

Extensions of the Paivio, Yuille, and Madigan (1968) norms

JAMES M. CLARK

University of Winnipeg, Winnipeg, Manitoba, Canada

and

ALLAN PAIVIO

University of Western Ontario, London, Ontario, Canada

The Paivio, Yuille, and Madigan (1968) norms for 925 nouns were extended in two ways. The first extension involved the collecting of a much more extensive and diverse set of properties from original ratings and other sources. Factor analysis of 32 properties identified 9 orthogonal factors and demonstrated both the redundancy among various measures and the tendency for some attributes (e.g., age of acquisition) to load on multiple factors. The second extension collected basic ratings of imagery, familiarity, and a new age of acquisition measure for a larger pool of 2,311 words, including parts of speech other than nouns. The analysis of these ratings and supplementary statistics computed for the words (e.g., number of syllables, Kučera–Francis frequency) demonstrated again the relative independence of various measures and the importance of obtaining diverse properties for such norms. Implications and directions for future research are considered. The full set of new norms may be downloaded from www.psychonomic.org/archive/.

Thirty-five years ago, Paivio, Yuille, and Madigan (1968) (hereafter, PYM) published a set of word norms that became a widely used resource in the field of cognitive psychology. The PYM norms included imagery ratings, concreteness ratings, and meaningfulness values for 925 words (all nouns), with familiarity ratings also available at a later date. In addition, frequency values were provided for the same items from the Thorndike–Lorge (Thorndike & Lorge, 1944) and Kučera–Francis (Kučera & Francis, 1967) norms, and such physical properties as word length and number of syllables were readily computed. Many subsequent researchers, up to the present day, have selected materials for their cognitive studies from the PYM words, or from a larger unpublished set of words made available by Paivio (1974).

A statistic reported by Rubin (1981) clearly demonstrates the usefulness of these norms. Specifically, the PYM norms were cited 350 times in the 5-year period 1977–1981. The general importance of such stimulus norms for cognitive psychology is similarly clear from

an analysis by Proctor and Vu (1999). They reported that Psychonomic Society journals alone had published a total of 142 norms and cited considerably earlier reviews (not limited to Psychonomic Society journals) reporting equally large numbers of norms. The extent of experimenter and publisher resources devoted to norms is testament to their value in furthering our understanding of human cognition.

Methodologically, the study of such item attributes can be conceptualized as one of three general approaches to the understanding of human cognition, with the other two being experimental manipulations and individual differences among subjects (Paivio, 1971). Of particular theoretical interest are possible interactions between the different classes of investigation. For example, the effects of an attribute such as concreteness may vary both for participants who differ in their self-reported use of imagery across diverse tasks (an individual difference factor) and for different instructional conditions (an experimental factor). The robustness of this multipronged approach has been clearly demonstrated (Paivio, 1971, 1986).

Empirically, imagery and concreteness have been repeatedly demonstrated to be potent variables in a wide range of psychological tasks, including learning and memory tasks (e.g., Paivio, 1969), semantic retrieval tasks (e.g., Paivio, Clark, Digdon, & Bons, 1989), and diverse other cognitive tasks (for reviews, see Paivio, 1971, 1986, 1991). Although the empirical effects are well founded, their theoretical basis remains controversial, as researchers advanced other correlated properties such as context availability (Schwanenflugel, 1991; Schwanenflugel &

Various aspects of this work were supported by research grants from the Natural Sciences and Engineering Research Council of Canada. We thank a number of students and research assistants who contributed to this work over the years, most notably Trudy Bons and Jo Unger. We also thank Michael Cortese and David Rubin for comments on an earlier version of this article, and John Benjafield, David Rubin, and the American Psychological Association for permission to reproduce some of the normative values included in the databases. Correspondence should be addressed to J. Clark, Department of Psychology, University of Winnipeg, 515 Portage Avenue, Winnipeg, MB, R3B 2E9, Canada (e-mail: clark@uwinnipeg.ca).

Shoben, 1983), and as theoretical critics questioned the proposed nature of imagery (e.g., Pylyshyn, 1973).

Research on item attributes continues to become more sophisticated in a variety of ways, including consideration of larger numbers of properties (e.g., Paivio et al., 1989; Rubin, 1980), and serious efforts to model and simulate the effects of stimulus attributes (e.g., Ellis & Lambon Ralph, 2000). Computers have played a central role in both of these developments. The multivariate and simulation studies that might ultimately contribute to the emergence of theoretical models with strong empirical foundations has increased demand for large item pools and increased numbers of properties.

The need for a greater number of properties and items follows from the essentially nonexperimental nature of item attribute research. Although researchers may experimentally assign different types of materials to different subjects, they generally do not and in some cases cannot experimentally manipulate the property or properties of interest (although such experimental approaches have been used in some cases). Rather, the properties are generally measured in some fashion, and these measures will invariably correlate with other properties that might produce spurious effects or mask the effects of the target attributes. The primary way to address this problem, as in any nonexperimental research, is to identify diverse potentially contaminating constructs, obtain reliable and valid measures, and either control them in the selection of materials or include them in statistical analyses that can accommodate correlated factors (e.g., multiple regression, factor analysis, structural equation modeling).

In addition to serving this control function, the collection of a large number of properties can itself provide information useful in the conceptualization of various item attributes. We can illustrate with one controversial question—the relative importance of frequency and age of acquisition in picture naming and other semantic retrieval tasks (Morrison & Ellis, 2000). Paivio et al. (1989) obtained picture naming and imagery latencies for a moderate-sized pool of pictures and their most common labels. Information on a wide range of properties, including age of acquisition, was also obtained. Factor analysis of the results indicated that age of acquisition loaded on several different factors (e.g., familiarity, concreteness, name length), all of which contributed to picture naming latencies. One interpretation of this result is that age of acquisition is a multidimensional measure that taps a number of distinct properties of words and pictures, hence its superiority to single-component predictors in multiple regression analyses. More speculatively, one might hypothesize that people rating age of acquisition are actually making judgments of how concrete, short, and familiar items are, and that children in fact first learn words that tend to be concrete, short, and familiar.

