al., 2007). The distribution of density values had severe skew, so we converted it to a two-level factor: Words with at least two neighbors \((n = 54)\) formed one level, and those with zero or one neighbor formed the other \((n = 52)\). We analyzed the accuracy data separately, as in Experiment 1.

Results and discussion

The same three items discarded in Experiment 1 were discarded here. RTs were log transformed to improve normality. The results of the RT analysis can be seen in the left portion of Table 2.

Experiment 2 produced fewer significant main effects than Experiment 1, but a similar number of significant interactions. Words with higher familiarity values and those that were seen later in the task were associated with faster RTs.

Gender interacted significantly with both Danger and Ancientness. This interaction with Danger was also significant in Experiment 1, although it showed up in the error likelihoods there and the RTs here. As Figure 6 (left panel) shows, increasing Danger was associated with poorer performance for women. This matches what was seen in the error likelihoods in Figure 4 (left panel). However, whereas
men’s accuracy improved with increasing Danger in Experiment 1, here there was a negligible trend (translating to about 10 ms) in the other direction.

The right panel of Figure 6 shows that the relationship between Ancientness and RT was stronger for men than for women. In Experiment 1 the Ancientness main effect was significant, but it did not interact with gender.

The interaction between Danger and Ancientness was again significant (Figure 7, left panel). The slope relationships are similar to those seen in Experiment 1 (cf. Figure 3, left panel). There is facilitation (weaker this time) for words toward the Ancient end of the dimension, and inhibition (stronger this time) for words toward the Modern end. Our prediction for this interaction was that the Danger effect

<table>
<thead>
<tr>
<th>Variable</th>
<th>RT analysis</th>
<th>Error analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regression coefficient ($B$)</td>
<td>Standard error of $B$</td>
</tr>
<tr>
<td>Length</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Familiarity</td>
<td>−0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Trial Number</td>
<td>−0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Age of Acquisition</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>Orthographic N</td>
<td>−0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Meaningfulness</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Concreteness</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Male</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Danger</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Usefulness</td>
<td>−0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Ancientness</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Male x Danger</td>
<td>−0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Male x Usefulness</td>
<td>−0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Male x Ancientness</td>
<td>−0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>Danger x Usefulness</td>
<td>−0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>Danger x Ancientness</td>
<td>0.26</td>
<td>0.11</td>
</tr>
<tr>
<td>Usefulness x Ancientness</td>
<td>−0.20</td>
<td>0.11</td>
</tr>
<tr>
<td>Male x Danger x Usefulness</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Male x Danger x Ancientness</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>Male x Usefulness x Ancientness</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>Danger x Usefulness x Ancientness</td>
<td>0.08</td>
<td>0.36</td>
</tr>
<tr>
<td>Male x Danger x Usefulness x Ancientness</td>
<td>0.18</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$
would be stronger for more Ancient items, a prediction that was obviously not borne out. The prediction did hold in the auditory experiment (Figure 3, left), but even there the more Modern items did show a slight inhibitory trend. In fact, Wurm et al. (2009) found an interaction in auditory lexical decision times that looked almost exactly like Figure 3 (left), including this weak upward trend for more Modern items.

All of this suggests to us that the inhibitory effect for more Modern items might be replicable and thus require an explanation. Our best guess about this involves the idea noted above that the performance decrements seen in such interactions might be due to behavioral uncertainty (Wurm & Seaman, 2007). If our overall framework is correct, we can plausibly surmise that there should be less behavioral uncertainty associated with more Ancient concepts. It thus seems

![Figure 6](image1.png)  
**Figure 6.** Predicted log RT as a function of gender, log Danger ratings (left panel), and Ancientness ratings (right panel). Data are from Experiment 2 (visual lexical decision). “M” means male and “F” means female.

![Figure 7](image2.png)  
**Figure 7.** Predicted log RT as a function of Ancientness ratings, log Danger ratings (left panel), and log Usefulness ratings (right panel). Data are from Experiment 2 (visual lexical decision). “A” means more ancient words and “M” means more modern words (defined as one standard deviation below and above the mean, respectively). Ancientness was not dichotomized for the statistical analysis.
possible that our initial hypothesis was wrong and that a more promising hypothesis for future research would be that Ancient and Modern items will show different effects of Danger and Usefulness. Which effect is strongest will depend on other factors. We will return to this point below.

