Published in final edited form as:

Behav Res Methods. 2012 December ; 44(4): 1028–1041. doi:10.3758/s13428-012-0215-z.

Perceptual and Motor Attribute Ratings for 559 Object Concepts

Ben D. Amsel, Thomas P. Urbach, and Marta Kutas

Abstract

To understand how and when object knowledge influences the neural underpinnings of language comprehension and linguistic behavior, it is critical to determine the specific kinds of knowledge people have. To extend currently available normative data, we report a relatively more comprehensive set of object attribute rating norms for 559 concrete object nouns, each rated on seven attributes corresponding to sensory and motor modalities: color, motion, sound, smell, taste, graspability, and pain, in addition to familiarity (376 raters, mean 23 raters per item). Mean ratings were subjected to principal component analysis, revealing two primary dimensions plausibly interpreted as relating to survival. We demonstrate the utility of these ratings in accounting for lexical and semantic decision latencies. These ratings should prove useful for the design and interpretation of experimental tests of conceptual and perceptual object processing. The complete stimuli and norms may be downloaded from <http://brm.psychonomic-journals.org/content/supplemental>.

The representation of object concepts in long-term memory and the recruitment of this knowledge during language comprehension have long been central topics in cognitive science, and continue to receive considerable attention (e.g., Binder, Desai, Graves, & Conant, 2009; Martin, 2007). Our knowledge of objects consists of several kinds of information, many of them (but not all) perceivable through the senses (e.g., how an object looks, moves, tastes, and feels). Converging evidence suggests that object concepts are not represented in a unitary brain region, but are instead distributed across several brain regions including but not necessarily limited to sensory and motor cortex (Martin, 2007; Patterson, Nestor, & Rogers, 2007). Current research about these issues includes assessments of knowledge retrieval of different object properties during language comprehension (Amsel, 2011; Kan, Barsalou, Solomon, Minor, & Thompson-Schill, 2003; Kellenbach, Brett, & Patterson, 2001), and how task-related context flexibly modulates activation of object knowledge (Grossman, Koenig, Kounios, McMillan, Work, & Moore, 2006; Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008). These types of experiments typically rely on the specification of one or more aspects of the content of semantic representations. If a researcher hypothesizes that verifying an object's color versus its shape would produce meaningful differences in behavioral and/or brain-based dependent measures, she would need to specify the color and shape of several objects in preparation for the experiment. If an experimenter aims to delineate the time course of neural activity involved in deciding whether an object is colorful versus loud, he must have measures of colorfulness and loudness for stimulus selection.

In this report we provide ratings of eight object attributes for a large set of concrete nouns, as well as averaged response times associated with each attribute and each item. Our norms extend previous sets of object attribute ratings by a) incorporating a measure of response time for each attribute, b) utilizing a larger than typical set of words, and c) including not only standard perceptual attributes (e.g., color) but also less studied attributes (e.g., likelihood of pain, taste pleasantness). The inclusion of these attributes is important for

researchers interested in the full gamut of sensory modalities, and could motivate additional study of modalities that have received relatively less attention. We conducted a principal component analysis on the ratings revealing two major latent sources of variance. We found that certain of the ratings predict novel and unique portions of variance in decision latencies from previously reported lexical and concreteness tasks, highlighting the potential for the ratings to capture hitherto relatively unexplored kinds of semantic knowledge.

We now briefly review two major approaches to the specification of semantic content, namely, collection of feature norms and object attribute ratings. Feature production norms are generated by asking participants to list attributes of a given concept (e.g., <is red>, <used for cooking>), and retaining only attributes listed by at least 2-3 participants (e.g., McRae, Cree, Seidenberg, & McNorgan 2005; Vinson & Vigliocco, 2008). These datasets have been used, for example, to show that concepts with greater numbers of listed features are processed more quickly (e.g., Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008), and how feature correlations influence the organization of concepts in semantic memory (McRae, de Sa, & Seidenberg, 1997). Semantic features also have been categorized by knowledge type (e.g., visual, olfactory, encyclopedic; Cree & McRae, 2003; Wu & Barsalou, 2009) and used to assess the influences of different knowledge types on behavioral performance and neural activity (Amsel, 2011; Grondin, Lupker, & McRae, 2009). For example, Grondin et al. (2009) found that the number of shared features belonging to several different knowledge types could account for significant unique variance in lexical and concreteness decision tasks. Finally, at least two research groups have taken a somewhat different approach to semantic feature norming, whereby participants rate the degree to which a feature is experienced by each of the five senses (Lynott & Connell, 2009; van Dantzig, Cowell, Zeelenberg, & Pecher, 2011). From these data the authors compute a measure of modality-exclusivity—that is, the degree to which a semantic feature is experienced by a single sensory modality.

Another approach to revealing the content of object concepts is to ask participants to provide numeric or categorical ratings of various object criteria. This approach is less well-defined than feature norming; the purpose of discussing the studies in this section is to show that object attribute ratings are used extensively in perception and language experiments, which in turn motivates our collection of a single large-scale set of attribute ratings that span many of the above knowledge types. Oliver and colleagues (Oliver, Geiger, Lewandowski, & Thompson-Schill, 2009; Oliver & Thompson-Schill, 2003), for example, asked participants to rate object concepts on their shape, color, size, and tactile properties, and used these data to demonstrate modality-specific neural activation in ventral and dorsal processing streams during language comprehension. Moscoso del Prado Martin, Hauk, and Pulvermüller (2006) asked participants to make three judgments on a set of English words: “Does this word remind you of something you can visually perceive / a particular color / a particular form or visual pattern?” They found differences in event-related brain potential (ERP) amplitudes beginning at 200 ms to words rated high on color versus form-relatedness, taken to suggest rapid access (and differentiation) of semantic information during word recognition. Kellenbach et al. (2001) used objects that were colored or black and white, could or could not make noise spontaneously, and were obviously small or large in a positron emission tomography (PET) study to demonstrate activation of modality-specific cortex during retrieval of each kind of knowledge. González and colleagues (2006) asked participants to rate words on the degree they referred to objects with a strong smell, and found odor-related words (e.g., ‘garlic’) activated distributed circuits including typical language areas as well as primary olfactory cortex. Taken as a whole, these studies highlight the importance of specifying sensory-based semantic content for understanding how modality-specific processing is engaged by linguistic stimuli.

