Timed picture naming: Extended norms and validation against previous studies

ANNA SZÉKELY
Eötvös Loránd University, Budapest, Hungary

SIMONETTA D'AMICO
University of Rome "La Sapienza," Rome, Italy
and Libera Università Maria SS. Assunta, Rome, Italy

ANTONELLA DEVESCOVI
University of Rome "La Sapienza," Rome, Italy

KARA FEDERMEIER and DAN HERRON
University of California, San Diego, California

GOWRI IYER
University of California, San Diego, California
and San Diego State University, San Diego, California

THOMAS JACOBSEN
University of Leipzig, Leipzig, Germany

and

ELIZABETH BATES
University of California, San Diego, California

Factors affecting word retrieval were compared in a timed picture-naming paradigm for 520 drawings of objects. In prior timed and untimed studies by Snodgrass and Vanderwart (1980) and Snodgrass and Yuditsky (1996), concerns were raised that participants could not reliably name large numbers of items in a single session. We show that reliable results are obtained in a single session for 520 items and validate our method against previous findings by Snodgrass and colleagues for overlapping items. For these items, comparable levels of name agreement and latency are obtained, and we replicate effects of length, frequency, both objective and subjective age of acquisition, and visual complexity on reaction time (RT) and name agreement measures. Name agreement is unaffected by order of presentation, although there is a gradual increase in RTs across the session, requiring use of multiple random orders. Current extensions of our method include cross-linguistic, bilingual, developmental, and neuro-psychological studies and comparisons of action naming and object naming.

In 1980, Snodgrass and Vanderwart introduced a set of 260 pictures standardized for name agreement by native speakers of English. They determined the dominant name for each picture (defined empirically as the response given by the largest number of participants) as well as the number and nature of the alternative names provided for each picture (yielding measures of conceptual agreement and redefinitions of *name agreement*, in which synonyms and/or morphological variants of the dominant name were included). For the dominant names elicited by each picture, they examined the effects of variables that were known to play a central role in memory and retrieval, including various indices of frequency, age of acquisition (AoA), familiarity, imageability, image agreement, and visual complexity. Naming latencies were established for this picture set by Snodgrass and Yuditsky (1996), and contributions of the same lexical and pictorial variables to latency as well as to name agreement were determined. This set has been used in many picture-naming studies, sometimes with additional items (e.g., Barry, Hirsch, Johnston, & Williams, 2001; Barry, Morrison, & Ellis, 1997; Dell'Acqua, Lotto, & Job, 2000;

Feyereisen, Demaeght, & Samson, 1998; Morrison, Chappell, & Ellis, 1997; Morrison, Ellis, & Quinlan, 1992; Nation, Marshall, & Snowling, 2001; Nisi, Longoni, & Snodgrass, 2000; Sanfeliu & Fernandez, 1996).

Although there is now a large body of literature on picture naming both in and out of context (see Johnson, Paivio, & Clark, 1996, for a review), most of the work to date has focused on names for common objects depicted in drawings or photographs. Because the same normed stimuli have been used in so many studies, integration of findings across studies is facilitated. In the present study, we provide an extended set of norms for timed picture naming using 520 pictures of common objects, including a subset of the pictures from the original Snodgrass studies. We include many of the same predictor variables and measures of naming behavior that have been employed in previous studies and validate our results against prior findings.

In virtually all naming studies, written or spoken word frequencies are reported to be inversely related to reaction times (RTs; see, e.g., Forster & Chambers, 1973; Humphreys, Riddoch, & Quinlan, 1988; Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965; Preston, 1935; Thorndike, 1931). From the 1970s, AoA (based on adult estimates of the age at which a word was learned) has challenged frequency as the critical factor in predicting naming latency (Barry et al., 2001; Barry et al., 1997; Carroll & White, 1973a, 1973b; Gilhooly & Gilhooly, 1979; Morrison et al., 1992; Rochford & Williams, 1962a, 1962b). Lachman and his colleagues (Lachman, 1973; Lachman, Lachman, Thronesbery, & Sala, 1980; Lachman, Shaffer, & Hennrikus, 1974) reported that name agreement is another important determiner of RT in naming tasks, especially if it is low. They used the information statistic U as a measure of name agreement, on the basis of the frequency distribution and the number of alternative names for each picture. The same measure was adopted by Snodgrass and Vanderwart (1980). These norms and measures, and others like them, have proven useful in many different applications of picture naming, including neuropsychological studies with clinical populations (Chen & Bates, 1998; Druks, 2002), comparative studies of children and adults (Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; D'Amico, Devescovi, & Bates, 2001), and recent studies investigating covert or overt picture naming using fMRI (Hernandez, Martinez, & Kohnert, 2000; Murtha, Chertkow, Beauregard, & Evans, 1999) and event-related brain potentials (Schmitt, Münte, & Kutas, 2000; Schmitt, Schiltz, Zaake, Kutas, & Münte, 2001; Van Turennout, Hagoort, & Brown, 1997, 1999).

In our laboratories, we have used timed picture naming to assess cross-linguistic differences in lexical retrieval under different contextual conditions (Bates, Devescovi, & Wulfeck, 2001). For example, we have shown that grammatical gender can affect the time required to name a picture embedded in a grammatically congruent (facilitative) or incongruent (inhibitory) auditory phrase, in Spanish (Wicha et al., 2001), Italian (Bentrovato, Devescovi, D'Amico, & Bates, 1999), German (Jacobsen, 1999), and Swahili (Alcock & Ngorosho, in press). In the same vein, we have investigated syntactic priming effects on naming of actions versus naming of objects in English (Federmeier & Bates, 1997) and Chinese (Lu et al., 2002) using short phrases that are congruent or incongruent with form class (e.g., "Here is the table" vs. *"He likes to table"). We have also used timed picture naming to assess degrees of fluency or dominance in the two languages of Spanish–English bilinguals (Kohnert, Bates, & Hernandez, 1999; Kohnert, Hernandez, & Bates, 1998).

Although we have made use of the Snodgrass pictures and norms to derive many of our items, it became clear to us early on that a much larger list of items would be required in order to achieve a balance over languages in the experimental parameters of interest (e.g., target name length, target name frequency, and other factors that assess "nameability"). With this goal in mind, we initiated a cross-linguistic norming study several years ago (Bates et al., 2000), assembling a larger set of pictures than has been employed in previous studies. Hence, it was important for us to establish the validity and reliability of timed picture naming using a set of this size. In their pioneering studies of both timed (Snodgrass & Yuditsky, 1996) and untimed (Snodgrass & Vanderwart, 1980) picture naming, Snodgrass and colleagues were concerned that participants would not be able to name the full set of 260 items in a single session. For that reason, they broke the list down into two separate lists and administered them to different groups of participants. For a variety of reasons, both pragmatic (to facilitate testing of large numbers of participants) and scientific (to reduce interparticipant variability across the list of items), we hoped that it would be possible to obtain norms for all 520 items from individual participants in a single session. Pilot studies suggested that this would indeed be possible, in a briskly paced timed naming paradigm lasting approximately 45 min (with occasional breaks). All of our cross-linguistic norms have been collected in this fashion (Bates et al., 2003; Bates et al., 2000; Székely et al., in press).

The present study serves as a methodological cornerstone for a series of cross-linguistic, developmental, and neuropsychological studies that have just been completed or are still under way using our variant of the Snodgrass method. Here, we introduce our methodology and the norms that we have obtained for English for all 520 items. A subset of our items ($N = 161$) overlapped with the picture stimuli used by Snodgrass and Vanderwart (1980) and Snodgrass and Yuditsky (1996) and elicited the same target names in our study and in theirs. We therefore took advantage of the detailed published norms provided by these investigators to compare our results with theirs for this overlapping subset, to validate our findings against previous work. Comparisons focus on dependent variables (name agreement, number and variety of alternative names, mean and variance of naming latency for dominant and alternative names) as well

as predictor variables (effects associated with characteristics of the pictures and with characteristics of the dominant name for each picture). Finally, we present evidence of the reliability of our method. This includes an examination of the effect of order of presentation on name agreement and latency within the 10 different randomized orders that we used, to assess any potential effects of session length on naming performance.