The interaction between Usefulness and Ancientness was also significant. As mentioned above, we have found a significant facilitative main effect of Usefulness in many previous studies (including Experiment 1 of the current study). However, as Figure 7 (right panel) shows, the Usefulness effect is significantly stronger for words with more Modern relevance. This pattern runs counter to what we originally expected, and our refined view that incorporates behavioral uncertainty does not make the right prediction here either. We do not have a compelling explanation for this particular interaction, but we believe we should be careful not to overinterpret it as it was significant only one of the four times it was assessed — it was not significant in the error analysis (below) or in either analysis from Experiment 1.

The right portion of Table 2 shows the results of the error analysis. Words that were longer or had higher familiarity values had lower error likelihoods. Words that were seen later in the task had higher error likelihoods. With the exception of length these main effects were also significant in the RT analysis.

Gender again interacted significantly with Usefulness. As can be seen in Figure 8, with increasing Usefulness both genders have a decreasing probability of making an error, but the decrease is greater for women than men. Although the effects are somewhat attenuated, this is the same pattern seen in Experiment 1 (see Figure 4, right panel).

Finally, the three-way interaction between gender, Danger, and Ancientness was again significant. This interaction is shown in Figure 9. As in Experiment 1

![Figure 8](image-url)  
**Figure 8.** Predicted probability of error as a function of gender and log Usefulness ratings. Data are from Experiment 2 (visual lexical decision). “M” means male and “F” means female.
Ancient semantic processing

Figure 9. Predicted probability of error as a function of gender, Ancientness, and log Danger ratings. Data are from Experiment 2 (visual lexical decision), and are plotted separately for men (left panel) and women (right panel). “A” means more ancient words and “M” means more modern words (defined as one standard deviation below and above the mean, respectively). Ancientness was not dichotomized for the statistical analysis.

(c.f. Figure 5), women showed a stronger effect than men on Modern words. With visual presentation, though, there was little difference between the genders on more Ancient words. This interaction highlights one more time that our original hypothesis stands in need of refinement. For men the more Ancient items showed the stronger Danger effect, but women showed an unexpectedly strong performance decrement for more Modern items as Danger rose. This decrement can perhaps be explained by the idea of behavioral uncertainty, as noted above, but why it should only apply to women is unclear.

General discussion

While the hypothesized interaction between Danger, Usefulness, and Ancientness was not seen in either experiment, the current study adds to the literature showing an interaction between Danger and Usefulness in auditory lexical decision (Wurm, 2007; Wurm & Seaman, 2008; Wurm & Vakoch, 2000; Wurm et al., 2007), with a stimulus set different from any of the previous studies. The only previous visual lexical decision study to include these variables did find the same interaction (Wurm et al., 2004b), but Experiment 2 failed to show it. It may be that differences in the time course of information uptake make the interaction much easier to detect with auditory presentation. The visual task presents a stimulus in its entirety at onset, whereas the stimulus unfolds over time in the auditory task. These differences are reflected in the mean lexical decision times (626 vs. 908 ms
for the visual and auditory experiments, respectively). Future research should investigate the possibility that such interactions are sensitive to the time-course of task-specific information processing.

Aside from this interaction, though, a large number of effects were found in both experiments. In three of four instances (Figure 3; Figure 7, left panel) the hypothesized interaction between Danger and Ancientness was significant. For words that were judged to be more Ancient, increasing Danger was associated with faster or more accurate performance. For words toward the Modern end of the dimension, performance was slower or less accurate. This finding lends support to the idea of an evolution-based fear module. Research on fear conditioning has shown that ancient threats are easier to associate with negative outcomes and are more resistant to extinction compared to modern threats (Öhman & Mineka, 2001), and images of ancient threats are quickly and automatically detected (Öhman et al., 2001). Under this account, more Ancient concepts are better able to tap into this hypothesized module.

The examples given in most evolutionary accounts such as Öhman and Mineka (2001; Öhman et al., 2001) are concepts such as snakes and spiders. We note here that our account does not distinguish between biologically-defined and non-biologically-defined concepts. The item set used in the current study is not ideal for testing this — our stimulus selection procedure was blind to this factor, and nearly three-quarters of the items we selected are non-biologically-defined. In our view the distinction should not be central, but this remains an intriguing question for future research. One could also think of differentiating between ancient things that are or are not still part of the current environment. Still other possibilities can be imagined.