In addition to sensory-based content, several groups have collected ratings of different aspects of human-object interaction. Magnié, Besson, Poncet, and Dolisi (2003) had participants rate the degree an object could be uniquely pantomimed. Campanella, D'Agostini, Skrap, & Shallice (2010) used these manipulability ratings to show that participants with damage to posterior middle temporal gyri had particular difficulty with naming objects that were highly manipulable—consistent with sensory/motor models of semantic memory. They subsequently showed an explicitly semantic influence of manipulability in word-to-picture matching tasks, and argued manipulability should be considered a semantic dimension (Campanella & Shallice, 2011). Salmon, McMullen and Filliter (2010) argued that manipulability should be subdivided into the independent dimensions of graspability, and functional usage. Consistent with their claim they found that ratings for each of these dimensions were uncorrelated.

Whereas the above studies largely concern the interaction of objects and finger, hand, and arm effectors, body-object interaction (BOI) ratings (Bennett, Burnett, Siakaluk, & Pexman, 2011; Tillotson, Siakaluk, & Pexman, 2008; Siakaluk, Pexman, Aguilera, Owen, & Sears, 2008a; Siakaluk, Pexman, Sears, Wilson, Locheed, & Owen, 2008b) are designed to index the extent that a person interacts with an object using any part of their body. Siakaluk and colleagues (2008a, 2008b) found that words with higher BOI values are responded to more quickly in lexical and semantic decision tasks even after controlling for imageability and concreteness. Whereas BOI ratings are thought to specifically index physical interactions with objects, Juhasz, Yap, Dicke, Taylor, and Gullick (2011) collected Sensory Experience Ratings (SER), designed to reflect the degree a word evokes any kind of sensory experience. Importantly, although SER were correlated with imageability, they predicted lexical decision latencies in a large data set with imageability controlled. These studies suggest that information initially learned via motor interaction with objects may be recruited not only in the service of perception and action, but also during lexical and semantic tasks.

In addition to their utility in designing and interpreting controlled experiments, empirically-derived semantic content also has enabled important advances in the development of distributional models of word meaning. Johns and Jones (2011) developed a distributional model that initially contained linguistic information derived from large text corpora and perceptual information derived from feature norms (i.e. Lynott & Connell, 2009; McRae et al., 2005; Vinson & Vigliocco, 2008), but was able to infer the 'perceptual' representations of all words in its 'memory' from the human-generated features available for a small subset of those words. Interestingly, their model was also able to predict the dominant sensory modalities of a new set of words. Another advance is due to Andrews, Vigliocco, and Vinson (2009), who created a probabilistic Bayesian model that treats distributional and experiential data as a unitary joint distribution. Their model accounts for several behavioral measures (e.g., picture naming and lexical decision latencies) more accurately than models trained on either distributional or experiential data alone. Important for present purposes, the innovation of these models is made possible in part by human-derived content.

At least one group has collected a set of object attribute ratings encompassing a variety of knowledge types. The Wisconsin Perceptual Attribute Ratings Database (Medler, Arnoldussen, Binder, & Seidenberg, 2005) consists of four types of perceptual ratings (sound, color, manipulation, and motion) and an emotional valence rating for 1402 words ranging from very abstract ("advantage") to very concrete ("airplane"). Three hundred forty-two participants used an online form to rate how important each perceptual attribute was to the meaning of each word on a 7-point scale from "not at all important" to "very important." The current study builds upon this dataset and the work presented above by including several additional attributes, providing response times for each kind of rating, and

demonstrating the utility of our norms in accounting for decision latencies in lexical and semantic decision-making.

Current study

The main purpose of the current study is to provide a relatively more comprehensive source of information about several object attributes for use in psycholinguistic, cognitive, perceptual, and computational research. Rather than relying on categorical judgments of object knowledge, we assess each of the above dimensions on a scale ranging from 1 to 8, which upon averaging becomes a near-continuous rating scale. Our motivations for examining the present eight types of attributes are based on a determination of use in previous research, and our aim to include a more comprehensive set of measures than previous norms. Each of the five traditional Aristotelian sensory modalities (vision, touch, hearing, smell, and taste) is represented, in addition to the sensation of pain. We assess two kinds of visual knowledge, color and motion, which are represented in different brain regions proximal to corresponding sensory cortex (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Simmons, Ramjee, Beauchamp, McRae, Martin, & Barsalou, 2007). Ratings of taste and smell intensity are anticipated to be highly redundant, which motivates the collection of separate intensity and pleasantness judgments in the olfactory and gustatory domains, respectively (c.f., De Araujo, Rolls, Kringelbach, McGlone, & Phillips, 2003). Tactile object information is assessed with graspability judgments, which reflect knowledge of physical object properties and learned sensorimotor programs. The motivation for this dimension is based on the importance of grasping behavior in our interaction with the environment, and the sustained research focus on its neural substrates (Chao & Martin, 2000; Davare, Kraskov, Rothwell, & Lemon, 2011; Goodale, Meenan, Bulthoff, Nicolle, Murphy, & Racicot, 1994). Last but not least, we assess the likelihood that each object would cause the perception of pain, which is usually triggered by activation of specific nociceptors (Millan, 1999). Like other senses, the ability to sense pain may be adaptive—congenital insensitivity to pain is linked to shorter life expectancy (Nagasako, Oaklander, & Dworkin, 2003).

With the mean attribute ratings in hand, we examine the distributions and response times associated with each. We conduct a principal component analysis to uncover shared variance among attributes, revealing two major sources of shared variance readily interpretable as related to survival. Finally, we utilize the ratings to account for portions of unique variance in published decision latencies in a concreteness and lexical decision task, which reveals multiple attributes as successful predictors of decision latency.

Method

Participants

Four hundred and twenty undergraduate students (308 female; 109 male; 3 declined to state) were recruited from the departments of Psychology, Linguistics, and Cognitive Science at the University of California, San Diego and were awarded course credit upon successful completion of the experiment. Participants were native English speakers between 18 and 30 years of age ($M = 20.7$, $SD = 1.8$), had completed on average 15 years of education, and reported normal vision and no major neurological or general health problems. Three hundred seventy seven participants were right-handed, 33 were left-handed, and the remainder declined to state.