METHOD

Participants

Fifty right-handed college students at the University of California at San Diego (35 female and 15 male, ranging in age from 18 to 25 years) participated in the picture-naming study. They were all native speakers of English, reported normal or corrected-to-normal visual acuity, and reported no exposure to languages other than English before age 12. They received course credit for their participation. In addition, on-line subjective rating studies were obtained from a separate set of 73 participants (all English-speaking, right-handed undergraduate students at the University of California at San Diego): 20 for the goodness-of-depiction ratings and 53 for the age-of-acquisition ratings.

Materials

Picture stimuli were black and white line drawings of 520 common objects (see Table 1), including 174 pictures from the original Snodgrass and Vanderwart (1980) set. They were scanned and stored digitally for presentation within the PsyScope Experimental Control Shell (Cohen, MacWhinney, Flatt, & Provost, 1993). Pilot studies were carried out for the selection of these pictures. Several different sources of picture material were used, resulting in more than 1,000 pictures in the pretest set. Item selection was subject to several constraints, including picture quality, visual complexity, and cross-cultural validity of the depicted item. The pilot naming studies indicated that normal adult participants were able to name

Table 1
Sources of Picture-Naming Stimuli

Source	N
Snodgrass & Vanderwart, 1980[1]	174
Alterations of Snodgrass & Vanderwart, 1980[1]	2
Peabody Picture Vocabulary Test, 1981[2]	62
Alterations of Peabody Picture Vocabulary Test, 1981[2]	8
Dronkers set[3]	39
Abbate & La Chappelle, "Pictures, Please," 1984[4,5]	168
Max Planck Institute for Psycholinguistics[6]	20
Boston Naming Test, 1983[7]	5
Oxford "One Thousand Pictures"[8]	25
Miscellaneous	17

[1]Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 174-215.
[2]Dunn, Lloyd M., & Dunn, Leota M. (1981). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: American Guidance Service.
[3]Picture set used by Dronkers, N. (personal communication).
[4]Abbate, M. S., & La Chappelle, N. B. (1984a). *Pictures, please! An articulation supplement*. Tucson, AZ: Communication Skill Builders.
[5]Abbate, M. S., & La Chappelle, N. B. (1984b). *Pictures, please! A language supplement*. Tucson, AZ: Communication Skill Builders.
[6]Max Planck Institute for Psycholinguistics, Postbus 310, NL-6500 AH Nijmegen, The Netherlands.
[7]Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia: Lee & Febiger.
[8]*Oxford junior workbooks*. Oxford: Oxford University Press.

520 items in a single 45–60-min session including occasional breaks.

The picture stimuli used in this study, with their empirically determined dominant and alternative names and main dependent variables, are available at http://www.crl.ucsd.edu/~aszekely/ipnp/method.html. The main independent variables for the picture and the dominant response (objective and subjective AoA measures, goodness of depiction ratings, word frequency, and objective visual complexity) are included as well. With regard to word attributes, all variables are characteristics of the dominant response (i.e., the name given by the largest number of participants in the study). This includes presence/absence of a fricative or affricate in the initial consonant, which is a variable that has been reported to influence the time required for a response to register on the voice key.

An objective measure of AoA was derived from published norms for the American version of the MacArthur Communicative Development Inventories, or CDI (Fenson et al., 1994). The CDI is based, among other things, on concurrent parent reports of vocabulary development in very large samples of children, collected in a recognition-memory format with a large checklist of words that are likely to be acquired between the ages of 8 and 30 months. For our present purposes, the CDI yields a simple 3-point scale: 1 = *words acquired (on average) between 8 and 16 months*; 2 = *words acquired (on average) between 17 and 30 months*; and 3 = *words that are not acquired in infancy (> 30 months)*. Although it could be argued that this parent report measure is simply another kind of subjective rating scale, the measure itself was based on diary and free-speech studies, and large numbers of validation studies have shown that the developmental effects detected by the CDI correlate highly with free speech and other forms of language behavior by children in the 8–30-month age range. Hence, we view the CDI as an objective measure of AoA, to complement the subjective ratings of AoA that have been used in other studies (for examples of how the contribution of these two AoA measures may differ for picture naming and reading, respectively, see Bates, Burani, D'Amico, & Barca, 2001).

Frequency counts were taken from frequency dictionaries of written English (the CELEX Lexical database [Baayen, Piepenbrock, & Gulikers, 1995] as well as Kučera & Francis, 1967, and Francis & Kučera, 1982). In accordance with Snodgrass and Vanderwart (1980) and Snodgrass and Yuditsky (1996), ln (1 + raw frequency count) was applied to normalize frequency measures for use in correlational analyses.

The "shared name" variable reflects the fact that some dominant names were used for more than one picture. For example, the word *bottle* was used as the dominant name for the picture of a baby bottle as well as for that of a wine bottle. Items that share the same dominant name with another picture were specified by a dichotomous variable (1 = *shared name*; 0 = *not a shared name*). The dominant names that were shared by two or more pictures are *bird* (45 and 317), *boat* (48 and 391), *bottle* (18 and 53), *brush* (64 and 191), *chest* (92 and 474), *fence* (152 and 175), *glass* (180 and 511), *gun* (189 and 354), *hat* (80 and 199), *needle* (280 and 436), *priest* (265 and 338), and *stroller* (19 and 428).[1]

Another dichotomous variable is word complexity, which was assigned to any item on which the dominant response was a plural, a compound word, or a periphrastic (multiword) construction. Sixteen percent of the object items, such as *railroad tracks* and *ice cream cone*, had complex dominant names.

In addition to predictor variables associated with the target names, estimates of objective visual complexity were obtained for the picture itself, on the basis of the size of the digitized stimulus picture files (for further details, see Székely & Bates, 2000).

Procedure

The participants were tested individually in a dimly lit, quiet room. Before the picture-naming task began, voice sensitivity was calibrated for each participant with the help of a reading list of words featuring various initial phoneme patterns (none of these

words were appropriate as names for the pictures in the main experiment). The participants were instructed to name the pictures that would appear on the screen as quickly as they could without making mistakes and to avoid coughs, false starts, hesitations (e.g., "umm"), articles, other extraneous material (e.g., "*a* dog" or "*That's a* dog"), or anything other than the best and shortest name they could think of for the depicted object. To familiarize the participants with the experiment, a practice set of pictures depicting geometric forms such as a triangle, a circle, and a square were given as examples in object naming.

During testing, the participants wore headphones with a sensitive built-in microphone (adjusted to optimal distance from the participant's mouth) that were connected to the Carnegie Mellon button box, an RT-measuring device with 1-msec resolution designed for use with Macintosh computers. The pictures were displayed on a 15-in. VGA computer screen set to 640×480 resolution (pictures at 300×300 pixels). The participants viewed the centered items from a distance of approximately 80 cm. On each trial, a fixation plus sign (+) appeared centered on the screen for 200 msec, followed by a 500-msec blank interval. The target picture remained on the screen for a maximum of 3,000 msec. An interval of up to 4,000 msec was allowed for a response, which ended the trial. The picture disappeared after 3,000 msec or as soon as a response was made; at the same time, the appearance of a dot signaled voice detection—a clue for the error-coding procedure. The intertrial interval varied between 1,000 and 2,000 msec.

The RT associated with each response was recorded. Ten different scripts were used, each with a unique randomization of picture items. For each list, a printout of the items (the expected picture names) served as a score sheet for coding purposes during the experiment. The experimenter took notes on the score sheet according to an error-coding protocol (see details below). Alternative namings were also recorded manually on the score sheet. No pictures were preexposed or repeated during the test, and, hence, no training of the actual targets occurred. A short rest period was included automatically after 104 trials, but the participants could ask for a pause in the experiment at any time. The experimental sessions lasted 45 min on average and were tape-recorded for subsequent off-line checking.