Our understanding of these and previous data patterns involves the allocation of attentional resources, but in a way different from the traditional view that negative stimuli automatically command the perceiver’s attention (e.g. Pratto & John, 1991). In our view, the survival relevance of any incoming stimulus is routinely and automatically evaluated along the dimensions of Danger and Usefulness, whether it is particularly negative or not. This evaluation is carried out during processing on the basis of whatever information is available, even if it is only partial stimulus information — and it even applies to novel stimuli (Duckworth, Garcia, Bargh, & Chaiken, 2002; Song & Schwarz, 2009; Wurm, 2012). Converging evidence from a number of areas such as automatic attitude activation, affective priming, the mere exposure effect, and recent work in memory supports this conclusion.

We believe along with other grounded accounts that perception is for action. Because lexical processing is so fast, easy, and automatic, the cognitive system can already during processing allocate resources to the readying of an appropriate real-world response such as approaching or withdrawing. Stimuli that call for
both kinds of behaviors (high on both Danger and Usefulness) engage conflicting response tendencies, slowing RTs. Stimuli that are toward the Modern end of the scale and also high on Danger may also slow RTs via behavioral uncertainty, although this *post hoc* analysis will require additional research to verify.

If our description is correct, it should be possible to alter participants’ observed data patterns through the use of attentional manipulations like those found in dual-task paradigms, or by manipulating task difficulty. Indeed, Wurm and Seaman (2008) conducted two perceptual identification experiments in which spoken stimuli were embedded in noise. The participants’ task was to repeat each stimulus as quickly as possible. A noise level was chosen to produce a mean recognition probability of about 0.6, making the task quite difficult. Under these circumstances the typical Danger x Usefulness interaction reversed. Speed and accuracy were *best* for items high on both Danger and Usefulness.

A standard naming experiment, with pristine auditory stimuli in a sound-controlled environment, produced the expected interaction. Performance was worsening for items high on both Danger and Usefulness, akin to Figure 2 of the current study. Wurm and Seaman (2007) interpreted this as evidence supporting the idea that under normal listening conditions, in which word recognition is fast and almost perfectly accurate, processing resources are allocated toward planning an appropriate response. This is prevented when the highly-practiced, automatic task is made very difficult as in perceptual identification, because all processing resources are needed for stimulus disambiguation (i.e. recognizing the word).

In addition to the Danger x Ancientness interaction, the current study revealed a number of interesting interactions involving gender. The gender x Danger interaction was significant in the error analysis of Experiment 1 and the RT analysis of Experiment 2 (Figures 4 and 6, left panels), and the gender x Usefulness interaction was significant in the error analyses of both experiments. These effects may reflect the early gender roles thought to be held by our predecessors (Eagly & Wood, 1999; Silverman et al., 2007). As Danger ratings increased, women became slower (Experiment 2) or less accurate (Experiment 1), while men showed less slowing and increased accuracy. Usefulness increased accuracy for all participants, but significantly more so for women (Figure 4, right panel, and Figure 8). These differing patterns were predicted above, based on research on evolution and gender roles and the supposed hunter/gatherer division of labor. These results also fall in line with Taylor et al.’s (2000) model described above, a model that is based on the notion that men and women evolved different response systems to stress.

Some of these two-way interactions were also part of a significant three-way interaction between gender, Danger, and Ancientness. For Ancient words, men’s Danger effects were as strong as (Experiment 2) or stronger than (Experiment 1) those seen for women. For Modern words, women’s Danger effects were
consistently stronger than those seen for men. The effect was restricted to the error analyses but was found in both experiments and the effects appear largely similar (Figures 5 and 9).

The current study has not only shown that the rated Ancientness of a word can significantly affect both lexical decision RT and accuracy, but has shown how such a dimension can be used to test hypotheses on the evolution of human perception and cognition. It has also shown how Ancientness may allow for the teasing apart of gender differences in perceptual processing, supporting previous work (e.g. Taylor et al., 2000; Wurm et al., 2007). Cognitive psychology does not have a long history of examining participant gender effects, but our research suggests that gender can be an important predictor of performance in basic cognitive tasks (see also Wurm et al., 2007).

The current study also underscores the utility of analyzing error data. Researchers do not routinely do this, in part because error data often show smaller and fewer effects than RT data. With statistical analyses like the multilevel mixed-effects models used here, this need not be the case. Such models have impressive power compared to the more traditional regression models in which individual participant data are collapsed into a mean accuracy for each item (see also Wurm et al., 2007).

The current study also underscores the utility of analyzing error data. Researchers do not routinely do this, in part because error data often show smaller and fewer effects than RT data. With statistical analyses like the multilevel mixed-effects models used here, this need not be the case. Such models have impressive power compared to the more traditional regression models in which individual participant data are collapsed into a mean accuracy for each item (see also Wurm et al., 2007).