Stimuli

Nouns—Each of the 560 normed words corresponds to an English noun denoting an object concept. Nouns were chosen primarily from the two largest existing feature production norms for concrete nouns (McRae et al., 2005; Vinson & Vigliocco, 2008), and included 47 additional nouns chosen by the experimenters. We endeavored to include a wide range of nouns that have been used in previous psycholinguistic experiments and will be most likely to serve in future experiments. We included exemplars of several common categories (i.e., buildings, creatures, fruits & vegetables, places, plants, musical instruments, tools, vehicles).

Attribute ratings—Appendix A contains the full text for each question. The rating scale for each question was pegged to two labels anchoring either extreme of the scale (i.e., 1 and 8). Eight response choices were chosen because the most reliable ratings data are typically obtained from scales with between 6 and 10 response options (Preston & Colman, 2000; Weng, 2004). An even number of response options were provided to preclude participants from making neutral responses.

Design

Fourteen versions of the experiment were created. The 560 experimental words were randomly divided into two stimulus sets (A, B) each containing 280 words. Each stimulus set was randomly divided into 14 lists, each containing 20 words. Each list was then paired with one of the seven ratings questions (excluding the familiarity rating) with the constraint that each question be selected twice (i.e., each of the seven questions was paired with two lists). Seven different list-question pairings (i.e., blocks) were created such that each list cycled through each of the seven questions. That is, every seven participants received the same pairing, and every second participant received the same stimulus set. The order of presentation of each block, however, was randomized across participants.

Procedure

Upon signing up for the experiment online, each participant was e-mailed a unique password with which they could login to the experiment website at their convenience. The e-mail emphasized the importance of setting aside one hour to complete the experiment undisturbed, and reiterated the inclusion criteria. Upon logging into the secure website, participants were asked to provide informed consent by typing their names and the date. If they agreed to participate they were redirected to a form asking several demographic questions, followed by a page explaining the upcoming training session.

Participants then performed a training session that was designed to familiarize them with quickly and accurately pressing the number keys from 1 to 8 on a computer keyboard. They were instructed to place their index fingers and pinky fingers on the 4, 5, and 1,8 keys respectively. They then completed 66 practice trials where a prompt stated “What is the number shown?” and a number between 1 and 8 appeared above the prompt. The first 16 trials consisted of 1 to 8 and 8 to 1 presented sequentially. The remaining 50 trials were randomly selected. At the completion of this training block participants were informed of their accuracy rate. If they correctly responded to 65% or more trials, they were given the option to either continue to the experiment or repeat the practice session. If they correctly responded to less than 65% of trials they repeated the practice session as many times as needed to pass this criterion (no participant needed more than three attempts).

Following the practice session participants were instructed that they would be asked to make several judgments about “words that refer to objects such as tools, animals, vehicles, fruits, etc.” They were informed that for each word they would first rate their familiarity with the object the word refers to on a scale from “Extremely familiar” to “Not at all familiar.”

Second, they were asked to “please rate the object on a particular characteristic (e.g., how it looks, feels, smells, etc.).” Participants then viewed the second part of instructions, which contained an example of each rating question, the scale they would use to make their rating, and a brief description of what a typical judgment at either end of the scale might entail (see Appendix A). In the likelihood of pain example we included additional examples at the middle of the scale because pilot testing suggested participants might require further explanation. The wording of the taste pleasantness question differed slightly from the other likelihood questions (i.e., “The taste of this object is most likely?”) because we wanted participants to focus on the perception of taste, rather than pleasantness—which could involve other modalities, or perhaps a more abstract judgment. Finally, participants were encouraged to respond as accurately and quickly as possible, and were informed that they could not change an answer once registered, that some trials would be more difficult than others, and that there were no “correct” answers.

Each experimental block was preceded by an example trial identical to what would appear in that block; example trial stimuli did not re-appear in the experimental trials. Each trial consisted of the target noun presented in 18 pt black Arial font, below which appeared the ratings question and scale presented in 14 pt Arial font. Note that these are relative size measures. The actual size of presented stimuli for each participant was determined by the screen size and resolution of their monitor. These stimuli remained on the screen until a response was entered. Participants responded by typing a single numeric character into a 2-character wide text box directly below the rating scale, after which the response and the response latency were automatically entered into the database (i.e., participants did not have to press *Enter*). Response latency was defined as the elapsed time (ms) between the simultaneous presentation of the target word and rating scale, and the registration of a key input. The subsequent trial was presented after a 500 ms delay. The experiment took participants between 40 and 60 minutes to complete.

Analysis

We discarded all data from 36 participants with RTs less than 250 ms on at least 15% of trials. We discarded all data from an additional 8 participants that typed the same response in succession for 20 or more trials. Next, we removed single trials with response latencies less than 250 ms or greater than 6000 ms (5% of remaining trials). Finally, responses (1 to 8) and response times (ms) were averaged across participants (mean number of participant ratings for each item is 23) for each question type and each noun. Each noun was then associated with a single mean rating and response time for each question. We unintentionally collected data for “onion” and “onions,” and retained only “onion” in the final dataset consisting of 559 items.

Results & Discussion

The full set of stimuli, attribute ratings, associated response times, and principal component scores (see below), are available as supplementary material (see descriptions in Appendix B). Examples of items at both extremes of each rating are provided in Table 1. No item appears more than once in this table, highlighting the diversity of knowledge types. The distributions of ratings varied considerably (Figure 1). For instance, whereas graspability and visual motion were approximately bimodal, the remaining ratings were positively skewed.

The mean response times differed to some extent between ratings (Table 2); the range between the fastest and slowest attribute was 103 ms. Familiarity ratings were considerably slower than the others (most likely because this rating accompanied the first exposure to each word), and should not be taken as an accurate reflection of the time course of

familiarity judgments. Given the web-based format, the mean response times associated with each attribute should be taken as crude approximations of time course information. That said, these by-item response times may be useful in designing experiments. An experimenter could match a set of stimuli not only on a given attribute rating, but also the response times associated with the rating, which may be able to account for some amount of previously unmeasured variance in task performance.