To obtain subjective ratings of goodness of depiction and AoA, the same 520 object pictures were presented to independent samples of participants who were not part of the norming study itself. However, a keypress recording procedure allowed the participants to use a broad time interval in which to make their responses, since the stimulus remained on the screen until the participants responded by pressing one of the keys representing the scales. For goodness of depiction, the participants were asked to rate how well the picture fit its dominant name on a 7-point scale. Average ratings were calculated for each picture. In the AoA rating task, we applied two different experimental conditions (the participants were randomly assigned to either condition). In the word-only condition, the dominant names of the 520 objects were presented (the paradigm was adapted from Carroll & White, 1973a). In the picture–word condition, on the other hand, words were accompanied by pictures, the assumption being that the picture would facilitate the conceptualization of the word (the paradigm was adapted from Snodgrass & Yuditsky, 1996). In both conditions, the participants were told that they would see a series of stimuli presented one at a time on the computer screen. They were instructed to rate each stimulus they saw on a 9-point scale (2, 3, 4, 5, 6, 7–8, 9–10, 11–12, and 13+ years), which was marked on the keyboard, on the basis of their estimates of the age at which they had learned the word in either spoken or written form. The two subsets resulted in very similar ratings ($r = 0.98$, $p < .0001$); thus, subsequent analysis is based on the average rating of all 50 participants (for further details on background, procedure, and findings associated with AoA, see Iyer, Saccuman, Bates, & Wulfeck, 2001).

Scoring of the Picture Naming Study

Our scoring criteria were modeled closely on procedures adopted by Snodgrass and Vanderwart (1980), with a few exceptions. The target name for each picture was determined empirically, in two steps.

First, the data were subjected to error coding to determine which responses could be retained for both naming and RT analyses. Three error codes were found to be possible:

1. *Valid response* refers to all responses with a valid (i.e., codable) name and usable, interpretable RTs (no coughs, hesitations, false starts, or prenominal verbalizations such as "*that's a* ball"). Any word articulated completely and correctly was kept for the evaluation, except for expressions that were not intended namings of the presented object, such as "I don't know."

2. *Invalid response* refers to all responses with an invalid RT (i.e., coughs, hesitations, false starts, prenominal verbalizations) or a missing RT (i.e., the participant did produce a name, but it failed to register with the voice key).

3. *No response* refers to any trial in which the participant made no verbal response of any kind.

Only the valid responses were used for determining the target name and for further analyses. Once the set of valid responses had been determined, the target name was defined as the *dominant response*— that is, the name that was used by the largest number of participants. In the case of ties (two or more responses uttered by exactly the same number of participants), three criteria were used to choose one of the tied responses as the target: (1) closeness to the intended target (i.e., the hypothesized target name used to select stimuli prior to the experiment), (2) singularity, if singular and plural forms were tied, and (3) number of phonological variants.

Second, every response was coded into one of four different lexical categories:

Lexical Code 1: The target name (dominant response, empirically derived).

Lexical Code 2: Any morphological or morphophonological alteration of the target name, defined as a variation that shares the word root or a key portion of the word without changing the word's core meaning. Examples would include diminutives (e.g., "bike" for *bicycle*, "doggie" for *dog*), plural/singular alternations (e.g., "cookies" when the target name was *cookie*), reductions (e.g., "thread" when the target name was *spool of thread*), and expansions (e.g., "truck for firemen" when the target name was *fire truck*).

Lexical Code 3: Synonyms for the target name (which differ from Code 2 in that they do not share the word root or key portion of the target name). With this constraint, a synonym was defined as a word that shared the same truth-value conditions as the target name (e.g., "couch" for *sofa* or "chicken" for *hen*).

Lexical Code 4: This category was used for all names that could not be classified in Codes 1–3, including hyponyms (e.g., "animal" for *dog*), semantic associates that share the same class but do not have the target name's core meaning (e.g., "cactus" for *desert*), part–whole relations at the visual-semantic level (e.g., "finger" for *hand*), and so forth. Although some of these responses could be viewed as frank errors, we took an intentionally neutral stance regarding the "right" or "wrong" answer in this norming study. Hence, Lexical Code 4 is best viewed simply as "other." This "other" category included superordinate names such as "animal" or "food," or names that were reasonable names for the picture but did not relate to the dominant response (e.g., "cactus" where the dominant response was *desert*).

Name Agreement

Percent name agreement (*Lex 1dom*) was defined as the proportion of all valid trials (a codable response with a usable RT) on which the participants produced the target name. The number of alternative names for each picture was determined by number of types (i.e., number of different names provided on valid trials, in-

Table 2
Comparison of the Results of the Present Study with Snodgrass and Vanderwart
(1980) and Snodgrass and Yuditsky (1996) for the 161 Overlapping Items

Variable	Mean ($N = 161$) Snodgrass Studies	Mean ($N = 161$) CRL	Pearson's Correlation
H statistics[a]	0.486	0.495	+.69*
Name agreement[b]	89%	89%	+.67*
Lenient name agreement[c]	94%	93%	+.50*
Trimmed mean RT of target (msec)[d]	887	891	+.87*
Subjective AoA (1–9)[e]	3.87	3.68	+.95*
Visual complexity[f]	2.95	13,056 (bytes)	+.72*

[a]Snodgrass & Vanderwart (1980) *H* statistics of name agreement (based on written namings) and CRL *H* statistics (based on oral namings), both calculated in the same way.
[b]Snodgrass & Vanderwart (1980) percent name agreement (percentage of dominant written responses) and CRL Lex 1dom (percentage of dominant oral responses with valid RT).
[c]Snodgrass & Yuditsky (1996) %corr$_{lib}$ correctness under the liberal criteria (dominant responses, responses that occurred at least twice, and synonyms) and CRL % of valid RT dominant responses and their phonological variants or synonyms.
[d]Snodgrass & Yuditsky (1996) and CRL trimmed mean RT of valid dominant responses.
[e]Snodgrass & Yuditsky (1996) off-line subjective AoA ratings of pictures and their dominant names and CRL on-line subjective AoA ratings (same scale).
[f]Snodgrass & Vanderwart (1980) mean of subjectively rated visual complexity (1–5) and CRL objective visual complexity measure based on the file size of the pictures (in JPG format).
*$p < .01$.

cluding the target name). In addition, following Snodgrass and Vanderwart (1980), we also calculated the *H* statistic (or *H* stat, also called *U statistic*), a measure of response agreement that takes into consideration the proportion of participants producing each alternative. Higher values of *H* indicate lower values of name agreement, and 0 refers to perfect name agreement (see Snodgrass & Vanderwart, 1980, for details).

Percent name agreement measures for each item were based on the four-category lexical code. For each item, Lex 1dom refers to the percentage of all codable responses with a valid RT on which the participants produced the dominant name. *Lex 2phon* is the percentage of all codable responses with valid RTs that were classified as morphological variants of the dominant name. *Lex 3syn* is based on the same denominator and refers to the ratio of codable responses on which a synonym was produced. Finally, *Lex 4oth* refers to the percentage of all codable responses with a valid RT on which the participants produced a response that could not be categorized into Lexical Codes 1–3.

Reaction Time

Several alternative RT measures were calculated. *Mean total RT* refers to mean RTs across all valid trials, regardless of the content of the response. *Mean target RT* refers to mean latency for dominant responses only. *Mean phon2 RT* is the mean latency averaged over responses that were classified as morphological variants of the dominant names. *Mean syn3 RT* refers to the mean RT of valid responses that are synonyms of the target response. Finally, *Mean oth4 RT* refers to the mean RT of valid responses that could not be categorized into Lexical Codes 1–3.

RESULTS

Our methods for eliciting object names differ in several respects from those used in previous studies, including factors such as timing, pace, and number of items administered. For example, both Snodgrass and Vanderwart (1980, untimed written responses) and Snodgrass and Yuditsky (1996, timed oral responses) were

concerned that 260 items would prove too tiring to administer in one session. Hence, they split the full list into two sets of 130 items that were administered to separate groups. By contrast, we administered all 520 items to every object-naming participant in a single fast-paced RT session lasting approximately 45 min (with breaks). We will describe four sets of results to determine whether these methodological differences reduced the validity of our findings and/or their comparability to earlier work. First, we compare our object-naming results with those obtained by Snodgrass and Vanderwart and by Snodgrass and Yuditsky, on the basis of 161 items that overlap in their two experiments (which we will refer to collectively as the Snodgrass studies) and ours (henceforth the Center for Research in Language, or CRL, study). Second, we examine differences between those items that overlap with those of the Snodgrass studies (henceforth *old items*) and the novel items that we added to reach a total of 520 (henceforth *new items*). Third, we examine the effect of order of presentation on both name agreement and RTs in the CRL study, in order to determine whether our results are skewed by fatigue, practice, and/or any shift in strategies that might occur in the course of a 520-item session. Fourth and finally, we examine the reliability of our findings with 520 items by examining Cronbach's alpha over participants.