Generalization across items provides yet another reason to continue the development of item norms for use in cognitive research and theorizing. No single set of norms

will ever suffice, because results can depend on the particular pool of items that have been included in the norms. Despite the large corpus of materials on which the Kučera and Francis (1967) frequency norms are based, for example, abstract words are still probably overrepresented just because of the types of text that dominate the corpus (e.g., literary and academic materials). It is therefore important to continue to develop additional norms to permit evaluation of the generality of findings across diverse word pools.

Another facet of the generalization issue is the possibility of generational or cohort differences across extended periods of time. With respect to word familiarity, for example, exposure to and knowledge of particular words might differ today from ratings, like those in the PYM norms, collected during the 1960s. New norms and replication of existing properties allow researchers to determine the continuing validity of norms collected years and in some cases decades ago.

This article reports two extensions of the PYM norms. Part 1 reports a marked expansion of the number of properties available for the original 925 PYM items, and Part 2 reports an expansion of the number of items for which basic properties are available. We also provide results of factor analyses for both extensions, with the analysis in Part 1 being particularly informative about interrelationships among a diverse collection of word properties.

PART 1 Expanded Property Set

The set of properties covered in the original PYM norms included word concreteness, imageability, and meaningfulness. Norms were soon available for ratings of word familiarity as well. Over the years, a number of researchers have reported additional attributes for the core PYM items, several of which are included in the following analyses.

In selecting additional attributes to measure, our goal was to build on one systematic conceptualization of item attributes that emerged from Paivio's (1971, 1986) dual-coding theory framework and was applied to good effect in the Paivio et al. (1989) picture naming–imaging study. There we found that different measures coalesced in a reasonable way to suggest specific phases of processing (e.g., picture or word identification, image or name activation, response production), and that these different phases mapped nicely onto what Paivio (1971) has called *representational*, *referential*, and *associative levels of processing*.

Representational processing involves the activation of internal verbal and imaginal representations by external stimuli, although similar processes may contribute to the end stages of internal activation via associative and referential processes. Prototypical tasks would be word and picture identification. Referential processes involve activation of representations of the other type (e.g., verbal–imaginal

and imaginal-verbal). Prototypical tasks would be imagining to words and naming pictures, although representational processes would contribute to stimulus identification. Associative processes involve the activation of representations of the same type (e.g., verbal-verbal associations and imaginal-imaginal associations). A prototypical task to assess verbal associative processing would be a word association task, although again representational processes would contribute to identification of stimulus words. Other tasks (e.g., defining words) would implicate all three types of processing to some degree, perhaps varying as a function of word concreteness.

Given verbal stimuli, as in the present norming studies, this model suggests at least three phases of processing: representational activation of the words, associative activation of other words, and referential activation of imaginal representations. Distinct properties could contribute to each stage; for example, word identification processes might be influenced by word length, familiarity, spelling, and number of neighbors, and imaging processes by concreteness, familiarity with the word and object, and image complexity.

This dual-coding perspective and existing research, such as the Paivio et al. (1989) work on naming and imaging, provided the general framework for the selection and generation of new measures for the existing PYM norms. The orientation differed somewhat from the prior work in that there was no specific criterion task in mind (i.e., no equivalent to naming and imaging latencies).

One goal was to obtain more diverse measures of physical attributes that could affect item processing, either at the identification or production stage of processing. Toward that end, we sought to expand on such traditional measures as length in letters (Len) and syllables (Syl). Separate groups of subjects were asked to estimate the number of words that began with the same initial letters as the target word (InLet), the number of similar appearing words (SimAp), the number of words that began with the same initial sound (InSou), the number of similar sounding words (SimSou), and the number of rhyming words (Rhy). We hypothesized that several of these measures would reflect the number of competing representations from which the target must be chosen, analogous to the orthographic neighborhood effects that have been studied mostly with short words (but see Rubin, 1981). As a more aggregate measure of word-level processing, we also asked participants to rate the pronounceability (Pron) of the words.

A final measure in this representational processing category was related to the Let measure. Specifically, we calculated the number of words beginning with the same first two letters of each word in the norms, using just the top (in terms of frequency) 5,001 words of the Brown (1984) corpus (LetBr). We chose the top 5,001 words as an approximate way to emphasize readily available words in semantic memory, on the assumption that most of the

competition at this level would arise from relatively accessible words.

A second major category concerned familiarity and related constructs. This domain is not so clearly restricted to a specific level of processing in the dual-coding scheme, inasmuch as familiarity could impact representational, referential, and associative processes. It is probably difficult, for example, to identify very unfamiliar words, and to generate either meaningful semantic associates or images for such words. Familiarity (Fam) ratings were already available for the PYM items (see Part 2 for a description), as were Kućera-Francis and Thorndike-Lorge frequencies (FrqKF, FrqTL). In addition, we were able to obtain replication familiarity ratings collected more recently (Fam2).

We also included several published measures of word availability. Availability is the ease with which words are brought to mind by other words or contexts. This construct can be measured by how often words are used, for example, to define other words. Two measures of availability were included in a factor analysis reported in the results. Rubin and Friendly (1986) reported availability measures for the PYM items from the Kiss, Armstrong, and Milroy (1972) norms (AvlRF). Keenan and Benjafield (1994) determined *availability* values for the PYM items on the basis of the frequency with which each word was used to define words in the *Oxford English Dictionary* (AvlKB).

Partly overlapping with the construct of familiarity and frequency are norms related to the age of acquisition of words—that is, the age at which words were learned. This construct is measured in two ways in the present norms. First, we obtained ratings for age of acquisition (Age) as one indicator of this factor. Second, Clark (2002) has recently completed a dictionary project to generate a proxy age of acquisition measure for a large set of words. Values for the number of children's dictionaries in which a word appears (maximum = 50) were obtained for each of the PYM items (Dict).

In the associative and referential processing domains of semantic memory, the original PYM norms included imagery (Img), concreteness (Con), and meaningfulness (Mng). We were able to replicate and update the imagery ratings with more recent unpublished norms (Img2). We also expanded the semantic set in several ways, partly to address questions about the causal role of imagery versus other semantic factors. Schwanenflugel (1991) and her colleagues have proposed that context availability accounts for many of the effects that have been attributed to concreteness or imagery. Accordingly, ratings of context availability (ConAv) were obtained. Other participants rated the number of associates that came to mind (Assoc), an estimate of the richness of the semantic network evoked by each word. We also had participants rate ease of definition (Def) as one index of the overall accessibility of the meaning of the word in semantic memory. Since some letter strings are ambiguous with respect to meaning (i.e., they are homonyms or ho-

mographs), another group of participants rated the 925 words for number of meanings (Amb).

