Dimensions of Speech Perception: Semantic Associations in the Affective Lexicon

Lee H. Wurm and Douglas A. Vakoch
State University of New York at Stony Brook, USA

The affective lexicon has been explained in terms of three underlying dimensions: Evaluation, Activity, and Potency. We assessed the importance of these dimensions during online speech perception. Participants made speeded lexical decisions about emotion words that were heard in a tone of voice that was either congruent or incongruent with the word’s meaning. The denotative semantic category from which words were chosen was significantly related to lexical decision times ($P < 0.001$). Tone of voice did not influence decision times, nor did it interact with semantic category. Regression analyses showed that lexical decision times were significantly predicted by dimension weights on Potency, and by the three-way interaction between dimension weights on Evaluation, Activity, and Potency (both $Ps < 0.001$). The implications of this study for models of knowledge representation and perception are discussed.

INTRODUCTION

The connotative meanings of words can be described in terms of a small number of underlying dimensions or semantic primitives (e.g. Johnson-Laird & Oatley, 1989; Osgood & Suci, 1955). The early work of Osgood and colleagues is one important way in which this has been done (Osgood & Suci, 1955; Osgood, Suci, & Tannenbaum, 1957). From this work came the dimensions of Evaluation, Potency, and Activity. Evaluation can be thought of as a dimension ranging from “good” to “bad”, activity as ranging from “active” to “passive”, and Potency as ranging from “strong” to “weak”. These dimensions (going by various names) have been studied extensively and discovered to play a role in a variety of contexts (e.g.

Requests for reprints should be sent to Lee H. Wurm, Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11794–2500, USA.

A portion of this work was presented at the 127th meeting of the Acoustical Society of America, 9 June, 1994, in Cambridge, MA. We would like to thank Arthur Samuel for allowing this project to be completed in his laboratory. We would like to thank Annmarie Cano, Arthur Aron, and David Cross for their help with statistical analyses.

© 1996 Psychology Press, an imprint of Erlbaum (UK) Taylor & Francis Ltd
Apple & Hecht, 1982; Daly, Lancee, & Polivy, 1983; Green & Cliff, 1975; Heise, 1965; Morgan & Heise, 1988). Evaluation, Potency, and Activity have even been proposed as cross-cultural universals (Osgood, May, & Miron, 1975; but see also Russell, 1983, who believes the case is strong only for Evaluation). All three dimensions are frequently found in judgement tasks in which subjects rate the similarity of a range of emotion states (Bush, 1973; Daly et al., 1983; Morgan & Heise, 1988).

Researchers have made efforts to determine whether the outputs of cognitive processes are best understood in categorical or dimensional terms. Often, although the best explanation for performance data may at first seem to be categorical, on further study they can be explained dimensionally. This is true for such diverse phenomena as object classification (Ahn & Medin, 1992) and categorical perception of phonemes (Massaro, 1987). One advantage of using dimensional rather than categorical information is that it is computationally cheaper to arrange information dimensionally, where all that must be stored for any given item is a weight on each of a small number of dimensions. Anderson (1991) points out that each category maintained creates extra computational cost for the information-processing system (that is, the organism). This is particularly important when information processing is seen in an evolutionary context. According to evolutionary theory, characteristics that yield even a small advantage in survival are more likely to be passed on to future generations (Darwin, 1859/1968). To the extent that organisms can process information about the external environment necessary for their survival in an efficient manner, their limited resources are more readily available for other activities central to evolutionary success such as self-preservation and reproduction.