1. Validating Object-Naming Results Against Findings from Earlier Studies

To compare our results with the classic works by Snodgrass and Vanderwart (1980) and Snodgrass and Yuditsky (1996), we analyzed a subset of 161 object pictures. Of the 174 items that were used in both studies, these were the pictures that also elicited same domi-

nant responses in the Snodgrass studies and in our study. Of the 13 items that did not result in the same target names, 5 were the same for CRL and Snodgrass and Vanderwart (1980) but were dropped in Snodgrass and Yuditsky (1996), and 1 elicited a different target name between Snodgrass and Vanderwart (1980) and Snodgrass and Yuditsky (1996). The remaining 7 elicited different target names in our study in comparison with the Snodgrass studies. We also note in this regard that there were a total of eight discrepancies in the target names elicited by Snodgrass and Vanderwart (1980) versus Snodgrass and Yuditsky (1996), involving 7 pictures that we did not choose for our study. The reasons for such discrepancies are unknown; they could be due to unreliability of the pictures, to historical changes in the language, or perhaps to regional differences in culture and language use.

The mean values for the 161 items that did overlap for the Snodgrass studies and the CRL study are compared in Table 2, for a representative set of dependent and independent variables. Correlations between the Snodgrass studies and the CRL study were also calculated for each of these variables.

Mean values proved to be remarkably similar across studies, despite differences in method and a separation of more than 20 years in data collection. Correlations between the studies for these variables were also significant and relatively large. The strong correlation for visual complexity is especially interesting, since Snodgrass and colleagues assessed visual complexity through subjective ratings, whereas we used an objective measure of complexity on the basis of the file size of the scanned picture stimulus file (see Székely & Bates, 2000, for details).

The basic name agreement measure (percentage of participants producing the dominant name for each item) was highly correlated in the two studies, which is important, since Snodgrass and Vanderwart (1980) used written responses, whereas we (like Snodgrass and Yuditsky, 1996) elicited oral responses with RTs recorded by a voice key. Hence, despite substantial differences in timing and modality, these 161 pictures resulted in a similar degree of name agreement. However, the correlation was somewhat lower for the more lenient name agreement measure. In this regard, note that the lenient measure of name agreement was calculated quite differently in the Snodgrass and CRL studies. In our scoring system, the lenient measure collapsed across Lexical Codes 1–3, including the dominant response, any synonyms of the dominant response, and morphophonological variants of the dominant response (e.g., "thread" for *spool of thread* or "doggie" for *dog*); entries in Lexical Code 4 (other) were excluded from this count. By contrast, the Snodgrass studies included any response produced by at least 2 participants within the lenient measure of naming accuracy. Because of this difference in coding, our lenient measure is much more homogeneous, both semantically and phonologically. This difference may explain why the correlation between studies for the lenient measure was somewhat lower and why the respective lenient naming measures also behave differently in relation to other dependent and independent variables (see below).

With regard to naming latencies, we reached a decision not to trim RTs in our own norming. Given the heterogeneous nature of the populations to which our norms will be applied (children, aphasic patients, and both normal and language-disordered speakers in many languages), we believe that there is no valid a priori basis for deciding on an appropriate trimming cutoff. However, in order to compare our RT results more closely with those of Snodgrass and Yuditsky (1996) for our purposes here, the RTs in Table 2 were recalculated using their trimming procedure. Specifically, the mean RT of the dominant responses was trimmed by eliminating all RTs more than 2 SDs from each item's mean (this refers to values in Table 2 only). When similar trimming procedures are used, we find results for RTs similar to those reported by Snodgrass and Yuditsky. Nevertheless, for all subsequent analyses (see Tables 3–6), we will continue to use the untrimmed results, in line with our general approach to the norming enterprise.

In Table 3, the intercorrelations among the variables above are summarized for each study, again on the basis of the overlapping 161 items (the correlation matrix for CRL is presented in the upper diagonal; the corresponding statistics for Snodgrass & Vanderwart [1980] and Snodgrass & Yuditsky [1996] are presented in the lower diagonal). Although these correlations are all in the same direction, some of them are substantially larger in the CRL data set. This is especially true for correlations

Table 3
Linear Correlations of Variables Within Each Set for the 161 Overlapping Items

	H Stat	NA	LNA	Mean RT	SubjAoA	VC[1]
H statistic	–	−.96*	−.80*	+.72*	+.45*	n.s.
Name agreement	−.95*	–	+.79*	−.63*	−.39*	n.s.
Lenient NA	−.32*	+.40*	–	−.71*	−.40*	n.s.
Mean RT of target[a]	+.41*	−.49*	−.77*	–	+.59*	n.s.
Subjective AoA (1–9)	+.20†	−.28*	−.52*	+.61*	–	n.s.
Visual complexity[b]	n.s.	n.s.	n.s.	+.15‡	+.28*	–

Note—Upper triangle: naming correlation results of CRL. Lower triangle: naming correlation results of the Snodgrass studies. [a]Trimmed mean RT of valid dominant responses for Snodgrass and Yuditsky (1996) but untrimmed mean for CRL. [b]Subjectively rated visual complexity measure in the Snodgrass studies, but objective visual complexity measure in the CRL study. VC: visual complexity. *$p < .01$. †$p < .05$. ‡$p < .1$.

between name agreement and RT. One possible explanation for this difference in magnitude may lie in the fact that Snodgrass and Yuditsky relied on H statistics and name agreement measures from Snodgrass and Vanderwart, comparing the older name agreement variables (based on written responses) with RTs obtained in the newer study (based on oral responses). Hence, their naming and RT measures are taken from separate groups of participants and distinct modalities, whereas our naming and RT means are all taken from the same people, responding in the oral modality.

There is one important difference between our RT results (for both the Snodgrass and the CRL items) and the results and conclusions reported by Snodgrass and Yuditsky (1996). In their discussion of the role of name agreement in naming latency, they report the following:

> ". . . [H]aving a variety of acceptable names for an object does not in and of itself increase the naming time for the most-often used name. Furthermore, naming times for the nondominant name are about the same as naming times for the dominant name." (p. 523)

These conclusions were supported by a comparison of RTs for the 48 pictures that "showed a greater than 20% difference between %Corr$_{lib}$ and %Corr$_{strict}$ and that were given alternate names by at least 3 subjects" (Snodgrass & Yuditsky, 1996, p. 523). These were broken down first into the 32 pictures that had just one alternative name. For these pictures, there was no significant difference in RTs between the dominant and the single nondominant alternatives. Another 16 pictures had two or more dominant names. Although RTs were longer for the nondominant names in this case, the difference did not reach significance.

The conclusion reached by Snodgrass and Yuditsky (1996) is puzzling in view of the rather large differences observed in Table 4, in which means on all major dependent variables are compared for the old items that overlap with theirs and for our new items. It is clear from the table that RTs for the dominant response appear to be considerably shorter than RTs to produce alternative names in any of the other three categories.