Despite our caution in interpreting the bases of these response times, it is worth noting that taste judgments were substantially faster than any others. A significant difference existed ($t(1116) = -4.36, p < .001$) between the by-item taste judgment times ($M = 1121$ ms) and the second fastest sound judgment times ($M = 1186$ ms). Although we can only speculate about the mechanisms underlying this advantage, it is intriguing to note that perceiving pictures of high versus low calorie foods (which presumably reflects taste pleasantness to some extent) may generate increased activation of neural reward networks (Killgore, Young, Femia, Bogorodzki, Rogowska, & Yurgelun-Todd, 2003), and could modulate image-locked electrophysiological brain potentials as early as 165 ms following picture onset (Toepela, Knebela, Hudry, le Coutre, & Murray, 2009). Whether a neural reward network sensitive to taste pleasantness can be engaged using words versus images, and if so, how quickly, remains to be determined.

Assessing latent structure

Several pairs of attribute ratings were significantly correlated (Table 3), suggesting the presence of latent structure. We assessed the shared variance across the 7 attribute ratings (excluding familiarity) with principal component analysis (PCA), a useful statistical technique for finding latent patterns in high-dimensional data. The PCA was used to aid in interpreting the shared knowledge underlying each attribute. In addition, the resulting component scores—which reflect weighted mixtures of particular sets of attributes—were compared with several of the ratings described in the Introduction. These analyses shed some new light on the kinds of knowledge that may underlie the different rating variables available in the literature.

Upon conducting a principal component analysis with varimax rotation we inspected the resulting scree plot, which revealed a marked decrease in the proportion of original variance explained after the 2nd eigenvalue, and thus suggested that a two-factor solution provides a parsimonious decomposition of the original ratings. The first and second factors accounted for 34% and 26% of the variance in the original variables, respectively. The varimax-rotated solution is visually depicted in Figure 2, in which sound intensity, visual motion, and likelihood of pain cluster together on the first component, and color, taste, and smell cluster on the second component. The component loadings are provided in Table 4.

The first component reflects both living and nonliving objects (e.g., missile, lion, train, bull) that capture our attention via multiple sensory modalities. Graspability has a substantial negative loading on this first component, consistent with the observation that loud, potentially harmful objects likely to be in motion are relatively unlikely to be graspable in one hand. The second principal component loads on vividly colored objects that are likely to emit a strong smell and taste good. It transparently reflects foods—both biological and otherwise (e.g., orange, cake, lollipop). These two components may reflect information about two requisites for survival and thus successful gene transmission: namely, avoiding death and locating nourishment. The primacy of the first component could reflect the possibility that visual, auditory, and nociceptive sensory organs are adaptations conferred by evolution. Vision may have evolved to exploit the kind of electromagnetic energy that does not pass through objects, thus providing the organism with information about the location of potentially harmful moving objects. Under this interpretation, the visual system did not

evolve to provide the organism with knowledge *per se*, but to provide useful knowledge (Marr, 1982). Similarly, the auditory system may have evolved in part to detect sounds that are useful for identifying the current location of objects in the environment, including predators (Stebbins & Sommers, 1992). Finally, as Dawkins (2009) points out, nociception may have been favored by natural selection over a less unpleasant warning system for noxious stimuli, as long as the ability to experience pain increased the likelihood of survival.

Comparisons with other ratings studies

Additional support for the above speculations appears in Wurm (2007), who reported mean ratings for danger and usefulness on a set of words including 104 nouns (i.e., participants rated the extent that a word denotes an entity that is “Not at all useful/dangerous for human survival” versus “Extremely useful/dangerous for human survival”) on an 8-point scale. Wurm used these ratings to predict lexical decision latencies and found an interaction between the factors (see previous similar results cited within) that may reflect competing pressures to both avoid dangerous objects and approach valuable resources (e.g., food). Although only 29 nouns were shared between his and our datasets, the correlation between our 1st principal component scores and his danger ratings was significant ($r = .67, p < .01$), as was the relationship between our 2nd component scores and his usefulness ratings ($r = .53, p < .01$). Examination of correlations with specific ratings provides an even more transparent explanation. The strongest association with his danger ratings and usefulness ratings, respectively, was with our likelihood of pain ratings ($r = .89, p < .01$) and taste pleasantness ratings ($r = .63, p < .01$).

Next, we determined which of the current ratings are most strongly associated with the established concreteness and imageability variables (Coltheart, 1981). Among 358 shared items, only taste pleasantness ($r = .30, p < .001$) and smell intensity ($r = .31, p < .001$) had notable correlations with concreteness. Among 361 shared items, only color vividness ($r = .33, p < .001$) and familiarity ($r = .33, p < .001$) had notable correlations with imageability.

We compared the Medler et al. (2005) perceptual attribute ratings with the current ratings, in which 355 items overlapped. The highest agreements among the three directly comparable ratings were sound ($r = .94, p < .001$) and motion ($r = .92, p < .001$), suggesting these ratings capture a common latent variable, followed by color ($r = .72, p < .001$). Next, our graspability ratings were designed to capture the degree that an object affords grasping by a single hand, which is not the same as manipulation. Medler et al. defined manipulation as follows: “a physical action done to an object by a person. Note that a manipulation is something that is DONE TO an object, NOT something that the object does by itself.” As expected given this difference, their manipulation and our graspability ratings were only moderately correlated ($r = .38, p < .001$), suggesting a substantial difference in the type of knowledge brought to bear on each decision. Finally, our likelihood of pain and their emotional valence had a substantial negative correlation ($r = -.50, p < .001$), which would be expected.