Although some models of conceptual organisation and perception have addressed the fact that seemingly categorical perceptual effects can be described in dimensional terms (e.g. Goldstone, 1994; Masarro, 1987), none has provided much detail about what those dimensions are. The three Osgoodian dimensions, which have been particularly useful for examining the affective lexicon, may be candidates. Each of them is associated with psychological and physiological responses of the organism to emotion-eliciting contexts (Scherer, 1986), and the same dimensions seem to be firmly established as early as age 8 (Russell & Ridgeway, 1983). In addition, Osgood (1969) has proposed a model in which these dimensions are important evolutionarily. It is advantageous in terms of survival, according to this view, to be able to categorise items quickly along the three dimensions: Is something good or bad? Fast or slow? Strong or weak? To the extent that this analysis holds, we would expect Evaluation, Activity, and Potency (or dimensions like them) to be hardwired into the perceptual system.
The present study investigates whether the dimensions of Evaluation, Potency, and Activity are related directly to perception, or whether they emerge only at higher levels in the cognitive system. Before describing the current study, however, it is necessary to clarify some distinctions and define some basic terms.

The *lexicon* is a special part of semantic memory. Traditionally, the lexicon is viewed as dealing only with *words*, which serve as labels for the more complex *concepts* presumed to be the main contents of semantic memory. *Lexical access* is the process by which an acoustic signal makes contact with the representation in memory corresponding to the spoken word. It is a basic perceptual process that occurs automatically whenever acoustic information sufficient for activation of such a representation reaches an attending listener.

In the context of our research, *perception* is the involuntary, low-level process whereby a physical stimulus makes initial contact with a mental representation. From the point of view of perception researchers, the question of *when* variables have an influence is as important as *whether* they do. This leads us to the important distinction between processing that is truly perceptual (called “online” processing) and that which occurs post-perceptually. In general, an effect is considered online if the basic act of perceiving a stimulus (not interpreting it or making a judgement about it) is itself influenced. Word frequency is an example of such an effect: High frequency words can be accessed more quickly than low frequency words (e.g. Forster & Chambers, 1973; Oldfield, 1966; Segui, Mehler, Frauenfelder, & Morton, 1992).

Variables that influence later decision processes (i.e. those that occur “higher up” in the system) are considered post-perceptual. Such variables have an effect only *after* the basic perceptual act has been completed. The perceptual vs. post-perceptual issue places strong constraints on models of perception, and consequently the matter is being vigorously debated.

The main purpose of the current study is to examine the relationship between perceptual processing and the dimensions of Evaluation, Potency, and Activity. The dimensional work described above has not done this, but instead has involved post-perceptual judgements (e.g. deciding to which category a particular stimulus belongs).

In addition to the question of dimensional relationships in the lexicon, we were interested in the possibility that an auditory Stroop effect could influence speech perception. Although most familiar in the visual domain, many researchers have demonstrated auditory Stroop effects as well (e.g. Cohen & Martin, 1975; Green & Barber, 1981, 1983; Palef & Nickerson, 1978; Shor, 1975; Walker & Smith, 1984, 1985, 1986). However, none of these studies has defined mismatching stimulus characteristics in terms of meaning and tone of voice.
We used emotion words for this investigation for two reasons. First, in this domain we could easily define matching and mismatching conjunctions of words’ denotative semantic categories and the tones of voice in which they were uttered, which allowed us to look for Stroop interference. Secondly, in proposing that cognitive psychologists and social cognitive psychologists do more to take notice of each others’ work, Barsalou (1990) suggests that cognitive psychologists begin to use categories for which inferences are truly important. Surely emotions are such categories.

METHOD

Subjects

Sixty-eight undergraduates from the psychology subject pool at the State University of New York at Stony Brook participated. All were native speakers of English and all had normal hearing. They received course credit for their participation.

Materials

Forty-eight words were selected from Morgan and Heise’s (1988) study of pure emotion words (emotion words that are relatively free of cognitive and behavioural connotations (Clore, Ortony, & Foss, 1987; Ortony, Clore, & Foss, 1987). Hyphenated words were not allowed. Four words were chosen from each of three maximally distinct denotative semantic categories (disgusted, petrified, and happy). The remaining 36 were randomly chosen from the full list of emotion words so that our pool of items would not be concentrated near the end-points of the dimensions.