To investigate this point further, we compared RTs for the group of items with no alternative names (100% agreement), one alternative name, two alternative names, three alternative names, and four or more alternative names. The distribution of these items is presented in Table 5, which shows that there is indeed a larger concentration of items with many competitors in the new picture set—further evidence that these are harder items ($\chi^2 = 17.66$, $p < .001$). However, within this breakdown we also had enough items within each group to permit a 2 (picture

Table 4
Comparison of the Mean Values of the Overlapping Set
($N = 161$) With the Nonoverlapping Set ($N = 359$)

Variables	Overlapping	Nonoverlapping	$p <$
	Dependent		
%Valid response	96.77%	95.86%	.10
% No response	1.76%	2.58%	.10
% Invalid response	1.47%	1.56%	n.s.
Number of types	2.78	3.61	.01
H statistic	0.495	0.753	.01
% Lex 1dom	89.39%	83.03%	.01
% Lex 2phon	2.62%	4.14%	.05
% Lex 3syn	1.20%	2.88%	.01
% Lex 4oth	6.80%	9.95%	.01
Mean total RT	950	1,082	.01
Mean target RT	932	1,058	.01
Mean phon2 RT	1,154 ($N = 39$)	1,264 ($N = 124$)	n.s.
Mean syn3 RT	1,289 ($N = 19$)	1,345 ($N = 74$)	n.s.
Mean oth4 RT	1,328 ($N = 81$)	1,343 ($N = 249$)	n.s.
	Independent		
Length in syllables	1.78	1.73	n.s.
Length in characters	5.91	5.89	n.s.
Initial frication	28%	28%	n.s.
Ln freq (CELEX)	2.72	2.40	.05
Ln freq (K&F 1967)	2.45	2.15	.10
Ln freq (F&K 1982)	2.67	2.34	.05
Picture goodness ratings	6.03	5.70	.01
Objective AoA	2.07	2.44	.01
Subjective AoA	3.68	4.17	.01
Objective visual complexity	13,056	18,386	.01
Items with shared name	3%	5%	n.s.
Complex words	14%	18%	n.s.

Note—K&F 1967, Kučera and Francis (1967); F&K 1982, Francis & Kučera (1982); Ln freq, log natural transformation of raw frequency counts.

Table 5
Breakdown of Old Items From the Snodgrass Studies Versus New Items by Number of Alternative Names

Number of Alternatives	Old Items ($N = 161$)	New Items ($N = 359$)	Total
0	58	71	129
1	27	62	89
2	27	68	95
3	19	53	72
4–18	30	105	135

source) $\times$ 5 (number of alternatives) between-items analysis of variance (ANOVA) on target naming times. The results included a large main effect of number of alternative names [$F(4,519) = 100.63$, $p < .0001$], reflecting a monotonic increase in RTs with each increase in the number of alternative names. There was also a main effect of picture source, with longer RTs for the newer pictures regardless of number of alternatives [$F(1,519) = 25.07$, $p < .0001$], reflecting the point that we have already made: The new CRL pictures are more difficult overall than the pictures that we used from the Snodgrass studies. However, the interaction did not approach significance ($F < 1$), which means that the monotonic effect of number of competitors on RTs is the same for both picture sets. Despite the absence of a significant interaction, the results are graphed separately for the 161 Snodgrass items and the 359 new CRL items in Figure 1. This figure makes it clear that the same monotonic effects hold for the Snodgrass pictures, at least when they are presented within the context of our larger set of 520 items.

We also asked whether there is a difference in RT between the dominant and nondominant names within items and whether this difference varies as a function of type of response (morphophonological variant, synonym, or other). In our full set of 520 items, there were 32 items that elicited alternative names in all coding categories. For these items (which tend to be especially difficult), mean RTs were 1,143 msec ($SD = 220$) for the dominant response, 1,316 msec ($SD = 269$) for morphophonological variants, 1,381 msec ($SD = 416$) for synonyms, and 1,311 msec ($SD = 389$) for items coded as "other." A within-items ANOVA showed that this difference is significant [$F(3,93) = 4.45$, $p < .009$, Greenhouse–Geisser corrected]. In line with the analyses presented by Snodgrass and Yuditsky (1996), we went on to eliminate the category "other" and restrict the analysis above to items with alternatives that would be considered correct by our lenient code (dominant response, morphophonological variants, and synonyms). Of the 40 items that met this criterion, the mean RTs were 1,105 msec ($SD = 221$) for the dominant responses, 1,308 msec ($SD = 412$) for morphophonological variants, and 1,363 ($SD = 409$) for synonyms, a difference that is significant by a within-items ANOVA [$F(2,78) = 8.28$, $p < .001$, Greenhouse–Geisser corrected]. In short, for these two subsets of admittedly rather difficult items, which elicit alternative names in

several coding categories, the dominant response is significantly easier to produce (as is shown by shorter RTs) than any of the alternatives.

Putting these between- and within-items analyses together, we have to disagree with the conclusions reached by Snodgrass and Yuditsky (1996) regarding the effect of alternative names on naming latency. Our results suggest that alternative names may well be serving as competitors, increasing the time required to produce the dominant response. In fact, this conclusion is compatible with the numerical difference that Snodgrass and Vanderwart (1980) report for 16 pictures that had two or more dominant names. It is likely that we were able to detect these competitor effects more clearly because we have a larger and more variable data set and a somewhat different coding scheme for competing alternatives.

To summarize so far, the CRL naming study and the pioneering studies by Snodgrass and Vanderwart (1980) and Snodgrass and Yuditsky (1996) yield largely comparable results for a set of 161 overlapping pictures and dominant names. Differences in the magnitude of these effects can be explained by variations in method, including the fact that all of our dependent variables were collected orally from the same group of participants. Also, our lenient measure of name agreement was restricted to a more homogeneous set of alternative names (characterized by both semantic and phonological homogeneity) in comparison with the more lenient naming measure based on a more varied set of alternative names that was adopted in the Snodgrass studies. All of our dependent variables are based on the behavioral measures of spoken picture naming performed by the same groups of participants. The one substantive difference between our findings and the conclusions reached by Snodgrass and Yuditsky revolves around "codability" and the effect of alternative names. We find strong evidence that alternative names are associated with longer RTs, in line with models that postulate competitive effects of alternative
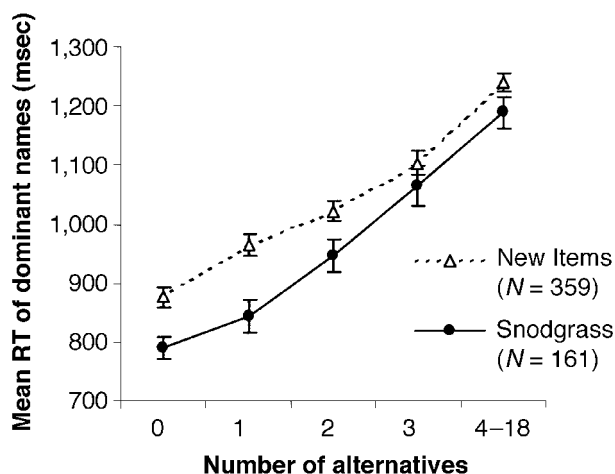

Figure 1. Effects of number of alternative names on target-naming latencies.

names during word selection. These conclusions hold for the old items that were used in the Snodgrass studies as well as for the new items that we added to the stimulus set.

2. Comparing Results for Old and New Items

Snodgrass and Vanderwart's (1980) materials have been used successfully in a large number of studies, and it is clear from the accumulated results that most of their stimuli are "good" items that are readily recognized by native speakers in the languages in which they have been used. By raising the number of picture stimuli from 260 to 520, we ran a risk of adding items that were much lower in nameability and that might differ from the original set in other ways. It was expected that the variance would be substantially larger along many dimensions that are important for lexical access, in comparison with that of the Snodgrass studies based on the smaller set. First, we asked whether the subset of 161 pictures from the Snodgrass studies (validated above) differ systematically from the remaining 359 items of our larger picture set, by comparing means on major dependent and independent variables using the conventional independent-sample t test (Table 4).

Naming was indeed significantly faster in our study for the 161 Snodgrass items, which also elicited fewer alternatives. The concepts elicited by the Snodgrass subset were also acquired earlier (as measured by both subjective and objective AoAs) and were rated as easier to depict. The Snodgrass pictures proved to be less complex than the rest of the CRL pictures, on the basis of our own objective measure of visual complexity. The dominant responses for the Snodgrass subset were significantly higher in frequency, yet they were the same length and shared the same word characteristics as the remaining 359 dominant responses in the CRL stimulus set. In short, the Snodgrass items are easier, but acceptable levels of agreement and RT are obtained for most of the new items as well. It seems that, in the initial development of their widely used stimulus set, Snodgrass and colleagues selected object concepts that are especially easy to depict. The items that we added to this picture set are still valid and useful, but they are, on average, more difficult.