We compared our graspability ratings with Bennett et al.’s (2011) body-object-interaction ratings which were only moderately correlated ($r = .62, p < .001$) among 266 shared items, suggesting substantial differences in the underlying knowledge bases—perhaps because BOI reflects any part of the body, not just the hand. We then compared each of our attribute ratings to Juhasz et al.’s (2011, 2012) SER variable, which is thought to reflect all sensory modalities. Among 337 shared items, we found five significant correlations, though no association was particularly strong: from largest to smallest, color intensity ($r = .25, p < .001$), smell intensity ($r = .24, p < .001$), taste pleasantness ($r = .21, p < .001$), sound intensity ($r = .14, p < .001$), and visual motion ($r = .11, p < .05$). Notice that the three largest associations are driven by the same three attributes that contributed to our 2nd principal

component. Indeed the strongest relationship here is between the 2nd component scores and SER ($r = .30, p < .001$), which suggests Juhasz et al.'s SER variable may be weighted more heavily by those knowledge types most salient in the conceptual representations of edible entities (c.f., Cree & McRae, 2003). For instance, the five words with the highest SER ratings (among all 5857 monosyllabic and disyllabic words) are “garlic”, “walnut”, “water”, “pudding”, and “spinach.”

Finally, we compared the present graspability ratings with Salmon et al.'s (2010; p. 85) graspability ratings (i.e., “please rate the manipulability of the object according to how easy it is to grasp and use the object with one hand”), which were made on photographs not words, originated from a subject pool in Atlantic Canada, and were conducted in a laboratory. Despite these differences, the ratings were highly correlated ($r = .97, p < .001$) among 161 shared items, which bolsters the validity of our web-based data collection.

Putting the ratings to use: Semantic richness effects

Concepts associated with greater amounts of semantic information are recognized faster and more accurately than relatively impoverished concepts (Pexman et al., 2008). The behavioral semantic richness effect has been shown with several measures, including the number of listed features for a given concept, which can influence decision latencies in lexical and semantic decision tasks (Pexman, Holyk, & Monfils, 2003; Pexman, Lupker, & Hino, 2002). More recently, Grondin et al. (2009) and Amsel (2011) demonstrated that specific types of number-of-feature measures (e.g., shared features, visual motion features, function features) account for unique portions of variance in behavioral decision latencies and electrophysiological activity, respectively. Certain types of object knowledge such as gustatory, olfactory, and auditory information, however, are not well represented by current feature norms—many concepts have no features of this type listed. The present attribute ratings may be better suited for capturing certain kinds of information because they are distributed among integers equal to or greater than one, and approximate a continuous variable after averaging. In addition, the nature of the information contained in the ratings likely differs to some extent from the feature counts. Number of visual color features and color vividness ratings, for example, may be respectively tapping into the salience of color information for a concept and the vividness of the color itself. For instance, “coconut” along with two other concepts, has the highest number of visual color features (4) in the entire McRae et al. norms, but its mean color vividness rating in the present norms is well below average (3.3). For these reasons we directly compared the predictive performance of the current ratings with the measures employed by Grondin and colleagues. If each kind of content (i.e., feature norms and attribute ratings) captures unique aspects of word meaning, we should find that variables from both datasets enter into the upcoming regression equations.

We report the results of two regression analyses designed to examine the ability of the current ratings to account for variance in lexical and semantic decision latencies from Grondin et al. (2009). We are especially interested in a direct comparison of the number-of-features measures to the attribute ratings. Two models were fitted to decision latencies on 245 items from lexical and concreteness decision tasks. Word frequency (natural log of HAL frequency), word length, and object familiarity from McRae et al. (2005) were forced into the models regardless of statistical significance. Next, variables from two sources competed for model inclusion. The first were number of shared (i.e., co-occur in three or more of 541 concepts in McRae et al., 2005) visual motion, color, visual form and surface, taste, smell, sound, tactile, and encyclopedic features. The second were the mean ratings for each of the 7 attributes in the current norms. We employed all-subsets regression followed by cross validation to select a best model (see McLeod & Xu, 2011, for implementation details). The best fitting model (i.e., largest log-likelihood) for every model size from one to

k variables was initially selected, where k is the total number of candidate variables. The single best model from these candidate models was then identified using delete- d cross-validation¹, which increases the likelihood that the selected model would account for decision latencies collected on a different random sample of concrete nouns. Results from each model fit are shown in Table 5.

Participants were faster to signal an object concept as concrete when the concept was associated with a more intense smell, and had more visual form and surface, encyclopedic, and tactile features. Participants were faster to signal an object concept as a valid English word when the concept was associated with a higher likelihood of visual motion, increased taste pleasantness, and had more encyclopedic and tactile features. The results of these re-analyses of the Grondin et al. data suggest both feature norms and attribute ratings capture important and non-redundant information about the content of object concepts. The significant effects of smell intensity and taste pleasantness in the concreteness and lexical tasks, respectively, are particularly interesting in that these types of knowledge have often been overlooked in studies of lexical and semantic processing. These results, including our analysis of Juhasz et al. (2011, 2012), bolster the suggestion that a richer array of perceptually-based semantic knowledge is made available during language tasks than previously thought. The significant benefits of taste pleasantness and visual motion on lexical decision performance are especially interesting because successful discrimination of a word from a nonword need not rely on any aspect of word meaning, let alone specific perceptual inputs like taste and motion. Future research will need to examine the extent to which different kinds of knowledge are brought to bear on lexical and semantic decisions, as well as the stability of such effects. Our ratings could be used to design controlled experiments aimed at testing specific claims about knowledge use during language comprehension. For example, a researcher could select a set of words rated low and high on color vividness or sound intensity but matched on relevant psycholinguistic variables, and determine whether and how much these variables influence performance on various language tasks.

The fact that different attributes entered each regression model and certain attributes entered neither model may reflect some degree of task-specific conceptual flexibility in the brain. The kinds of object knowledge recruited during lexical decisions could differ substantially from the knowledge recruited during concreteness decisions. Additional tasks such as pleasantness decisions, or even natural reading in different contexts, could involve the recruitment of different subsets of knowledge—perhaps including those knowledge types that did not influence lexical and concreteness latencies. Some support for this notion of conceptual (in)flexibility has been provided by Grossman and colleagues (Grossman et al., 2006; Peelle, Troiani, & Grossman, 2009), who found that for the same set of nouns, typicality judgments versus pleasantness judgments, and similarity-based strategies versus rule-based strategies, resulted in markedly different patterns of neural activation. Similarly, Hoenig and colleagues (2008) found that neural activation in vision and motion-related regions was sensitive to whether participants verified visual or action-related properties of words denoting object concepts.