Twenty-four of these 36 words were changed into nonwords. We did this by changing the phoneme at the point in the acoustic signal where the word in question diverges from all other words in English (the uniqueness point; see Marslen-Wilson, 1984). The phoneme at the uniqueness point was changed to a different phoneme from the same broad class (i.e. fricatives replaced fricatives, vowels replaced vowels, and so on). For example, the word “depressed” was changed to the nonword “deprussed”. The words and nonwords used in this study are listed in the Appendix.

Words from the Happy category had a higher mean frequency (43.25) than those from the other categories (3.25 for Disgusted, 21.00 for Petrified, and 10.75 for Other; frequencies are from Francis & Kučera, 1982). This poses a problem, because frequency affects lexical access speed, as we noted earlier. However, it could not be avoided in the current study because we did not have a large pool of pure emotion words from which to choose a frequency-balanced subset.
The location of each of the 24 stimuli (words) in three-dimensional connotative lexical space was specified by the dimension weights for each word on Evaluation, Potency, and Activity, which were reported by Morgan and Heise (1988), averaged across male and female subjects. Mean weights were −0.76 for Evaluation (SD = 2.4), −0.51 for Potency (SD = 2.0), and 0.52 for Activity (SD = 1.4).

Each stimulus item was read at the end of the carrier phrase: “When that happened, I felt ______.” The carrier phrase provided a natural and fluent prosodic context while keeping the potential effects of semantic context to a minimum. The carrier phrase was read four times for each stimulus item, once in each of the following tones of voice: happy, petrified, disgusted, and neutral. Tones of voice were produced according to specifications of two perceptual parameters found to differentiate the four tones of voice across numerous past studies: speech rate and average pitch (Murray & Arnott, 1993). In total, there were 192 stimulus sentences, each of which was digitised at a sampling rate of 10kHz (low-pass filtered at 4.8 kHz) and stored in a disc file.

Manipulation Check

To determine whether the tone-of-voice manipulation actually worked, we conducted a rating study. Fifteen undergraduates at the State University of New York at Stony Brook listened to the 192 stimulus sentences, presented in a random order. The word or nonword that ended each sentence was digitally stripped from the carrier phrase, so that the tone-of-voice ratings would not be influenced by word meanings. Subjects were asked to give each sentence a rating from 1 to 7 on each of three dimensions, with anchor points corresponding to the following extremes: slow vs. fast rate of presentation; low vs. high pitch; and good vs. bad. These dimensions correspond to the two perceptual parameters by which the stimuli were modelled (speech rate and average pitch) as well as the Evaluation dimension.

The results of this rating study are shown in Table 1. Consistent with Murray and Arnott’s (1993) review of previous studies characterising these four tones of voice, compared to the neutral tone of voice: (1) disgust was slower and lower pitched; (2) petrified was very much faster and very much higher pitched; and (3) happy was faster and higher pitched. Moreover, petrified was rated as being more negative on the Evaluation dimension than either happy or neutral ($P < 0.005$ for all comparisons). We can thus be reasonably sure that the tones of voice were perceived as intended.
Procedure

In order to avoid the possible effects of repetition priming, the 192 stimulus sentences were divided into four different lists of 48 sentences. Each word was presented in exactly one tone of voice per list. For example, the word “scared” was heard in a happy tone of voice in one list, in a petrified tone of voice in another list, and so on. Participants were randomly assigned to hear one of the four lists of stimuli. Within each list the order of presentation was randomised. Each group of participants got a different random order.

Within each list, 25% of the stimuli were from each of the four tones of voice. Across the entire experiment, each word was heard in each tone of voice an equal number of times. For each participant, half of the items were words and half were nonwords.

Participants were directed to make a speeded lexical decision about the item in sentence-final position. Participants used their dominant hands to make responses on a button board, pushing one button for words and another for nonwords. Reaction times (RTs) were measured from the onset of the phoneme at the uniqueness point of each word. Participants were tested in groups of one to four in a sound-attenuating chamber. Digitised speech files were played for the participants over headphones at a comfortable listening level.