We also asked whether our enlargement of the picture set resulted in a shift in conceptual content, which could, in turn, result in differences in both word and picture characteristics. For this purpose, all the picture items were classified into one of nine semantic categories: people, animals (including insects), body parts of people or animals (e.g., eye, foot, mouth, paw), vehicles, foods, things to wear, small artifacts (which people would normally lift and operate with their hands), large artifacts (including large appliances, furniture, buildings, or other constructions), and objects found in nature (e.g., flower, tree, volcano). Table 6 presents the numbers and percentages of items within each semantic category, for the Snodgrass set and for the additional CRL items.

The distribution of old (Snodgrass) and new (CRL) items across these nine semantic categories was significantly different ($\chi^2 = 20.5$, $p < .01$). In comparison with the items from the Snodgrass studies, the new CRL items contain a lower proportion of animals and foods, a higher proportion of people or human-like characters (e.g., genie, snowman), and large artifacts. A univariate ANOVA was performed to determine whether there was a significant interaction of semantic category and picture source (the 161 old vs. the 359 new pictures) on naming behavior. This analysis yielded significant main effects of both semantic category and picture source on both total and target RTs, but no significant interaction. There was also a main effect of picture source on name agreement, but no effect of semantic category (that is, none of the categories was particularly "nameable") and no interaction. We may conclude that RTs are longer and name agreement is lower for the items that we added to the Snodgrass picture set, and this expansion involved a search within categories that had lower proportional representation in the original Snodgrass set. However, it appears that the conceptual content was not skewed in ways that interact with naming difficulty (either in terms of name agreement or in terms of RT). An informal inspection of the dominant responses elicited by old versus new items indicates that the two response sets tended to occur at the same basic level of picture description. (See the dominant response for each picture at http://www.crl.ucsd.edu/~aszekely/ipnp/method.html.) It was not the case that

Table 6
Distribution of Items in the Nine Semantic Categories Within the Overlapping Set ($N = 161$) and the Nonoverlapping Set ($N = 359$)

Category	N (Full Set)	% (Set of 161)	% (Set of 359)
People	27	1	7
Animals	81	20	14
Body parts	28	5	5
Vehicles	35	6	7
Foods	46	13	7
Things to wear	39	7	8
Small artifacts	163	35	30
Large artifacts	77	10	17
Objects and phenomena in nature	24	3	5
Total	520	100	100

speakers resorted more often to vague superordinate terms (e.g., "animal") for the new items, nor was there a noteworthy increase in the use of subordinate terms (e.g., "Scotch Terrier" for *dog*). Our additional items are harder, but they are named according to similar principles.

3. Effects of Session Length

Another aspect of our expanded picture set is the marked increase in session length, which could result in a build-up of fatigue, practice effects, and/or other strategies in the course of the session. On the basis of earlier work suggesting that participants became fatigued after approximately 100 items, Snodgrass and Yuditsky (1996) split their picture set in half, with individual participants naming only 130 items presented in a single session with an average duration of 30 min. In the CRL study, the participants were asked to name 520 pictures (four times the number named in Snodgrass and Yuditsky, 1996), but timing parameters were much faster. Hence, our sessions lasted approximately 45 min—50% longer than the sessions in Snodgrass and Yuditsky (1996)—although individual participants named four times as many pictures.

To investigate the effect of session length and order of presentation, we carried out a regression analysis of mean accuracy and mean RT values of naming events as a function of their order of appearance in the session (note that each time point reflects the contribution of 10 different pictures from 10 different randomized lists). For mean RT of all valid responses, R^2 was .211 [mean RT = 958.1 + 0.286 (item position)]. For the mean RT of valid responses, R^2 = .214 [mean RT = 921.4 + 0.268 (item position)]. Analogous regression analyses failed to reach significance for percent dominant response or for percentage of participants producing any codable response with a valid RT. In other words, there is a significant increase in naming latency across the session (an average increase of 100 msec across the session), but accuracy levels are unaffected by session length. The mean

RTs of all valid responses to the pictures are presented as a function of item position in Figure 2. There is a small but reliable drift toward longer latencies spread evenly across the session (including the first 100 items, the number used by Snodgrass & Yuditsky, 1996). The RT effect justifies our decision to use at least 10 different random orders in the naming experiment, whereas the absence of a linear relationship between order and accuracy testifies to the validity of our within-participants, single-session design.

4. Reliability

Finally, we conducted a reliability analysis over participants to assess the stability of individual naming performance that might interfere with our item-based data analysis. Cronbach's alpha values were computed across participants for the two main RT measures (missing values were replaced by item means). The reliability statistics were high (alpha = .97 for RT across all valid trials, and alpha = .96 for RT values of dominant responses only). These results attest to the reliability of an item-based analysis of the naming results, collapsed over participants.

SUMMARY AND FUTURE DIRECTIONS

In the CRL picture-naming norming project, we have used a much larger set of items (520 pictures) than had been adopted in previous norming projects, and we administered all these items in a single session (averaging 45 min in length, with breaks). In the present study, we validated our methodology against the important early work by Snodgrass and Vanderwart (1980) and Snodgrass and Yuditsky (1996), and we have shown that comparative results are obtained for 161 overlapping items even though these items were (in our study) embedded within a much larger set. To the extent that our findings differ from those of the Snodgrass studies, small varia-
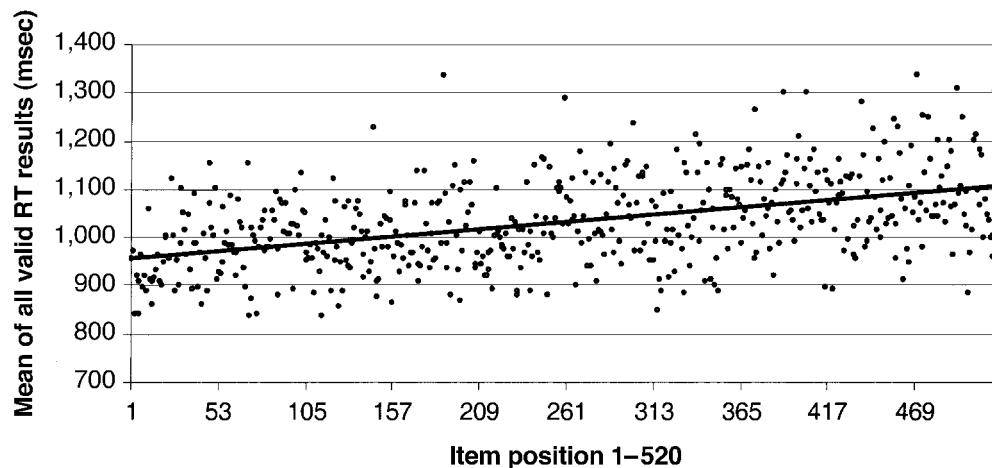

Figure 2. Reaction time (RT) as a function of item position.

tions in methodology appear to be sufficient to explain the difference. For example, we used more homogeneous criteria for lenient name agreement in our study, which led to higher correlations between lenient name agreement and other dependent and independent variables.

The largest discrepancy between our study and the earlier reports by Snodgrass and colleagues revolves around the relationship between RTs and name agreement (i.e., number of alternative names). On the basis of analyses of a relatively small subset of pictures (those with only one competitor and those with two or more competitors), Snodgrass and Yuditsky (1996) concluded that there is no disadvantage for items with more competing alternatives and that the dominant name is no faster than alternative names for the same picture. Our results contrast sharply with this conclusion. We find a strong monotonic increase in RTs as a function of number of alternative names, an effect that is apparent both in the 161 old items that we share with the Snodgrass studies and in the 359 new items added to the stimulus set. Furthermore, within two subsets of items that had alternative names in several coding categories, we find that the dominant response is faster than any of the three alternative-name categories (morphophonological variants, synonyms, and other). In fact, the results leaned in this direction for the 16 items in the original Snodgrass and Yuditsky study that elicited two or more alternative names. We suspect that the difference between their findings and ours reflects the advantages that we gain by using a larger and more varied set of picture stimuli.