Lexical and concreteness decision tasks are just two of many tools to study linguistic and conceptual processing. Our attribute ratings also could be used in a larger variety of tasks to determine the degree of task-specific flexibility in the brain. For example, a cognitive

¹For each candidate model, a random sample of 78% of the by-item decision latencies was held-out (i.e., the “validation set”) while regression parameters were estimated from the remaining 22% of by-item decision latencies (i.e., the “training set”). This ratio was determined by equation 4.5 in Shao (1997). Mean-squared prediction error (MSE) was computed by subtracting the predicted y values of the training model, from the observed y values in the validation set, and taking the mean of the squares of the difference. This procedure was repeated 1000 times, resulting in a grand-average cross-validation error score (i.e., average MSE).

neuroscientist could select words rated as very low or high on graspability, and test whether the intensity and time course of neural activity underlying perception of these words differs as a function of whether or not the preceding context draws the comprehenders attention to graspability.

Conclusion

We reported the results of a large-scale, web-based object attribute rating study including a number of informative statistical analyses, and offer the ratings for future use. We discussed their relation to existing attribute ratings, and demonstrated their use as significant predictors of performance in word recognition experiments. The current set of attribute ratings include relatively unexplored dimensions of object knowledge such as pain perception and taste pleasantness, which may be useful for additional research into the interface between perception and semantics. Finally, at least 90% of the nouns in large-scale feature norms (McRae et al., 2005; Vinson & Vigliocco, 2008) were included in our ratings, resulting in a richer collective database for use in future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NICHD grant 22614 to M.K., and Center for Research in Language postdoctoral NIDCD fellowship T32DC000041 to B.D.A

Appendix

Appendix A:

How familiar are you with this object?

Not at all familiar								Extremely familiar
1	2	3	4	5	6	7	8	

How vivid (intense) is the color of this object?

Not at all vivid							Extremely vivid
1	2	3	4	5	6	7	8

For this question we are interested in the vividness of color. For example, I would rate 'bone' which is a dull color as 1 or 2 and a 'sunflower' which is brightly colored as a 7 or 8.

When one sees this object, how likely is it to be in motion?

Extremely unlikely							Extremely likely
1	2	3	4	5	6	7	8

For this question we are interested in the perception of motion. When someone sees an object, how likely is that object to be in motion? For example, though a 'chimney' could eventually crumble, it is not typically in motion and I would rate it 1 or 2. On the other hand, a 'hummingbird' could be sitting still on a branch, but it is typically in motion and I would rate it 7 or 8.

How likely is someone to grasp this object with one hand?

Extremely unlikely							Extremely likely
1	2	3	4	5	6	7	8

For this question we are interested in how likely one would grasp an object with one hand. For example, a 'comet' is too large to grasp and I would rate it 1 or 2, whereas a 'donut' is made to be picked up and put in one's mouth and I would rate it 7 or 8.

How likely is this object to cause pain?

Extremely unlikely							Extremely likely
--------------------	--	--	--	--	--	--	------------------

1 2 3 4 5 6 7 8

For this question we are interested in the likelihood that an object would cause pain. For example, although a 'feather' could be painful if I poked you with it, it is soft and light and does not typically cause pain so I would rate it 1 or 2. By contrast, an 'arrow' though it could be used for target practice, is made to inflict wounds and thus I would rate it high, maybe 7 or 8. And I would rate a 'finger' between 3 and 4 because it hurts to get poked but fingers do lots of other non-painful things.

How loud (intense) is the sound produced by this object?

Not at all loud 1 2 3 4 5 6 7 8 Extremely loud

For this question we are interested in the intensity of sound. For example, a 'toothpick' typically makes no noise and I would rate it 1 or 2, whereas an 'avalanche' makes lots of noise and I would rate it 7 or 8.

The taste of this object is most likely?

Extremely unpleasant 1 2 3 4 5 6 7 8 Extremely pleasant

For this question we are interested in the taste of an object. For example, although I've never tasted a 'sewer', I expect it would taste extremely unpleasant and I would rate it 1 or 2, whereas I love the taste of 'chocolate' and would rate it 7 or 8.

How intense is the smell produced by this object?

Not at all intense 1 2 3 4 5 6 7 8 Extremely intense

For this question we are interested in the intensity of smell. For example, the average 'mosquito' doesn't have much of a smell so I would rate it a 1 or 2, whereas 'manure' has a strong, pungent, odor which I would rate as a 7 or 8.

Appendix B: Explanations of variables in the object attributes file

Each row of the spreadsheet corresponds to one of the 559 rated items, named in the "Concept" column. Columns 2 to 9 contain the mean ratings for each attribute. Columns 10 to 17 contain the mean response times for each attribute. Columns 18 to 19 contain the principal component scores for the first and second extracted components (see text for more detail).

References

- Amsel BD. Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*. 2011; 49(5):970–983. [PubMed: 21219919]
- Andrews M, Vigliocco G, Vinson D. Integrating experiential and distributional data to learn semantic representations. *Psychological Review*. 2009; 116(3):463–498. [PubMed: 19618982]
- Bennett SDR, Burnett AN, Siakaluk PD, Pexman PM. Imageability and body-object interaction ratings for 599 multisyllabic nouns. *Behavior Research Methods*. 2011; 43(4):1100–9. [PubMed: 21681627]
- Binder JR, Desai RH, Graves WW, Conant LL. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*. 2009; 19(12):2767–2796. [PubMed: 19329570]
- Campanella F, D'Agostini S, Skrap M, Shallice T. Naming manipulable objects: Anatomy of a category specific effect in left temporal tumours. *Neuropsychologia*. 2010; 48(6):1583–1597. [PubMed: 20144630]
- Campanella F, Shallice T. Manipulability and object recognition: Is manipulability a semantic feature? *Experimental Brain Research*. 2011; 208(3):369–383.
- Cree GS, McRae K. Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology-General*. 2003; 132(2):163–201. [PubMed: 12825636]