The affective domain has a long history with respect to word attributes, dating back to Osgood's pioneering research, and certainly was critical to much of the early clinical interest in word association and related tasks. Goodness ratings (Gdn) were obtained from Rubin and Friendly (1986). We collected new ratings of the pleasantness (Pls) and emotionality (Emo) of the 925 words. In addition to direct ratings of emotionality, we derived two additional measures from the goodness and pleasantness ratings by calculating the absolute deviation from the average of each rating (EmoGd and EmoPl, respectively). These measures are based on the rationale that any deviation from the center of the pleasantness or goodness dimension represents an increase in emotionality, an assumption that we were able to test.

Finally, as part of a thesis under the supervision of the first author, norms had been obtained from two groups of participants on how gender laden the terms were; that is, to what extent words could be viewed as masculine at one pole and feminine at the other (Gend, Gend2). Although the research contexts in which such ratings might be of interest could be rather limited, they were included to determine whether other attributes might be confounded with this factor. For example, would more negative words tend to be masculine and more positive words feminine?

In summary, a total of 32 attributes were collected and analyzed for their underlying factor structure. The norms themselves contain values for many of these properties, with the rest being available from other sources.

Method

Materials. For the large majority of the rating tasks, the 925 PYM items were divided into three sets of 308, 308, and 309 words. An additional 60 common items were included in all of the lists. The words were printed in booklets using four distinct random orders for each set.

Procedures. All 17 rating tasks newly reported here were completed using a 7-point scale, with 1 indicating *low scores*, and 7, *high scores*. For 13 of the rating tasks, groups of varying sizes read the instructions for their particular task (see Appendix A for the instructions), and then rated the 368 or 369 words in their list. A total of 24 participants rated each word for each property. For the other 3 tasks (imagery replication, familiarity replication, and gender ladenness), three groups of 31 participants each rated all 925 words. In a smaller pilot study, an additional 10 participants completed the gender ladenness task. Instructions were similar but more extensive than those presented in Appendix A (e.g., recommending use of entire scale of values). Imagery instructions were similar to those of Paivio et al. (1968) and mentioned that the ease and speed of mental imagery was the central criterion, and that images could occur in various modalities. The familiarity instructions asked participants to rate their overall familiarity with the word across various contexts. Gender ladenness instructions asked participants to rate the words for how masculine (1) or feminine (7) the words were, with a neutral, gender-free value representing the midpoint of the scale (4).

Nine other properties were obtained from published norms: the PYM norms (imagery, concreteness, meaningfulness), the Paivio (1974) norms discussed more fully in Part 2 (familiarity), the Rubin and Friendly (1986) norms (RB availability, goodness), the Keenan

and Benjafield (1994) norms (KB availability), and published frequency norms (KF frequency, TL frequency).

From the Brown (1984) corpus, we obtained the most frequently occurring 5,001 words and determined the frequency with which words began with each of the 220 two-letter combinations that occurred for those words. The average number of words per letter combination was 22.73, ranging from a number of combinations with a single entry to 234 words for the "CO" combination. Each PYM item beginning with those letters was assigned the corresponding frequency. Word length in letters was computed directly from the word entry in the database and number of syllables was also determined and confirmed by several human judges.

From the goodness and pleasantness ratings, we computed two new emotionality measures (emotionality–goodness, emotionality–pleasantness) by determining the absolute deviation of each rating from the mean rating ($M_s = 4.0831$ and 3.5945 for goodness and pleasantness, respectively). Words receiving high scores on these derived scales would be either low on goodness–pleasantness or high on goodness–pleasantness. Words receiving low scores would have received average (i.e., neutral) ratings on goodness–pleasantness. The legitimacy of this procedure was clear from correlations reported later and a plot of the direct ratings of emotionality against the goodness and pleasantness ratings; specifically, the plots were V-shaped.

Clark (2002) entered the words from 50 children's English-language dictionaries into a computer file and standardized the spellings. There were 7,539 distinct words in the final corpus, and the age of acquisition proxy was the number of dictionaries in which each word occurred. Corrected split-half reliability was .960, and the norms have been shown to correlate well with a number of published age of acquisition measures (Clark, 2002).

For the count variables with skewed distributions, we used the logarithm to the base 10 of the original scores, with one added to the original scores to avoid the undefined $\log(0)$. The properties transformed in this way were: Dict, FrqKF, FrqTL, AvlKB, AvlRF, and LetBr.

Participants. In total, 1,039 participants completed the rating tasks satisfactorily. Most participants were introductory psychology students who participated as a course requirement for which other options were also available.

Results and Discussion

We will first examine statistics related to the individual properties, especially those that are newly reported here. Subsequent analyses examined the factor structure of the 32 properties.

Internal consistency. Table 1 summarizes descriptive statistics and (where new) reliabilities for the 32 properties. The observed reliabilities (Cronbach's alpha) measure internal consistency and are excellent with a few exceptions. People appear to judge many of the properties in a highly consistent manner. The lowest reliability of .579 is for judging the number of words with similar initial sounds (InSou). Although this may reflect in part the modest uncertainty in English between printed letters and spoken sounds, this explanation appears strained by the much better consistency for judging the number of similar sounding words (.897).

The second lowest reliability, .695, is for ratings of the number of meanings (Amb), a semantic property. This scale also has the lowest mean, however, perhaps reflecting the relative shortage of homonyms even in such a large word pool. This lack of variation could limit the expected value for correlational measures of reliability.