RESULTS

Subjects who had error rates greater than 0.25 or mean RTs greater than 1000msec were excluded from our analyses. Four subjects’ data were excluded by these criteria. For the remaining 64 subjects, data were dis-
carded for trials on which the word/nonword decision was made incorrectly (6.8% of all trials).\textsuperscript{1}

The RT data were not normally distributed, so we performed a square root transformation. The dependent variable in all reported analyses was the transformed RTs on correct trials (word targets only).

**Stroop Analysis**

Our first analysis was a repeated-measures analysis of variance (ANOVA), with the denotative semantic category of the words and the tone-of-voice category serving as the factors.

For the Disgusted, Petrified, Happy, and Other semantic categories, mean RTs were 551, 535, 378, and 484 msec, respectively (corresponding SEMs were 17.84, 17.57, 17.10, and 14.07). A repeated-measures ANOVA was performed, and revealed a significant effect of semantic category \([F(3,54) = 62.73, \ MSE = 6.25, \ P < 0.001]\). Subjects responded more quickly to words belonging to the Happy semantic category than those in the other three categories, by an average of 145 msec. As we have noted, this could have been due to the higher mean word frequency of those words. Other effects in the analysis were not reliable \((F < 1.0 \text{ for tone of voice and for the interaction})\). Hearing a word spoken in a tone of voice that is incongruent with its denotative meaning did not produce evidence for any Stroop-type interference in RTs.\textsuperscript{2}

**Regression Analysis**

Because the main analysis did not reveal any evidence of a tone-of-voice or of Stroop-type interference, we excluded the tone-of-voice factor from the second analysis. (The pattern of results for this analysis was the same

\textsuperscript{1} Error rates did not differ across the four tones of voice \([F(3,189) = 2.04, \ MSE = 0.009, \ P > 0.10]\), but there was a significant effect of semantic category \([F(3,189) = 34.08, \ MSE = 0.009, \ P < 0.001]\). Disgusted words had a higher error rate (16.8%) than the other three categories (1.6%, 3.9%, and 4.2% for Happy, Petrified, and Other, respectively).

\textsuperscript{2} One possible reason we did not observe a Stroop effect has to do with stimulus onset asynchrony (SOA), which is the difference in time of onset between the two characteristics of the stimulus under observation. Previous authors have noted that the basic Stroop phenomenon seems most robust with SOAs in the 100 msec range (Glaser & Glaser, 1982; McLeod, 1991). In our stimulus sentences, the onset of the tone-of-voice information always preceded the onset of the semantic information by at least 800 msec (and often by as much as 2500 msec in the slower sentences.)
whether tone of voice was included or not. Exclusion of the factor was done only in the interest of simplification.)

The second analysis was a supplementary hierarchical regression analysis, in which the connotative semantic content of the words was coded continuously in terms of dimension weights on Evaluation, Potency, and Activity (described earlier). This analysis provided a detailed investigation of the roles of each of the three dimensions, as well as their possible interactions. In addition, this analysis allowed us to control statistically for word frequency effects (described in Materials).

Regressors were entered in blocks, in the following order: (1) log word frequency and subjects; (2) dimension weights on Evaluation, Activity, and Potency; (3) the 2-way dimension weight interaction terms (i.e. Evaluation × Potency, Evaluation × Activity, and Activity × Potency); and (4) the three-way dimension weight interaction term. The results of this preliminary regression analysis revealed three suppressor variables: dimension weight on Evaluation; the interaction between dimension weights on Evaluation and Potency; and the interaction between dimension weights on Evaluation and Activity. Suppressor variables are those for which the direction (i.e. sign) of the regression coefficient is different from the sign of that variable’s bivariate correlation with the dependent measure. For example, the bivariate correlation between dimension weight on Evaluation and RT was significant and negative, whereas the regression coefficient for Evaluation was significant and positive. Because the effects of suppressor variables cannot be meaningfully interpreted, and because they enhance the effects of variables entered into the equation after them (Tabachnick & Fidell, 1989), these three variables were removed from the analysis. A second regression analysis was then run without them.3