- Davare M, Kraskov A, Rothwell JC, Lemon RN. Interactions between areas of the cortical grasping network. *Current Opinion in Neurobiology*. 2011; 21(4):565–570. [PubMed: 21696944]
- Dawkins, R. *The greatest show on earth: The evidence for evolution*. Free Press; New York: 2009.
- de Araujo IET, Rolls ET, Kringelbach ML, McGlone F, Phillips N. Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. *European Journal of Neuroscience*. 2003; 18(7):2059–2068. [PubMed: 14622239]
- Gonzalez J, Barros-Loscertales A, Pulvermuller F, Meseguer V, Sanjuan A, Belloch V, et al. Reading cinnamon activates olfactory brain regions. *NeuroImage*. 2006; 32(2):906–912. [PubMed: 16651007]
- Goodale MA, Meenan JP, Bulthoff HH, Nicolle DA, Murphy KJ, Racicot CI. Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology*. 1994; 4(7):604–610. [PubMed: 7953534]
- Grondin R, Lupker SJ, McRae K. Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language*. 2009; 60(1):1–19. [PubMed: 20046224]
- Grossman M, Koenig P, Kounios J, McMillan C, Work M, Moore P. Category-specific effects in semantic memory: Category-task interactions suggested by fMRI. *NeuroImage*. 2006; 30(3):1003–1009. [PubMed: 16413792]
- Hoenig K, Sim E, Bochev V, Herrnberger B, Kiefer M. Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience*. 2008; 20(10):1799–1814. [PubMed: 18370598]
- Johns, BT.; Jones, MN. Construction in semantic memory: Generating perceptual representations with global lexical similarity. In: Carlson, L.; Hölscher, C.; Shipley, T., editors. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*; Austin, TX. Cognitive Science Society; 2011. p. 767-772.
- Juhasz, BJ.; Yap, MJ. Sensory experience ratings (SERs) for over 5,000 mono- and disyllabic words. 2012. Unpublished manuscript
- Juhasz BJ, Yap MJ, Dicke J, Taylor SC, Gullick MM. Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology*. 2011; 64(9):1683–1691.
- Kan IP, Barsalou LW, Solomon KO, Minor JK, Thompson-Schill SL. Role of mental imagery in a property verification task: FMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*. 2003; 20(3-6):525–540. [PubMed: 20957583]
- Kellenbach ML, Brett M, Patterson K. Large, colorful, or noisy? attribute- and modality-specific activations during retrieval of perceptual attribute knowledge. *Cognitive Affective and Behavioral Neuroscience*. 2001; 1(3):207–221.
- Killgore WDS, Young AD, Femia LA, Bogorodzki P, Rogowska J, Yurgelun-Todd DA. Cortical and limbic activation during viewing of high-versus low-calorie foods. *NeuroImage*. 2003; 19(4):1381–1394. [PubMed: 12948696]
- Lynott D, Connell L. Modality exclusivity norms for 423 object properties. *Behavior Research Methods*. 2009; 41(2):558–564. [PubMed: 19363198]
- Magnie MN, Besson M, Poncet M, Dolisi C. The snodgrass and vanderwart set revisited: Norms for object manipulability and for pictorial ambiguity of objects, chimeric objects, and nonobjects. *Journal of Clinical and Experimental Neuropsychology*. 2003; 25(4):521–560. [PubMed: 12911106]
- Marr, D. *Vision: A computational investigation into the human representation and processing of visual information*. WH Freeman and Company; San Francisco: 1982.
- Martin A. The representation of object concepts in the brain. *Annual Review of Psychology*. 2007; 58:25–45.
- Martin A, Haxby JV, Lalonde FM, Wiggs CL, Ungerleider LG. Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*. 1995; 270(5233):102–105. [PubMed: 7569934]
- McLeod, AI.; Xu, CJ. [Retrieved December 2, 2011] Bestglm: Best subset GLM. R package version 0.33from <http://CRAN.R-project.org/package=bestglm>

- McRae K, Cree GS, Seidenberg MS, McNorgan C. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*. 2005; 37(4):547–559. [PubMed: 16629288]
- McRae K, deSa VR, Seidenberg MS. On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology-General*. 1997; 126(2):99–130. [PubMed: 9163932]
- Medler DA, Arnoldussen A, Binder JR, Seidenberg MS. The wisconsin perceptual attribute ratings database. 2005 from <http://www.neuro.mcw.edu/ratings/>.
- Millan MJ. The induction of pain: An integrative review. *Progress in Neurobiology*. 1999; 57(1):1–164. [PubMed: 9987804]
- Moscato del Prado Martin F, Hauk O, Pulvermüller F. Category specificity in the processing of color-related and form-related words: An ERP study. *NeuroImage*. 2006; 29(1):29–37. [PubMed: 16246594]
- Nagasako EM, Oaklander AL, Dworkin RH. Congenital insensitivity to pain: An update. *Pain*. 2003; 101(3):213–219. [PubMed: 12583863]
- Oliver RT, Geiger EJ, Lewandowski BC, Thompson-Schill SL. Remembrance of things touched: How sensorimotor experience affects the neural instantiation of object form. *Neuropsychologia*. 2009; 47(1):239–247. [PubMed: 18760292]
- Oliver RT, Thompson-Schill SL. Dorsal stream activation during retrieval of object size and shape. *Cognitive Affective & Behavioral Neuroscience*. 2003; 3(4):309–322.
- Patterson K, Nestor PJ, Rogers TT. Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*. 2007; 8(12):976–987.
- Pexman PM, Holyk GG, Monfils MH. Number-of-features effects and semantic processing. *Memory & Cognition*. 2003; 31(6):842–855.
- Pexman PM, Lupker SJ, Hino Y. The impact of feedback semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*. 2002; 9(3):542–549. [PubMed: 12412895]
- Pexman PM, Hargreaves IS, Siakaluk PD, Bodner GE, Pope J. There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*. 2008; 15(1):161–167. [PubMed: 18605497]
- Preston CC, Colman AM. Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*. 2000; 104(1):1–15. [PubMed: 10769936]
- Salmon JP, McMullen PA, Filliter JH. Norms for two types of manipulability (graspability and functional usage), familiarity, and age of acquisition for 320 photographs of objects. *Behavior Research Methods*. 2010; 42(1):82–95. [PubMed: 20160288]
- Shao J. An asymptotic theory for linear model selection. *Statistica Sinica*. 1997; 7:221–262.
- Siakaluk PD, Pexman PM, Aguilera L, Owen WJ, Sears CR. Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect. *Cognition*. 2008a; 106(1):433–443. [PubMed: 17258186]
- Siakaluk PD, Pexman PM, Sears CR, Wilson K, Locheed K, Owen WJ. The benefits of sensorimotor knowledge: Body-object interaction facilitates semantic processing. *Cognitive Science*. 2008b; 32(3):591–605. [PubMed: 21635348]
- Simmons WK, Ramjee V, Beauchamp MS, McRae K, Martin A, Barsalou LW. A common neural substrate for perceiving and knowing about color. *Neuropsychologia*. 2007; 45(12):2802–2810. [PubMed: 17575989]
- Solomon KO, Barsalou LW. Perceptual simulation in property verification. *Memory & Cognition*. 2004; 32(2):244–259.
- Stebbins, WC.; Sommers, MS. Evolution, perception, and the comparative method. In: Webster, DB.; Fay, RR.; Popper, AN., editors. *The evolutionary biology of hearing*. Springer-Verlag; New York: 1992. p. 211–227.
- Tillotson SM, Siakaluk PD, Pexman PM. Body-object interaction ratings for 1,618 monosyllabic nouns. *Behavior Research Methods*. 2008; 40(4):1075–1078. [PubMed: 19001398]
- Toepel U, Knebel J, Hudry J, le Coutre J, Murray MM. The brain tracks the energetic value in food images. *NeuroImage*. 2009; 44(3):967–974. [PubMed: 19013251]