Table 1
Descriptive Statistics and Reliabilities for Expanded Properties

Scale	Mean	SD	Alpha	Acronym
Age of acquisition	4.34	1.18	.941	Age
Number of associates	3.50	.72	.783	Assoc
Context availability	5.99	1.09	.936	ConAv
Ease of definition	5.14	.88	.869	Def
Number of meanings	2.42	.52	.695	Amb
Pronounceability	6.37	.56	.855	Pron
Same initial letters	4.05	.64	.773	Let
Similar looking words	3.02	.64	.777	Look
Same initial sound	3.80	.52	.579	InSnd
Similar sounding words	3.05	.93	.897	Sou
Number rhyming words	3.45	.69	.791	Rhy
Emotionality	3.39	1.05	.886	Emo
Pleasantness	3.59	1.23	.941	Pls
Emotionality–pleasantness	1.01	.70		EmoPl
Familiarity replication	5.97	1.00	.958	Fam2
Imagery replication	4.58	1.09	.918	Img2
Gender ladenness	3.74	.91	.938	Gend
Gender replication	3.69	.99		Gend2
Log children's dictionaries	.52	.58	.960	Dict
Imagery	4.97	1.39		Img
Concreteness	4.95	1.88		Con
Meaningfulness	5.89	1.10		Mng
Familiarity	5.05	1.13		Fam
Goodness	4.08	1.09		Gdn
Emotionality–goodness	.85	.68		EmoGd
Log R&F availability	1.41	.90		AvlRF
Log K&B availability	2.15	.81		AvlKB
Log L2 number	1.58	.412		LetBr
Length in letters	6.91	2.13		Len
Number of syllables	2.30	1.00		Syl
Log K&F frequency	1.13	.69		FrqKF
Log T&L frequency	1.20	.60		FrqTL

It would be informative to identify true homonyms in the PYM items to confirm and complement these ratings.

Four other ratings fall just below the .80 criterion sometimes used as a standard for internal consistency. Two (perhaps three) of these were ratings for physical attributes: number of words with same initial letters (Let), number of similar looking words (SimAp), and number of rhyming words (Rhy). The fourth scale with moderate reliability was number of associates (Assoc), a semantic dimension.

The preceding six scales, as well as others having stronger reliability, would benefit from the development of additional measures of the underlying constructs. These new scales might involve ratings and direct physical measures, such as the number of initial letters here derived from the top 5,001 words in the Brown corpus. To illustrate, number of associations might be measured objectively from free-association norms for the PYM items. The PYM words could also be classified as homonyms or not, either from existing norms or by judges.

Between-norms consistency. In addition to internal consistency, which measures consistency across different raters at a single point in time, we can measure the reliability (or consistency) between ratings obtained from different raters at different points in time. The *Img2* values obtained in 2001–2002 correlated .803 with the original PYM *Img* ratings, despite the intervening 30 years.

Similarly, *Fam2* correlated .852 with the corresponding Paivio (1974) ratings. Despite this high correlation, there was a systematic and interesting tendency for the original familiarity ratings to be somewhat lower than more current ratings, until the very highest levels were reached. The two measures of availability, *AvlKB* and *AvlRF*, correlated .742.

Validities within this property set. The overall factor structure of this item set provides information about the construct validity of our measures, but we first report some basic correlations between scales for which we would clearly predict convergent validity (the entire correlation matrix is presented in Appendix B). The expected correspondences were clearly observed, with one exception. The correlations were $-.818$ between *Age* and *Dict*, $.895$ between *Pls* and *Gdn*, and $.649$ between *Let* and *LetBr*. *Emo* correlated $.691$ with *EmoGd* and $.654$ with *EmoPl*, with the latter two correlating $.771$. The one exception to the strong convergent validities was a low correlation of $.199$ between *Assoc* and *Mng*. Clearly these measures do not tap the same underlying construct.

Factor structure of the expanded properties set. The 32 variables were subjected to a principal components factor analysis with varimax rotation. Seven factors had eigenvalues greater than 1.00, but the eigenvalue for an 8th factor was also very close to 1.00 (.997) and quite close to the eigenvalue for the 2nd factor (1.133). Several of the variables also had little of their variation explained by the 7- or 8-factor solutions. Moreover, extensive simulations of factor analysis by Wood, Tataryn, and Gorsuch (1996) showed more bias with under- than overextraction, and that overextraction was a particularly robust technique when singleton constructs were involved (singleton constructs are those for which a single variable is present in the data set). Overextraction is therefore particularly important for the identification of constructs having limited measures in the data set, a situation that probably holds for many normative sets, including the present one.

Because of these considerations, a 9-factor solution was extracted. The factor loadings appear in Table 2. The 9 factors accounted for over 84% of the variability in the norms. Factor loadings greater than .300 are in bold. Four of the factors (5–8) and their respective variables show particularly clean results, in the sense that variables loading on these factors are relatively pure measures of the constructs (i.e., they do not also have substantial loadings on other factors), and the factors themselves have substantial loadings only for those variables (i.e., other variables do not load on these factors). In turn, Factors 5–8 are readily identifiable as emotionality, initial letters, gender ladenness, and goodness–pleasantness.

Factor 9 is also a relatively clear measure of number of different meanings (i.e., homonymity), although additional measures of this construct would strengthen the case. The moderate loading for *Assoc* on Factor 9 makes sense in that the more meanings a word has, the more different associates people are likely to generate.

Table 2
Factor Analysis Results for 925 Nouns

	FAM 1	LEN 2	FRQ 3	IMG 4	EMO 5	INL 6	GEN 7	PLS 8	AMB 9
Fam2	89	09	28	12	12	10	01	07	11
ConAv	86	10	22	28	13	05	00	06	09
Def	82	10	16	39	05	04	01	05	04
Fam	74	13	51	12	08	10	04	12	13
Pron	71	34	32	13	02	08	-01	10	10
Age	-53	-31	-35	-52	-01	02	-14	-10	-11
Sou	18	90	18	10	01	08	-00	03	09
Rhy	20	86	09	02	03	13	00	04	-03
Loo	23	77	08	04	-06	17	-02	03	29
Syl	-02	-73	-28	-33	-03	05	-03	02	03
Len	03	-71	-37	-27	-03	20	-01	05	02
AvlKB	20	28	83	02	06	04	01	11	11
FrqTL	38	23	78	17	03	09	04	14	14
FrqKF	41	09	77	-08	03	11	-05	15	15
AvlRF	35	34	71	30	12	00	03	03	08
Dict	33	26	58	48	-07	-03	08	10	-03
Img	24	15	05	91	01	-04	01	02	-06
Con	05	13	-02	87	-28	-03	-03	-07	-14
Mng	35	12	19	70	10	-01	-08	09	-04
Img2	60	17	15	69	12	-02	-00	03	-01
EmoGd	04	00	-01	-05	92	02	00	-17	05
EmoPl	04	06	02	09	90	01	05	04	06
Emo	23	-05	11	-15	82	01	02	16	16
Let	09	03	02	-04	03	90	-02	01	-08
LetBr	-09	-08	06	08	02	85	-02	01	15
InSnd	24	25	06	-15	-01	76	-02	-01	02
Gend2	02	01	02	-00	02	-02	97	18	01
Gend	01	-06	01	-04	05	-03	96	22	00
Gdn	09	02	15	03	-08	-00	22	93	02
Pls	17	02	16	05	10	02	24	90	11
Amb	11	17	17	-18	15	07	01	07	85
Assoc	44	05	30	-03	32	01	02	10	63
%Var	16.13	12.32	12.08	11.63	8.29	7.12	6.27	6.13	4.48
Cum%Var	16.13	28.45	40.53	52.16	60.44	67.56	73.83	79.97	84.45

Note—Decimals omitted from main body of table.