The first step was to statistically remove (i.e. partial out) the effects of log word frequency and individual subjects’ variance [for word frequency, $\beta = -0.19$, $F(1,1367) = 72.42$, $MSE = 20.03$, $P < 0.001$]. After this was done, we found a significant increase in $R^2$ when we added words’ dimension weights on Potency and Activity [for these two regressors, $F(2,1365) = 41.29$, $MSE = 18.92$, $P < 0.001$]. Potency had a significant and negative relationship with RTs ($\beta = -0.22$, $P < 0.001$): Words with higher weights on Potency had faster lexical decision times. Dimension weight on Activity did not account for a significant proportion of the variance ($\beta = 0.03$, $P > 0.10$).

---

3 There is no wholly satisfactory way to deal with suppressor variables. However, because suppressors result in better predictive power for the regression model (Darlington, 1990), we opted for a conservative approach and excluded them. The only major difference between the analyses with and without suppressors was that the Activity × Potency interaction was significant with suppressors present, but not when they were removed.
The next regressor entered was the term for the two-way dimension weight interaction (Activity \times Potency). This interaction was not significant \([\beta = -0.04, F(1,1364) = 1.54, MSE = 18.91, P > 0.10]\).

Finally, we found that the three-way interaction between dimension weights on Evaluation, Activity, and Potency was significant \([\beta = 0.56, F(1,1363) = 181.51, MSE = 16.70, P < 0.001]\). This interaction can best be understood by examining Fig. 1. The figure shows mean RTs as a function of weights on the three dimensions. High and low values on each dimension shown in the figure were determined by median splits. The resulting dichotomies are used for illustration and were not used in the regression analysis, but they illustrate the general pattern of results.

**DISCUSSION**

This study shows that even at the very earliest stages of perception, the basic units that allow entry into semantic memory are themselves coded according to continuous dimensions of meaning. Furthermore, the current study demonstrates that the connotative semantic content of words allows us to predict the time course of online speech processing more accurately. Even after partialling out word frequency effects, we found that dimension weight on Potency was a significant predictor of RT. The higher a word’s weight on Potency, the more quickly it was accessed. Of course, any interpretation of the main effect of Potency must be qualified by a consideration of the significant three-way interaction between dimension weights, shown in Fig. 1. This interaction was very strong, accounting for more than twice as much variance as log word frequency, which was given first access to the variance.

One interpretation of the interaction of the three dimensions, based on an evolutionary model of the lexicon (Osgood, 1969), is offered here. Consider first the High Potency words (the top panel in Fig. 1). For High Evaluation words (i.e. those with a positive valence), Activity is not of any particular relevance. Although the objects referred to by such words are powerful, they do not represent a danger because they are pleasant. Thus, we would not expect their Activity level to matter much. If Evaluation is negative, however, the distinction between high and low Activity is absolutely critical. Survival itself could depend on the ability to make very fast decisions about things that pose an immediate threat (i.e. that are very bad, very strong, and very fast). Rapid processing is not required for things that are bad, strong, and *slow*.

An evolutionary interpretation of the bottom panel in Fig. 1, which shows the Low Potency words, is somewhat less clear. Because all of the words shown in this panel denote weak or ineffectual things, Activity should not play a particularly relevant role. We might expect RTs for Low
FIG. 1. Mean RT as a function of dimension weight on Evaluation, Activity, and Potency, in msec. The top panel shows High Potency words and the bottom panel shows Low Potency words. Error bars show +/- 1 SEM.
Evaluation words in this panel to be faster than those for High Evaluation words, and they are to a slight extent. However, because of the way it is organised, this area of connotative lexical space must be interpreted with caution. It turns out that very few words have high weights on Evaluation and low weights on Potency; ineffectual or weak words are generally not pleasant, whereas strong words generally are. Overall, we can see that mean RTs for Low Potency words are slower than for High Potency words, which again fits with an evolutionary perspective.