Within our own data set, the 161 items that overlapped with the Snodgrass set did prove to be easier (as is shown by higher name agreement and shorter RTs) than the 359 items that we added from other sources. Clearly, Snodgrass and Vanderwart (1980) and Snodgrass and Yuditsky (1996) had the wisdom to choose those concepts that are particularly accessible and easy to depict. However, the expanded set does appear to elicit acceptably high levels of name agreement and naming latency. Reliability measures yielded high levels of agreement, testifying to the internal reliability of the picture norms. We did find an increase in RTs across the course of the session and a gradual drift (i.e., no specific points at which fatigue set in) that justifies our decision to use 10 different randomized orders. However, order of presentation had no effect on any of the name agreement measures, suggesting that the same names (and range of names) are given at any point in the session.

This validation study serves as the cornerstone for a larger enterprise, opening the way for a wider range of studies of both object and action naming, in and out of context, across the range of languages that we had in mind when we undertook this project (see, e.g., Bates et al., 2003; Bates et al., 2000). We have completed a study comparing the data presented here on naming of objects with data on naming of 275 black and white drawings of transitive and intransitive actions by a different group of English speakers (Székely et al., in press). The CRL database is currently being used to develop stimuli for presentation in fMRI studies and event-related brain potential studies of both children and adults, in English, Italian, and Chinese (Dick et al., 2001). Other colleagues are using these stimuli in fMRI studies of Spanish–English bilinguals (Hernandez, Dapretto, Mazziotta, & Bookheimer, 2001), using covert picture naming in single-language presentation (Spanish only or English only) in comparison with picture naming in mixed-language presentation. The potential for interlanguage competition in bilingual picture naming has also been assessed with a subset of these stimuli in lifespan studies (from 5 to 85 years of age: Kohnert, 2000; Kohnert et al., 1999; Kohnert et al., 1998), including studies focused specifically on the effects of cognitive aging on the bilingual adult's ability to avoid interlanguage interference. Other bilingual offshoots of this project include (1) an objective study of cognate status (obtained by testing monolingual participants in a language they do not know in a picture–word verification paradigm, establishing the "guessability" of picture names based on cognate status), (2) a new study of AoAs, compared side by side for each language, in which norms are obtained from bilingual participants in both their languages, and (3) new measures of relative fluency or dominance in which bilinguals are asked to name the pictures in each language (in counterbalanced blocks) under speeded or visually degraded conditions, and to judge the match or mismatch between pictures and words using perceptually degraded auditory stimuli.

Developmental studies using these stimuli and the associated database are already under way, including one published study (D'Amico et al., 2001) in which similarities and differences in object naming of Italian-speaking 5-year-old children and adults were examined for a subset of 250 pictures that met various criteria for use with young children. In ongoing studies, we have started to compare timed picture naming and word reading for the same subsets of stimuli (Bates, Burani, et al., 2001), including preliminary results indicating that individual rates of speed and accuracy of picture naming may be excellent predictors of entry into literacy.

We are also using this database to compile a new timed naming test for aphasic patients, to explore dissociations between object naming (reported to be better preserved in nonfluent Broca's aphasics) and action naming (reported to be better preserved in fluent Wernicke's aphasics and some anomics—see Chen & Bates, 1998, for a review). This is a parallel effort in English, Italian, Spanish, and Chinese, part of the continuing project "Cross-Linguistic Studies of Aphasia." In this regard, we are particularly interested in studies showing that the kinds of naming errors produced by aphasic patients can also be elicited in normal adults forced to name pictures under perceptual degradation (Laws, Leeson, & Gale, 2002) or under increasing time pressure (Vitkovitch & Humphreys, 1991; Vitkovitch, Humphreys, & Lloyd-Jones, 1993). In the present study, the partici-

pants were asked to name the stimuli as fast as possible without making mistakes, and parameters were set to avoid inducing a speed/accuracy trade-off. However, we acknowledge that the data obtained with these timing parameters may differ markedly from findings under a more severe time pressure. The study of naming behavior under nonoptimal processing conditions is a promising avenue for comparative studies of normal and impaired naming behavior, within and across languages.

Finally, as we had hoped when we launched this project 3 years ago, these naming norms have already proven extremely useful in designing studies to examine sentential and grammatical priming effects on word production (as assessed through picture naming in context) and word comprehension (as assessed through cued repetition of the same target names, or through picture–word verification). Multiple projects using the picture norms in priming studies are already under way.

We plan to put the complete database (including all the item-specific characteristics that we have been able to derive for each language) on our website (http://crl.ucsd. edu/~aszekely/ipnp/), where these norms can be accessed and used, at no cost, by investigators around the world. We hope to convince our colleagues in other countries to obtain similar norms for their languages, and we will cooperate in every way we can to facilitate this process by providing digitized visual stimuli and PsyScope scripts to any interested users.

REFERENCES

ALCOCK, K. J., & NGOROSHO, D. (in press). Grammatical noun class agreement processing in Kiswahili. *Language & Speech*.

BAAYEN, R. H., PIEPENBROCK, R., & GULIKERS, L. (1995). *The CELEX Lexical Database (Release 2)* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

BARRY, C., HIRSCH, K. W., JOHNSTON, R. A., & WILLIAMS, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory & Language*, **44**, 350-375.

BARRY, C., MORRISON, C. M., & ELLIS, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency and name agreement. *Quarterly Journal of Experimental Psychology*, **50A**, 560-585.

BATES, E., BURANI, C., D´AMICO, S., & BARCA, L. (2001). Word reading and picture naming in Italian. *Memory & Cognition*, **29**, 986-999.

BATES, E., D´AMICO, S., JACOBSEN, T., SZÉKELY, A., ANDONOVA, E., DEVESCOVI, A., HERRON, D., LU, C. C., PECHMANN, T., PLÉH, C., WICHA, N., FEDERMEIER, K., GERDJIKOVA, I., GUTIERREZ, G., HUNG, D., HSU, J., IYER, G., KOHNERT, K., MEHOTCHEVA, T., OROZCO-FIGUEROA, A., TZENG, A., & TZENG, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, **10**, 344-380.

BATES, E., DEVESCOVI, A., & WULFECK, B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology*, **52**, 369-398.

BATES, E., FEDERMEIER, K., HERRON, D., IYER, G., JACOBSEN, T., PECHMANN, T., D´AMICO, S., DEVESCOVI, A., WICHA, N., OROZCO-FIGUEROA, A., KOHNERT, K., GUTIERREZ, G., LU, C. C., HUNG, D., HSU, J., TZENG, O., ANDONOVA, E., GERDJIKOVA, I., MEHOTCHEVA, T., SZÉKELY, A., & PLÉH, C. (2000). Introducing the CRL International Picture-Naming Project (CRL-IPNP). *Center for Research in Language Newsletter, 12(1)*. La Jolla: University of California, San Diego. Available at http://crl.ucsd.edu/newsletter/12-1/article.html.

BENTROVATO, S., DEVESCOVI, A., D´AMICO, S., & BATES, E. (1999). The effect of grammatical gender and semantic context on lexical access in Italian. *Journal of Psycholinguistic Research*, **28**, 677-693.

CARROLL, J. B., & WHITE, M. N. (1973a). Age of acquisition norms for 220 pictureable nouns. *Journal of Verbal Learning & Verbal Behavior*, **12**, 563-576.

CARROLL, J. B., & WHITE, M. N. (1973b). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, **25**, 85-95.

CHEN, S., & BATES, E. (1998). The dissociation between nouns and verbs in Broca's and Wernicke's aphasia: Findings from Chinese. Special issue on Chinese Aphasia. *Aphasiology*, **12**, 5-36.

COHEN, J., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, **25**, 257-271.

CYCOWICZ, Y. M., FRIEDMAN, D., ROTHSTEIN, M., & SNODGRASS, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, **65**, 171-237.

D´AMICO, S., DEVESCOVI, A., & BATES, E. (2001). Picture naming and lexical access in Italian children and adults. *Journal of Cognition & Development*, **2**, 71-105.