The story for Factors 1–4 is more complex, although again each factor includes variables that appear to be relatively pure measures. Factor 1 is a familiarity and/or context availability factor. Factor 2 represents some feature of the physical form of the word, perhaps a combination of being short in length and having many similar sounding or rhyming words. Interestingly, it is possible that these elements would have opposite effects on cognitive tasks; for example, word identification might be faster for short words, but slower for words with many competitors. Factor 3 is a frequency and/or availability factor. Factor 4 is clearly an imagery–concreteness factor.

The complexities for Factors 1–4 arise from variables that load on more than one factor, implicating a multidimensional underlying structure for these variables. There are several notable examples of this. First, the two measures of age of acquisition (Age and Dict) load rather evenly across these four factors. Age of acquisition appears to reflect words that are familiar, short, concrete, and occur frequently in the English language. The composite nature of age of acquisition provides one possible explanation for why it has proven to be such a robust predictor of picture naming when placed in direct competition with individual components.

There are several other examples of multicomponent variables. Number of associations loads on the ambiguity, familiarity, and emotionality factors. The Img2 ratings load about equally on imagery and familiarity. The familiarity loadings for these variables make theoretical sense in that it is difficult to generate associations or images to words with which we are unfamiliar. The Rubin and Friendly availability measure (AvlRF), based on the occurrence of words as free associations to other words, taps elements of familiarity, word length, and even concreteness. These influences, especially concreteness, are less apparent for the Keenan and Benjafield availability measure (AvlKB), which is based on the occurrence of words in the formal definitions of other words.

PART 2

Expanded Number of Items

Part 1 reported on a markedly increased set of properties for the original 925 PYM items. There are several reasons to also consider extending the number of items for which norms are available. First, if anything like the 9-factor solution reported above is correct, very large numbers of items may be required to select items that

represent orthogonal sets of properties; as one very extreme example, two levels for each of 9 factors would result in $2^9 = 512$ cells. Finer gradations of the dimensions or an increased number of dimensions would entail even more cells. Until the dimensionality of word properties related to human cognition is known, the most cautious assumption is probably that very large numbers of words will be needed for the kinds of studies and simulations likely to be undertaken in the future.

Our second extension of the PYM norms, therefore, involves a more than doubling of the number of items for which basic word properties are available, specifically to 2,311 words. Aspects of these norms have not previously been published, although they have been distributed widely by the second author to interested researchers (Paivio, 1974). One motivation for this extension was the desire to provide normative information for parts of speech other than nouns (the original PYM norms included virtually no other parts of speech). Given the important role of item attributes in language processing, restricting the word pool to nouns seriously limits the kinds of linguistic research that could be conducted with the original PYM norms.

Method

Participants. Familiarity was rated by four groups of from 47 to 49 participants each. Two additional groups of 16 participants each rated imagery value. All participants were introductory psychology students.

Materials and procedures. The word pool included a total of 2,311 words, including the original 925 PYM items. The additional 1,386 items were sampled from diverse parts of speech. Instructions for the imagery rating task followed those used for the original PYM norms. For the imagery ratings, the new items were divided randomly into two sets, with half the participants rating each set.

For the familiarity ratings, 133 items were repeated so as to permit calculation of additional information about the reliability of the familiarity ratings (for this reason, some references to these norms indicate 2,448 items). The words were divided into four sets of ap-

proximately the same size, with a different group of participants rating each set. Familiarity instructions were similar to those presented in Appendix A, but with somewhat more detail (e.g., about distributing scores across the full range of values, when appropriate).

For both rating tasks, groups of participants read the instructions and were asked whether they had any questions. Questions were answered with paraphrases of the original instructions. The participants then completed the rating task at their own pace.

Additional norms. In addition to the new ratings, several properties have been obtained for each word, including Kučera–Francis and Thorndike–Lorge frequencies (FrqKF, FrqTL). To obtain measures of the physical length of the words, the number of letters (Len) was computed by the database program in which the words were stored, and the number of syllables (Syl) was determined by several judges. The length of words in visual and spoken forms is likely to play an important role in the early identification of words and in their production. Scores for the frequency of occurrence in 50 children's dictionaries (Dict) were obtained from the Clark (2002) norms. For the following analyses, we used the logarithm to the base 10 of the frequency and children's dictionary measures, with one added to avoid the problem of an undefined log for words with a frequency of zero.

Results and Discussion

As in Part 1, we first consider normative information for the scales, and then examine the underlying factor structure.