- van Dantzig S, Cowell RA, Zeelenberg R, Pecher D. A sharp image or a sharp knife: Norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods*. 2011; 43(1): 145–154. [PubMed: 21287109]
- Vinson DP, Vigliocco G. Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*. 2008; 40(1):183–190. [PubMed: 18411541]
- Weng LJ. Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*. 2004; 64(6):956–972.
- Wu L, Barsalou LW. Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*. 2009; 132(2):173–189. [PubMed: 19298949]
- Wurm LH. Danger and usefulness: An alternative framework for understanding rapid evaluation effects in perception? *Psychonomic Bulletin & Review*. 2007; 14(6):1218–1225. [PubMed: 18229500]

Figure 1. Distributions of attribute ratings

By-items histograms for each attribute are depicted to estimate the continuous variable capturing each attribute rating. The x-axis depicts the full range of the rating scale (1 – 8) and the y-axis depicts the frequency of items falling into each discrete bin. Number of bins varies according to the range of the ratings for each attribute.

Figure 2. Principal component analysis: Varimax-rotated two-component solution

Words denoting the 7 original rating variables are placed at their coordinates on each component, and referenced by arrows originating at the zero-point of both components. The 1st and 2nd components accounted for 34% and 26% of the original variance, respectively. Grey data points signify coordinates of all 559 words; “A” denotes an artifact concept and “B” denotes a biological concept. Four individual words are shown in circles referenced by arrows to the word’s identity.

Table 1

Extreme scores in each dimension

Attribute	Lowest items	Highest items
Familiarity	Budgie, Starling, Oriole	Bed, Telephone, Bread
Likelihood of pain	Robe, Mittens, Pajamas	Whip, Grenade, Shotgun
Taste pleasantness	Shoes, Truck, Ambulance	Cake, Strawberry, Corn
Smell intensity	Thermometer, Clock, Baton	Cigarette, Garlic, Sardine
Graspability	Moon, Rainbow, Van	Penny, Olive, Cup
Sound intensity	Grapefruit, Caterpillar, Scarf	Jet, Cannon, Bomb
Visual motion	Wall, Church, Dresser	Dolphin, Butterfly, Ant
Color vividness	Envelope, Tissue, Freezer	Highlighter, Peacock, Orange

\$watermark-text

\$watermark-text

\$watermark-text

Table 2

Descriptive statistics of by-item response latency (ms)

	Familiarity	Pain	Smell	Color	Taste	Sound	Grasp	Motion
Mean	1729	1198	1187	1197	1121	1186	1224	1204
S.D.	153	239	239	228	245	252	313	262

Table 3

Correlations among attributes

	Familiarity	Pain	Smell	Color	Taste	Sound	Grasp
Pain	-0.17						
Smell	--	--					
Color	0.17	-0.21	0.22				
Taste	0.13	-0.30	0.58	0.33			
Sound	-0.16	0.55	--	-0.15	-0.27		
Grasp	0.31	-0.23	-0.11	0.12	0.24	-0.42	
Motion	-0.19	0.40	0.11	--	-0.15	0.53	-0.39

Pearson's correlation coefficients are shown only if statistically significant (i.e., if the 95% confidence interval (CI) around the estimate does not contain zero (CIs were determined with 10,000 runs of the accelerated bias-corrected bootstrap).

Table 4

Standardized component loadings

Attribute	CL1	CL2
Pain	0.69	−0.27
Smell	0.19	0.83
Color	−0.11	0.60
Taste	−0.26	0.83
Sound	0.83	−0.12
Grasp	−0.67	0.02
Motion	0.79	0.10

CL: Component loading.

Table 5

Best predictors of Grondin et al (2008) experiments (N = 245).

Variable	Concreteness decision task		Lexical decision task	
	β	p	β	p
Smell intensity	-.20	<.001		
Visual motion			-.23	<.001
Taste pleasantness			-.17	.003
Visual form/surface	-.15	.002		
Encyclopedic	-.13	.007	-.15	.012
Tactile	-.12	.015	-.14	.015

Beta symbol corresponds to standardized regression coefficients; p = p-value associated with a specific term in the model. Variables are sorted by effect size (standardized regression coefficient). Bold font variables are from the current norms, regular font variables are from Grondin et al. Familiarity, word frequency, and word length were forced into each regression model, and the remaining variables competed for inclusion in all-subsets regression analysis. The best models had the lowest mean squared prediction error computed from 1000 runs of the delete-*d* cross validation procedure (see text for details).