The current study represents an imperfect first attempt to study Ancientness in lexical processing, and our method produced mean ratings that were judged by experts to be at least generally suitable for the purpose at hand. These ratings relate to our dependent variables in ways that are interpretable and intriguing, and our results dovetail nicely with data from other areas of research. Nevertheless, much work remains to be done in fleshing out the Ancientness construct. How is it best conceptualized? Is it a continuous or dichotomous construct? Our view is reflected in the fact that we asked participants to use a two-point rating scale for Ancientness. However, the rating study did not produce the expected clear demarcation between concepts we should consider Ancient and those we should consider Modern. This lack of consensus among the rating participants informed our analytic strategy but whether it says anything about the proper way to conceptualize the construct is an open question. It does point to the need for an improved way of measuring Ancientness.

Wurm and Vakoch (2000) did not examine Ancientness but implicitly suggested one possible dichotomy: concepts whose existences pre-date spoken language are Ancient and other concepts are Modern. They wrote:

The idea that constructs such as Danger and Usefulness should influence the perception of spoken language may not be obvious. Language is, after all, a relative newcomer in evolutionary history. However, paying attention to the Danger and Usefulness of objects in the external world has served organisms well for millions of years. This strategy would not (and could not) be abandoned simply because
organisms stood erect and began to speak. Our position is that language developed on top of this existing, long-standing framework of survival goals and strategies (p. 187).

Such a dichotomy does not posit that the Danger and Usefulness of Modern things are irrelevant, but that the effects might differ. The strategy is Ancient, and we believe the effect of applying it to an Ancient concept might well differ from the effect of applying it to a Modern concept. This is an extreme view that seems difficult to test directly given the huge range of estimates on the timing of the emergence of spoken language.

Our instructions to the rating participants did not use the idea of “existing before spoken language.” It is unlikely that the participants would have arrived at a consensus about the items even if we had put it this way. It might prove beneficial to take an approach that makes use of formal methods from fields such as biological anthropology, historical linguistics, and evolutionary psychology.

As noted above, other refinements to the basic construct are possible, such as distinguishing between biologically-defined and non-biologically-defined concepts, or having categories for things that are “Ancient and still in existence” versus “Ancient but no longer in existence,” and so on. In addition, as noted above, our conceptualization of Ancientness takes into account only age. It says nothing about importance or significance, and future research can help determine whether it should.

Future studies on lexical processing might profitably focus on items whose classifications are more clear-cut. For example, water would have to be classified as Ancient and car would have to be classified as Modern under any scheme. Even this approach, though, is not as straightforward as it sounds and the number of concepts that can be unambiguously categorized appears to be fairly small. Small item sets raise serious concerns about statistical power (e.g., Baayen et al., 2007).

Traditional models do not have a place for variables such as Danger, Usefulness, or Ancientness. In most of them, such affective or semantic effects are ruled out a priori until after some moment of recognition. However, because related effects are now being seen in multiple paradigms (e.g. lexical decision, naming, perceptual identification, and both recall and recognition performance in memory), it is becoming apparent that what we are observing is a general principle underlying human cognition and perception.

Acknowledgement

We thank Anthony Ringuette and Natalia Tosti for help with data collection.
Notes

1. It should be noted that there is disagreement regarding calculation of $df$ for analyses involving multiple observations per participant. The value shown in most packages can be considered the upper bound for $df$ (the number of observations minus the number of model parameters), but this method can be anticonservative (Baayen, 2008a; Baayen et al., 2008). Baayen (2008b; Baayen et al., 2008) recommends an alternative method based on the posterior distributions of the model parameters. These distributions are obtained using Markov chain Monte Carlo sampling, with 10,000 samples generated. The $p$-values used to establish significant effects in the multilevel ANCOVA reported in Tables 1 and 2 are estimated based on these distributions. Only those effects that reached significance by both approaches were reported as significant in the current study.

2. Our error models were linear, with “log odds of an incorrect response” as the dependent variable. We have translated log odds into error probability on our plots for ease of interpretation. The relationship between log odds and probability is monotonic but not linear, so the plots will show some degree of curvilinearity. This becomes more pronounced as predicted probabilities asymptote toward zero.

References


**Corresponding address**

Devon Witherell  
Department of Psychology  
Wayne State University  
5057 Woodward Ave. (7th floor)  
Detroit, MI 48202  
USA  
Phone: 1-313-577-2800. Fax: 1-313-577-7636.  
dwitherell@wayne.edu
Copyright of Mental Lexicon is the property of John Benjamins Publishing Co. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.