Normative statistics. Table 3 shows means and *SDs* for the 2,111 words containing information on all properties. The words tend toward the midpoint of the imagery scale and somewhat toward the upper end of the familiarity scale. The sizable *SDs* for the imagery and familiarity ratings, especially the former, demonstrate the considerable internal consistency across raters (a large *SD* indicates a large average *r* between raters). For imagery, reliability was further confirmed by $r = .96$ across 123 nouns and a more modest $r = .67$ across 63 adjectives for which prior imagery ratings were available. The familiarity ratings also demonstrated excellent reliability. The split-half reliability was .92 for the entire set of words, and the two

Table 3
Descriptive Statistics, Correlation Matrix, and Factor Loadings for 2,111 Words of the Expanded Paivio Norms

	Img	Let	Syl	Fam	FrqKF	FrqTL	Dict	%Variance	Cumulative
Descriptive Statistics									
Mean	4.21	6.71	2.18	5.30	1.30	1.18	.55		
<i>SD</i>	1.44	2.18	.96	1.09	.82	.68	.59		
Correlations									
Letters	-.15								
Syllables	-.20	.82							
Familiarity	.05	-.31	-.31						
FrequencyKF	-.18	-.40	-.33	.76					
FrequencyTL	.17	-.48	-.40	.63	.67				
Dictionaries	.28	-.55	-.49	.62	.60	.74			
Factor Loadings and Variance Explained									
Factor 1	.02	-.26	-.19	.89	.87	.82	.75	41.20	41.20
Factor 2	-.10	.92	.92	-.09	-.20	-.28	-.38	28.20	69.40
Factor 3	.97	-.05	-.10	.01	-.29	.17	.31	16.62	86.02
Communalities									
	.96	.91	.90	.80	.88	.78	.80		

ratings for the 133 duplicated words correlated .948. Moreover, as is demonstrated in Table 3, familiarity ratings correlated highly with the related frequency measures from Thorndike–Lorge and Kučera–Francis.

Factor structure of the expanded item pool. A principal components factor analysis with varimax rotation revealed two factors with eigenvalues greater than 1.00, but the eigenvalue for a third factor was also very close to 1.00 (.951). Moreover, communalities for several scales were greatly improved by extracting a third factor, and, as noted in Part 1, overextraction is often a desirable strategy, especially when a single measure is available for one or more factors. A 3-factor solution was extracted, and the factor loadings appear in Table 1, along with the correlation matrix among the original measures and the communalities. The three factors accounted for 84.16% of the total variability across the seven measures.

The factor structure is quite self-explanatory. The first factor is clearly a frequency–familiarity factor, with high loadings for familiarity ratings and the two frequency measures. Frequency of words in children’s dictionaries also has a sizable loading on this factor. The second factor is obviously a word-length factor with loadings on number of letters and number of syllables. Dict also loads on this factor, albeit more modestly than the primary variables. The final factor is represented at present in this data set by the lone imagery measure, with a more modest loading for Dict.

Considering now the variables, the communalities demonstrate that the 3-factor structure captures much of the variability in the seven measures, with the children’s dictionary measure demonstrating the greatest unaccounted-for variance. As in Part 1, Dict is also the least pure of the measures, having loadings greater than .30 on all three factors and still having some considerable residual variance yet to be accounted for. This is consistent with the argument advanced in Part 1 that age of acquisition, for which Dict is a proxy, is a complex, multidimensional measure with several distinct components.

GENERAL DISCUSSION

Part 1 in particular clearly demonstrates the benefits of researchers collecting norms for some standard set of words. In addition to our own scales, we were fortunate to be able to include scales based on other researchers who have obtained property data for the original PYM items. This approach has not yet been exhausted, and we will continue to incorporate other norms and experimental results, both those that already exist (e.g., Rubin, 1981), and additional norms that future researchers might collect. The development of a norm archiving system associated with the Psychonomic Society will greatly enhance these efforts, especially given larger item pools, such as that presented in Part 2.

We strongly recommend, therefore, that researchers continue to coordinate their efforts toward the collection of item properties for well-established pools of words. This will result in the evolution of increasingly complete

and comprehensive norms, ideally for additional properties and stimulus materials. There are several advantages to this systematic approach.

First, the availability of increasingly comprehensive norms for a given pool of words allows for a deeper analysis and understanding of the relationships among the diverse properties that characterize words and, in some cases, their associated concepts. Our results demonstrate nicely both the unidimensional constructs that are relatively independent of other measures (at least of those measures that have been collected to date) and the multidimensional constructs that are more intertwined with diverse factors. This factor structure sheds light on the theoretical nature of the various measures.

Second, comprehensive norms allow researchers who posit a new item attribute to quickly determine the extent to which the new construct is unique, compared with existing constructs, as well as whether it demonstrates correlations with other constructs that would be expected theoretically. We believe that this will lead to earlier and more thoughtful consideration of the construct validity of the (perhaps) novel measure.

Third, comprehensive norms for sufficiently large numbers of items will allow ever more sophisticated selection of items for experimental use, along with greater control of competing properties. The use of comprehensive norms to identify sets of items that can be distinguished empirically on some attribute (and not on others) will also enable researchers to reflect more thoughtfully about how the construct of interest is (or is not) intertwined with other related constructs.

Methodological Caveats

Several caveats are in order with respect to research involving the collection and use of item properties. It is important that researchers appreciate fully the limitations of their data sets by virtue of the particular measures that are available. The amount of variability accounted for by a factor, for example, does not represent in any absolute sense the “importance” of that factor. Rather, it reflects the number of measures relevant to that construct that happen to be in the data set.

A related caveat is that researchers need to be cautious about excessive reduction of their variables into underlying constructs or factors. Unless they use several converging measures of all constructs of interest, overextraction should be the norm, as recommended by Wood et al. (1996). In addition to such standard considerations as an eigenvalue threshold, this implies examination of additional statistics (e.g., communalities for the measures) and deeper speculation about the entire domain of potential properties for a given stimulus set (e.g., whether certain constructs are inadequately represented in the existing norms).

Researchers also need to adopt statistical techniques that lend themselves to the analysis of richly interconnected predictors. Multiple regression, for example, has the unfortunate tendency to draw researchers’ attention toward the unique contribution of each predictor, hiding any

shared variability in the omnibus R and F test for the overall regression equation. Unless careful thought has been given in prior analyses to reducing the predictors to independent components (e.g., factor analysis), researchers may be led astray by a focus on idiosyncratic and perhaps noise-laden aspects of the stimulus properties.

Of particular importance is the need for researchers to be imaginative and persistent in their efforts to generate both new measures and more sophisticated theories for the constructs that underlie the complex similarities and differences among cognitive materials. The focus here has been on properties of words, but similarly complex and multidimensional issues arise with other cognitive materials, such as metaphors (e.g., Katz, Paivio, Marschark, & Clark, 1988) and pictures (e.g., Johnson, Paivio, & Clark, 1996).