DELL´ACQUA, R., LOTTO, L., & JOB, R. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers*, **32**, 588-615.

DICK, F., SACCUMAN, C., SERENO, M., MÜLLER, R.-A., BATES, E., & WULFECK, B. (2001, March). *Language production and comprehension in fMRI: Consistency and variability over individuals and group averages*. Poster presented at the Meeting of the Society for Cognitive Neuroscience, New York.

DRUKS, J. (2002). Verbs and nouns—a review of the literature. *Journal of Neurolinguistics*, **15**, 289-315.

FEDERMEIER, K. D., & BATES, E. (1997). Contexts that pack a punch: Lexical class priming of picture naming. *Center for Research in Language Newsletter, 11(2)*. La Jolla: University of California, San Diego. Available at http://crl.ucsd.edu/newsletter/11-2/.

FENSON, L., DALE, P., REZNICK, J., BATES, E., THAL, D., & PETHICK, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, **59** (5, Serial No. 242).

FEYEREISEN, P., DEMAEGHT, N., & SAMSON, D. (1998). Why do picture-naming latencies increase with age: General slowing, greater sensitivity to interference, or task-specific deficits? *Experimental Aging Research*, **24**, 21-51.

FORSTER, K. I., & CHAMBERS, I. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior*, **12**, 627-635.

FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

GILHOOLY, K. J., & GILHOOLY, M. L. (1979). Age-of-acquisition effects in lexical decision and episodic memory tasks. *Memory & Cognition*, **7**, 214-223.

HERNANDEZ, A. E., DAPRETTO, M., MAZZIOTTA, J., & BOOKHEIMER, S. (2001). Language switching and language representation in Spanish–English bilinguals: An fMRI study. *NeuroImage*, **14**, 510-520.

HERNANDEZ, A. E., MARTINEZ, A., & KOHNERT, K. (2000). In search of the language switch: An fMRI study of picture naming in Spanish–English bilinguals. *Brain & Language*, **73**, 421-431.

HUMPHREYS, G. W., RIDDOCH, M. J., & QUINLAN, P. T. (1988). Cascade process in picture identification. *Cognitive Neuropsychology*, **5**, 67-104.

IYER, G., SACCUMAN, C., BATES, E., & WULFECK, B. (2001). A study of age-of-acquisition (AoA) ratings in adults. *Center for Research in Language Newsletter, 13(2)*. La Jolla: University of California, San Diego. Available at http://crl.ucsd.edu/newsletter/13-2/article.html.

JACOBSEN, T. (1999). Effects of grammatical gender on picture and word naming: Evidence from German. *Journal of Psycholinguistic Research*, **28**, 499-514.

JESCHENIAK, J. D., & LEVELT, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and phonological form. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 824-843.

JOHNSON, C. J., PAIVIO, A., & CLARK, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, **120**, 113-139.

KOHNERT, K. (2000). *Lexical skills in bilingual school-age children: Cross-sectional studies in Spanish and English.* PhD dissertation, University of California, San Diego, and San Diego State University.

KOHNERT, K., BATES, E., & HERNANDEZ, A. (1999). Balancing bilinguals: Lexical-semantic production and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, & Hearing Research*, **42**, 1400-1413.

KOHNERT, K., HERNANDEZ, A., & BATES, E. (1998). Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain & Language*, **65**, 422-440.

KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

LACHMAN, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, **99**, 199-208.

LACHMAN, R., LACHMAN, J. L., THRONESBERY, C., & SALA, L. S. (1980). Object salience and code separation in picture naming. *Bulletin of the Psychonomic Society*, **16**, 187-190.

LACHMAN, R., SHAFFER, J. P., & HENNRIKUS, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency, and age-of-acquisition on lexical RT. *Journal of Verbal Learning & Verbal Behavior*, **13**, 613-625.

LAWS, K., LEESON, V., & GALE, T. (2002). The effect of "masking" on picture naming. *Cortex*, **38**, 137-148.

LU, C.-C., BATES, E., HUNG, D., TZENG, O., HSU, J., TSAI, C.-H., & ROE, K. (2002). Syntactic priming of nouns and verbs in Chinese. *Language & Speech*, **44**, 437-471.

MORRISON, C. M., CHAPPELL, T. D., & ELLIS, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, **50A**, 528-559.

MORRISON, C. M., ELLIS, A. W., & QUINLAN, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, **20**, 705-714.

MURTHA, S., CHERTKOW, H., BEAUREGARD, M., & EVANS, A. (1999). The neural substrate of picture naming. *Journal of Cognitive Neuroscience*, **11**, 399-423.

NATION, K., MARSHALL, C. M., & SNOWLING, M. J. (2001). Phonological and semantic contributions to children's picture naming skill: Evidence from children with developmental reading disorders. *Language & Cognitive Processes*, **16**, 241-259.

NISI, M., LONGONI, A. M., & SNODGRASS, J. G. (2000). Italian measurement on the relation of name, familiarity, and acquisition age for the 260 figures of Snodgrass and Vanderwart. *Giornale Italiano di Psicologia*, **27**, 205-218.

OLDFIELD, R. C., & WINGFIELD, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, **17**, 273-281.

PRESTON, K. A. (1935). The speed of word perception and its relation to reading ability. *Journal of General Psychology*, **13**, 199-203.

ROCHFORD, G., & WILLIAMS, M. (1962a). Studies in the development and breakdown of the use of names: I. The relationship between nominal dysphasia and the acquisition of vocabulary in childhood. *Journal of Neurology, Neurosurgery & Psychiatry*, **25**, 222-227.

ROCHFORD, G., & WILLIAMS, M. (1962b). Studies in the development and breakdown of the use of names: II. Experimental production of naming disorders in normal people. *Journal of Neurology, Neurosurgery & Psychiatry*, **25**, 228-233.

SANFELIU, M. C., & FERNANDEZ, A. (1996). A set of 254 Snodgrass–Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers*, **28**, 537-555.

SCHMITT, B. M., MÜNTE, T. F., & KUTAS, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology*, **37**, 473-484.

SCHMITT, B. M., SCHILTZ, K., ZAAKE, W., KUTAS, M., & MÜNTE, T. F. (2001). An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. *Journal of Cognitive Neuroscience*, **13**, 510-522.

SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 174-215.

SNODGRASS, J. G., & YUDITSKY, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, **28**, 516-536.

SZÉKELY, A., & BATES, E. (2000). Objective visual complexity as a variable in studies of picture naming. *Center for Research in Language Newsletter*, *12(2)*. La Jolla: University of California, San Diego. Available at http://www.crl.ucsd.edu/newsletter/12-2/article.html.

SZÉKELY, A., D´AMICO, S., DEVESCOVI, A., FEDERMEIER, K., HERRON, D., IYER, G., JACOBSEN, T., ARÉVALO, A., VARGHA, A., & BATES, E. (in press). Timed action and object naming. *Cortex*.

THORNDIKE, E. L. (1931). *A teacher's word book of the twenty thousand words found most frequently and widely in general reading for children and young people.* New York: Columbia University, Teachers College.

VAN TURENNOUT, M., HAGOORT, P., & BROWN, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 787-806.

VAN TURENNOUT, M., HAGOORT, P., & BROWN, C. M. (1999). The time course of grammatical and phonological processing during speaking: Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, **28**, 649-676.

VITKOVITCH, M., & HUMPHREYS, G. (1991). Perseverant responding in speeded naming of pictures: It's in the links. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 664-680.

VITKOVITCH, M., HUMPHREYS, G., & LLOYD-JONES, T. (1993). On naming a giraffe a zebra: Picture-naming errors across different object categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 243-259.

WICHA, N. Y. Y., BATES, E., OROZCO-FIGUEROA, A., REYES, I., HERNANDEZ, A., & GAVALDÓN DE BARRETO, L. (2001). *When zebras become painted donkeys: Interplay between gender and semantic priming in a Spanish sentence context.* Manuscript submitted for publication.

NOTE

1. The unique picture numbers for each of the relevant pictures are presented in parentheses.