The relative paucity of weak, good words is in accord with Morgan and Heise’s (1988) conclusion that Potency is an important dimension primarily for distinguishing between negative emotions. Morgan and Heise (1988) explain the importance of Potency in terms of its adaptiveness for the organism. Thus, an organism must decide in a hedonically negative situation whether it is being threatened, and if so, whether to respond with “fight” or “flight”.

Models that stress the role of affective-motivational states in cognitive development provide an account that is consistent with the evolutionary account given earlier (Izard & Malatesta, 1987). From the time we are children, we hear certain words more often uttered in an emotion-laden context (e.g. “poison”). This covariation is a valuable source of information for listeners, and our data show that it is coded in the lexicon. Studies of children’s emotion concepts indicate that Evaluation and Activity are important dimensions for post-perceptual classification of emotions as early as third grade in elementary school (Russell & Ridgeway, 1983). The time-scale within which these dimensions become important for children in organising their knowledge of the world must be investigated more thoroughly before any final conclusions can be drawn about the relative contributions of ontogenetic and phylogenetic models in accounting for the origins of the dimensional structure of the lexicon.

Bower (1987) concluded that perception is only weakly affected (if at all) by emotion, and only if the perceiver has a long-standing emotional condition (e.g. depression or an anxiety disorder). The current study suggests a more pervasive impact of the connotative meaning of emotion words, even for a population of normal subjects (i.e. subjects not chosen on the basis of having long-standing emotional states of one kind or another).

The current study also has implications for models of speech recognition. Existing models have not addressed the possible role of connotative meaning, but our work shows that connotative meaning is relevant at the earliest stages of perception. In general, network models of perception (e.g. McClelland & Elman, 1986) could easily be modified to accommodate our results; presumably the resting activation levels of words could reflect the words’ Evaluation, Potency, and Activity dimension weights, or,
alternatively, the connection weights between units in the network could be altered to reflect the dimension weights.

This study reveals some of the complexity of lexical representations. We found that lexical access time for any given emotion word is dependent, in part, on the characteristics of a larger class of words known to the listener. We also found a previously unknown link: Dimensions derived post-perceptually are also significantly related to online perception.

The framework used here allows us to locate easily any particular word relative to other words on the basis of three dimensions of connotative meaning. There is currently no similarly comprehensive and yet parsimonious system for categorising the denotative content of words. The dimensions of Evaluation, Potency, and Activity have a fairly obvious relationship to emotion words; that, in fact, is why we chose to study the particular words we did. The extent to which our results will generalise beyond emotion words is not immediately obvious, but given that all words can be located within the connotative meaning space that we examined (e.g. Osgood et al., 1975), it seems plausible that access to words in the general lexicon can be described in terms of the same dimensions. If this is the case, then the construct of a specifically affective lexicon could be called into question. Future studies will be required to determine whether these dimensions have comparable relevance when we consider words more representative of the entire lexicon.

The fact that lexical access can be influenced by word meanings has been demonstrated previously using various sorts of priming tasks, where access time is facilitated after prior presentation of a word related in meaning (e.g. Warren, 1977). The present finding is particularly important because we have not used a priming paradigm; the connotative semantic effects we identify are coded directly into the mental representations of words themselves and exist independently of priming. Thus, the present study provides a novel approach to uncovering the structure of semantic mental representations. The same approach would be applicable to examining other dimensions that could influence lexical access.

Manuscript received 27 March 1995
Revised manuscript received 4 October 1995

REFERENCES


APPENDIX

Stimuli

Words
Disgusted: disgusted, annoyed, displeased, irked
Petrified: petrified, terrified, afraid, scared
Happy: happy, pleased, glad, contented
Other: empty, shaken, outraged, delighted, passionate, unhappy, apprehensive, resentful,
    agitated, overjoyed, fearful, deflated

Nonwords
melantoly, lonefome, deprussed, sickemed, eloted, emballassed, micherable, exsouted,
lonery, cheerlace, reliejed, downhorted, joywess, regletful, hurk, jealoos, frightened,
lovezhick, prud, mab, irete, charned, overyelmed, horrisied