Our orientation, which has demonstrated some success, has been to conceptualize item attributes (and related performance on cognitive tasks) around dual-coding theory. Although we would encourage others to consider this framework for their own work on item attributes, the more important general admonition is to adopt a broad theoretical framework for the conceptualization of the rich array of properties characteristic of the stimulus materials used in cognitive research. This comprehensive orientation is likely to be more successful in the long run than a narrow focus on individual attributes and their correlates in cognitive tasks.

Conclusion

The development and evaluation of principled models for the tremendously diverse range of properties to which verbal and nonverbal materials lend themselves is an immense challenge with respect to the understanding and application of normative data in cognitive psychology. We had observed this earlier with respect to a smaller set of items and properties associated with concrete words and pictures of their referents (Paivio et al., 1989, see also Rubin, 1980), and have again demonstrated here the rich interrelationships among a mixture of possibly distinct and overlapping constructs. Since the examination of item attributes and their behavioral correlates continues to be one of the primary tools of cognitive psychology (and one that we would like to see amplified further), the discipline would benefit from researchers expanding the set of properties represented in existing norms (as represented here in Part 1), and expanding the number of items for which norms are available (as represented here in Part 2). This comprehensive and coordinated approach will help item attribute research to contribute more fully to our understanding of human cognition.

REFERENCES

- BROWN, G. D. A. (1984). A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. *Behavior Research Methods, Instruments, & Computers*, **16**, 502-532.
- CLARK, J. M. (2002, May). *Age of acquisition: A new measure and its correlates*. Presented at the Annual Meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Vancouver, BC.

- ELLIS, A. W., & LAMBON RALPH, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1103-1123.
- JOHNSON, C. J., PAIVIO, A., & CLARK, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, **120**, 113-139.
- KATZ, A. N., PAIVIO, A., MARSCHARK, M., & CLARK, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor & Symbolic Activity*, **3**, 191-214.
- KEENAN, T. R., & BENJAFIELD, J. G. (1994). An additional measure of availability derived from the *Oxford English Dictionary*. *Psychonomic Bulletin & Review*, **1**, 255-257.
- KISS, G. R., ARMSTRONG, C. A., & MILROY, R. (1972). *An associative thesaurus of English* [Microfilm version]. Wakefield, U.K.: E. P. Microform.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- MORRISON, C. M., & ELLIS, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology*, **91**, 167-180.
- PAIVIO, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, **76**, 241-263.
- PAIVIO, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston. (Reprinted 1979, Hillsdale, NJ: Erlbaum).
- PAIVIO, A. (1974). [Imagery and familiarity ratings for 2,448 words]. Unpublished norms. London, ON: University of Western Ontario, Department of Psychology.
- PAIVIO, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- PAIVIO, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, **45**, 255-287.
- PAIVIO, A., CLARK, J. M., DIGDON, N., & BONIS, T. (1989). Referential processing: Reciprocity and correlates of naming and imaging. *Memory & Cognition*, **17**, 163-174.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, **76**(1, Pt. 2), 1-25.
- PROCTOR, R. W., & VU, K-P. L. (1999). Index of norms and ratings published in the Psychonomic Society journals. *Behavior Research Methods, Instruments, & Computers*, **31**, 659-667.
- PYLYSHYN, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, **80**, 1-24.
- RUBIN, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning & Verbal Behavior*, **19**, 736-755.
- RUBIN, D. C. (1981). First-order approximation to English, second-order approximation to English, and orthographic neighbor ratio norms for 925 nouns. *Behavior Research Methods & Instrumentation*, **13**, 713-721.
- RUBIN, D. C., & FRIENDLY, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory & Cognition*, **14**, 79-94.
- SCHWANENFLUGEL, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 223-250). Hillsdale, NJ: Erlbaum.
- SCHWANENFLUGEL, P. J., & SHOEN, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 82-102.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College.
- WOOD, J. M., TATARYN, D. J., & GORSUCH, R. L. (1996). Effects of under- and over-extraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, **1**, 354-365.

ARCHIVED MATERIALS

The following materials associated with this article may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive. To access these files or links, search the archive for this article using the journal (*Behavior Research Methods, Instruments, & Computers*), the first author's name (Clark), and the publication year (2004).

FILE: ClarkPaivio1-BRMIC-2004.zip.

DESCRIPTION: The compressed archive file contains five files related to the extension of the 925 items in Paivio, Yuille, and Madigan (1968): cp2004a.var describes the variables, cp2004a.txt presents the norms in text format, cp2004a.dbf presents the norms in DBF format that can be read by database and spreadsheet programs, and two files that demonstrate reading and using the DBF file for SAS (cp2004.sas) and SPSS (cp2004a.spss).

FILE: ClarkPaivio2-BRMIC-2004.zip.

DESCRIPTION: The compressed archive file contains three files related to the extended norms for 2,311 words: cp2004b.var describes the variables, cp2004b.txt presents the norms in text format, and cp2004b.dbf

presents the norms in DBF format that can be read by database and spreadsheet programs, as well as by SAS and SPSS as illustrated for the other set of norms described above.

LINK: www.uwinnipeg.ca/~clark/cog/norms.

DESCRIPTION: Information about the above norms and additional related materials that become available can be accessed from this link. Follow the Clark & Paivio (2004) link.

AUTHOR'S E-MAIL ADDRESS: clark@uwinnipeg.ca.

AUTHOR'S WEB SITE: www.uwinnipeg.ca/~clark.

APPENDIX A
Instructions for 13 Additional Rating Tasks

Use the following scale to rate the words in the booklet on their AGE OF ACQUISITION—that is, on how old you were when you learned to understand and use the words. Words that you learned early in childhood get high ratings. Words that you learned later in childhood get low ratings.

Use the following scale to rate the words in the booklet on the NUMBER OF WORDS WITH THE SAME INITIAL LETTERS. Words that share their beginning letters with many other English words get high ratings. Words that begin with letter combinations that are uncommon get low ratings.

Use the following scale to rate the words in the booklet on the NUMBER OF WORDS WITH THE SAME INITIAL SOUND. Words that share their beginning sound with many other English words get high ratings. Words that begin with a sound that is relatively rare get low ratings.

Use the following scale to rate the words in the booklet on their NUMBER OF RHYMES—that is, on how many words there are in English that rhyme with or end with similar sounds as each word. Words that have many rhyming words get high ratings. Words that have few rhyming words get low ratings.

Use the following scale to rate the words in the booklet on their NUMBER OF ASSOCIATIONS—that is, on how many words there are in English that are associated with or related to each word. Words that have many associates get high ratings. Words that have few associates get low ratings.

Use the following scale to rate the words in the booklet on their EASE OF DEFINITION—that is, on how easy you would find it to define and explain each word. Words that are easy to define get high ratings. Words that are difficult to define get low ratings.

Use the following scale to rate the words in the booklet on the NUMBER OF SIMILAR SOUNDING words there are in the English language; that is, on how many words sound like the stimulus word. Words that have few similar sounding words receive low ratings, and words that have many similar sounding words get high ratings.

Use the following scale to rate the words in the booklet on the NUMBER OF SIMILAR LOOKING words there are in the English language; that is, on how many words look like the stimulus word. Words that have few similar looking words receive low ratings, and words that have many similar looking words get high ratings.

Use the following scale to rate the words in the booklet on their NUMBER OF MEANINGS—that is, on whether they have only one meaning or many meanings. Words that have many different meanings get high ratings. Words that have only one meaning or few meanings get low ratings.

Use the following scale to rate the words in the booklet on their EMOTIONALITY—that is, on the degree to which the words would evoke a positive or negative emotional response from people. Words that elicit strong feelings get high ratings. Words that are not emotional get low ratings.

Use the following scale to rate the words in the booklet on their PLEASANTNESS—that is, on the degree to which they would arouse positive feelings in people. Pleasant words that elicit positive feelings get high ratings. Words that arouse negative feelings get low ratings.

Use the following scale to rate the words in the booklet on their PRONOUNCEABILITY—that is, on how easy you find it to pronounce and speak the word. Words that are easy to pronounce and say get high ratings. Words that are difficult to pronounce and say get low ratings.

Use the following scale to rate the words in the booklet on CONTEXT AVAILABILITY—that is, how easy it is to think of a particular context or circumstance in which each word might appear. Think of a context for each word and then rate how easy it was to think of a context on a scale of 1 to 7, where 1 means “very hard to think of a context” and 7 means “very easy to think of a context.” For example, it may be easy to think of a context for the words “baseball” and “repentance” (e.g., the World Series or a Church) and they should receive high ratings of 6 or 7. It may be harder to think of a context for the words “inversion” and “sloop,” although eventually contexts might come to mind (e.g., a smog inversion in a large city or sailing on a lake). Such words should receive low ratings of 1 or 2. Rate words of intermediate ease of context availability between these extremes. Print your rating in the space provided below, making sure that the number of the space corresponds to the the number of the word you are rating.

(Continued on next page)

APPENDIX B
Correlation Among 32 Properties for 925 PYM Items

	Fam2	ConAv	Def	Fam	Pron	Age	Sou	Rhy	Loo	Syl	Len	AvlKB	FrqTL	FrqKF	AvlRF
ConAv	93														
Def	83	85													
Fam	85	80	76												
Pron	78	77	66	75											
Age	-64	-69	-74	-69	-66										
Sou	31	32	29	37	49	-49									
Rhy	28	26	26	33	42	-40	85								
Loo	33	31	34	37	43	-44	77	71							
Syl	-23	-28	-26	-28	-45	48	-66	-54	-49						
Len	-19	-24	-22	-27	-40	44	-66	-51	-47	80					
AvlKB	48	43	35	61	52	-48	45	37	40	-40	-44				
FrqTL	63	59	54	76	64	-64	45	38	41	-41	-42	84			
FrqKF	61	53	45	75	57	-47	32	26	31	-24	-28	73	79		
AvlRF	59	59	56	73	62	-71	53	44	43	-51	-55	74	79	69	
Dict	50	52	59	63	54	-82	46	37	36	-44	-46	57	69	55	76
Img	34	48	55	34	34	-61	28	22	19	-40	-35	16	33	08	44
Con	10	24	35	11	17	-44	19	14	12	-34	-30	03	16	-06	24
Mng	47	57	57	48	46	-59	29	24	21	-35	-33	31	43	28	50
Img2	69	77	80	60	60	-78	36	30	29	-40	-35	33	50	34	59
EmoGd	12	13	06	08	02	00	-00	02	-00	02	00	05	03	02	10
EmoPl	16	18	14	16	10	-15	07	08	05	-07	-05	10	11	08	19
Emo	32	32	18	29	21	-07	04	03	00	02	01	21	20	25	22
Let	17	12	09	15	14	-02	11	15	13	-00	14	06	11	14	05
LetBR	08	04	02	09	06	02	03	03	11	03	13	05	08	11	02
InSou	29	21	19	28	29	-15	29	32	37	-11	-01	20	23	27	17
Gend2	05	04	03	08	03	-15	02	03	-00	-04	-02	05	08	01	06
Gend	04	03	02	08	02	-12	00	01	-02	-01	00	05	08	01	04
Gdn	19	17	16	26	21	-23	08	07	08	-06	-05	24	29	27	19
Pls	29	27	26	35	29	-32	11	09	12	-08	-05	29	35	33	26
Amb	28	23	09	30	29	-16	25	17	34	-12	-12	32	31	35	26
Assoc	55	50	47	60	48	-42	24	17	33	-10	-11	43	51	53	46

APPENDIX B
(Continued)

Dict	Img	Con	Mng	Img2	EmoGD	EmoPL	Emo	Let	LetBR	InSou	Gend2	Gend	Gdn	Pls	Amb
54															
44	83														
54	72	56													
66	80	60	70												
-09	-04	-30	05	08											
05	08	-15	13	18	77										
-00	-06	-38	09	16	69	65									
-00	-03	-05	01	03	03	01	05								
-04	-04	-05	-01	-02	02	02	03	65							
11	-04	-06	02	08	02	02	07	64	47						
09	02	-05	-02	02	-01	07	08	-03	-03	-03					
08	-01	-09	-04	-01	01	09	10	-03	-04	-04	95				
24	07	-02	13	11	-21	-02	10	01	01	02	40	42			
29	10	-06	17	21	-04	16	28	03	03	06	40	44	90		
06	-15	-31	-03	02	19	19	33	07	10	20	05	05	12	20	
29	07	-17	20	32	32	34	51	03	08	16	06	07	15	32	57

(Manuscript received January 1, 2004;
revision accepted for publication July 18, 2004.)