

Perceptual and motor attribute ratings for 559 object concepts

Ben D. Amsel · Thomas P. Urbach · Marta Kutas

Published online: 23 June 2012
© Psychonomic Society, Inc. 2012

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 that people have. To extend the normative data currently available, 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, $M = 23$ raters per item). The mean ratings were subjected to principal-components 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.

Keywords Sensory · Perceptual · Conceptual · Object · Ratings · Norms · Semantic · Modality

The representation of object concepts in long-term memory and the recruitment of this knowledge during language

Electronic supplementary material The online version of this article (doi:10.3758/s13428-012-0215-z) contains supplementary material, which is available to authorized users.

B. D. Amsel (✉) · M. Kutas
Center for Research in Language,
University of California, San Diego,
San Diego, CA, USA
e-mail: bamsel@ucsd.edu

T. P. Urbach · M. Kutas
Cognitive Science, University of California, San Diego,
San Diego, CA, USA

M. Kutas
Department of Neuroscience, University of California,
San Diego, CA, USA

comprehension have long been central topics in cognitive science, and they 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 of how task-related context flexibly modulates activation of object knowledge (Grossman et al., 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, he or 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 or she 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 (1) incorporating a measure of response time for each attribute, (2) utilizing a larger than typical set of words, and (3) including not only standard perceptual attributes (e.g., color) but also less studied attributes (e.g., likelihood of pain or taste pleasantness). The inclusion of these attributes is important for

researchers interested in the full gamut of sensory modalities, and they could motivate additional study of modalities that have received relatively less attention. We conducted a principal-components 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 the attributes of a given concept (e.g., <is red>, <used for cooking>) and retaining only the attributes listed by at least two to three participants (e.g., McRae, Cree, Seidenberg, & McNorgan, 2005; Vinson & Vigliocco, 2008). These data sets 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 to show how feature correlations influence the organization of concepts in semantic memory (McRae, de Sa, & Seidenberg, 1997). Semantic features have also 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. 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 rated the degrees 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 computed 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 they 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?” The researchers found differences in event-related brain potential amplitudes beginning at 200 ms to words rated high on color versus form relatedness, which they took as suggesting rapid access (and differentiation) of semantic information during word recognition. Kellenbach et al. (2001) used objects that were either 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 et al. (2006) asked participants to rate words on the degrees to which they referred to objects with a strong smell, and found that 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 to which an object could be uniquely pantomimed. Campanella, D’Agostini, Skrap, and 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. These researchers subsequently showed an explicitly semantic influence of manipulability in word-to-picture matching tasks and argued that 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 studies above largely concerned 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, 2008; Siakaluk, Pexman, Sears, et al., 2008) are designed to index the extent that people interact with an object using any part of their bodies. Siakaluk and colleagues (Siakaluk, Pexman, Aguilera, et al., 2008; Siakaluk, Pexman, Sears, et al., 2008) 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 degrees to which a word evokes any kind of sensory experience. Importantly, although SER were correlated with imageability, they still predicted lexical decision latencies in a large data set when imageability was 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 that 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 do models trained on either distributional or experiential data alone. Importantly for the present purposes, the innovation of these models was 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 1,402 words ranging from very abstract (“advantage”) to very concrete (“airplane”). A total of 342 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 present study builds on this data set 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.

Present study

The main purpose of the present study was 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 assessed each of the dimensions above 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 on our aim to include a more comprehensive set of measures than previous norms have made available. Each of the five traditional Aristotelian sensory modalities (vision, touch, hearing, smell, and taste) is represented, in addition to the sensation of pain. We assessed two kinds of visual knowledge, color and motion, which are represented in different brain regions proximal to the corresponding sensory cortex (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Simmons et al., 2007). Ratings of taste and smell intensity were anticipated to be highly redundant, which motivated the collection of separate intensity and pleasantness judgments in the olfactory and gustatory domains, respectively (cf. de Araujo, Rolls, Kringelbach, McGlone, & Phillips, 2003). Tactile object information was assessed with graspability judgments, which reflect knowledge of physical object properties and learned sensorimotor programs. The motivation for this dimension derived from the importance of grasping behavior in our interaction with the environment and from the sustained research focus on its neural substrates (Chao & Martin, 2000; Davare, Kraskov, Rothwell, & Lemon, 2011; Goodale et al., 1994). Last but not least, we assessed 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 examined the distributions and response times associated with each. We conducted a principal-components analysis to uncover shared variance among the attributes, revealing two major sources of shared variance readily interpretable as related to survival. Finally, we utilized the ratings to account for portions of unique variance in published decision latencies in a concreteness and a lexical decision task, which revealed multiple attributes as successful predictors of decision latencies.

Method

Participants

A total of 420 undergraduate students (308 female, 109 male, and 3 who 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. The participants were native English speakers between 18 and 30 years of age ($M = 20.7$ years, $SD = 1.8$), had completed on average 15 years of education, and reported normal vision and no major neurological or general health problems. Of the participants, 377 were right-handed, 33 were left-handed, and the remainder declined to state.

Stimuli

Nouns Each of the 560 normed words corresponded to an English noun denoting an object concept. The nouns were chosen primarily from the two largest existing sets of 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 and vegetables, places, plants, musical instruments, tools, and 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 so as 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 and 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 log in

to the experiment website at their convenience. The e-mail emphasized the importance of setting aside 1 h to complete the experiment undisturbed, and it reiterated the inclusion criteria. Upon logging in to the secure website, the 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.

The 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 and 5 and the 1 and 8 keys on the keyboard, respectively. They then completed 66 practice trials in which a prompt stated “What is the number shown?” and a number from 1 to 8 appeared above the prompt. The first 16 trials consisted of 1 to 8 and 8 to 1 presented sequentially, and 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 of the 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 the 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, the 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 that the word referred 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 the instructions, which contained an example of each rating question, the scale that they would use to make their ratings, 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 that participants might require further explanation. The wording of the taste pleasantness question differed slightly from that of 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 on 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 it was 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; the example-trial stimuli did not reappear in the experimental trials. Each trial consisted of the target noun presented in 18-point black Arial font, below which appeared the rating question and scale, presented in 14-point Arial font. Note that these are relative size measures. The actual size of the presented stimuli for each participant was determined by the screen size and the resolution of their monitor. These stimuli remained on the screen until a response was entered. The participants responded by typing a single numeric character into a two-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 (in milliseconds) 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 min to complete.

Analysis

We discarded all of the data from 36 participants with response times less than 250 ms on at least 15 % of the trials. We discarded all of the data from an additional eight participants who had 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 6,000 ms (5 % of the remaining trials). Finally, responses (1 to 8) and response times (in milliseconds) were averaged across participants (the mean number of participant ratings for each item was 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 data set, consisting of 559 items.

Results and discussion

The full set of stimuli, attribute ratings, associated response times, and principal-component scores (see below) are available as supplementary materials (see the description in Appendix B). Examples of items at both extremes of each rating scale 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 (see Fig. 1). For instance, whereas graspability and visual motion were approximately bimodal, the remaining ratings were positively skewed.

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

The mean response times differed to some extent between ratings (Table 2); the range between the fastest and slowest attributes 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 on 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 ($M = 1,121$ ms) and the second fastest judgment times, for sound ($M = 1,186$ 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 et al., 2003) and could modulate image-locked electrophysiological brain potentials as early as 165 ms following picture onset (Toepel, Knebel, 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 seven attribute ratings (excluding familiarity) with principal-components analysis (PCA), a useful statistical technique for finding latent patterns

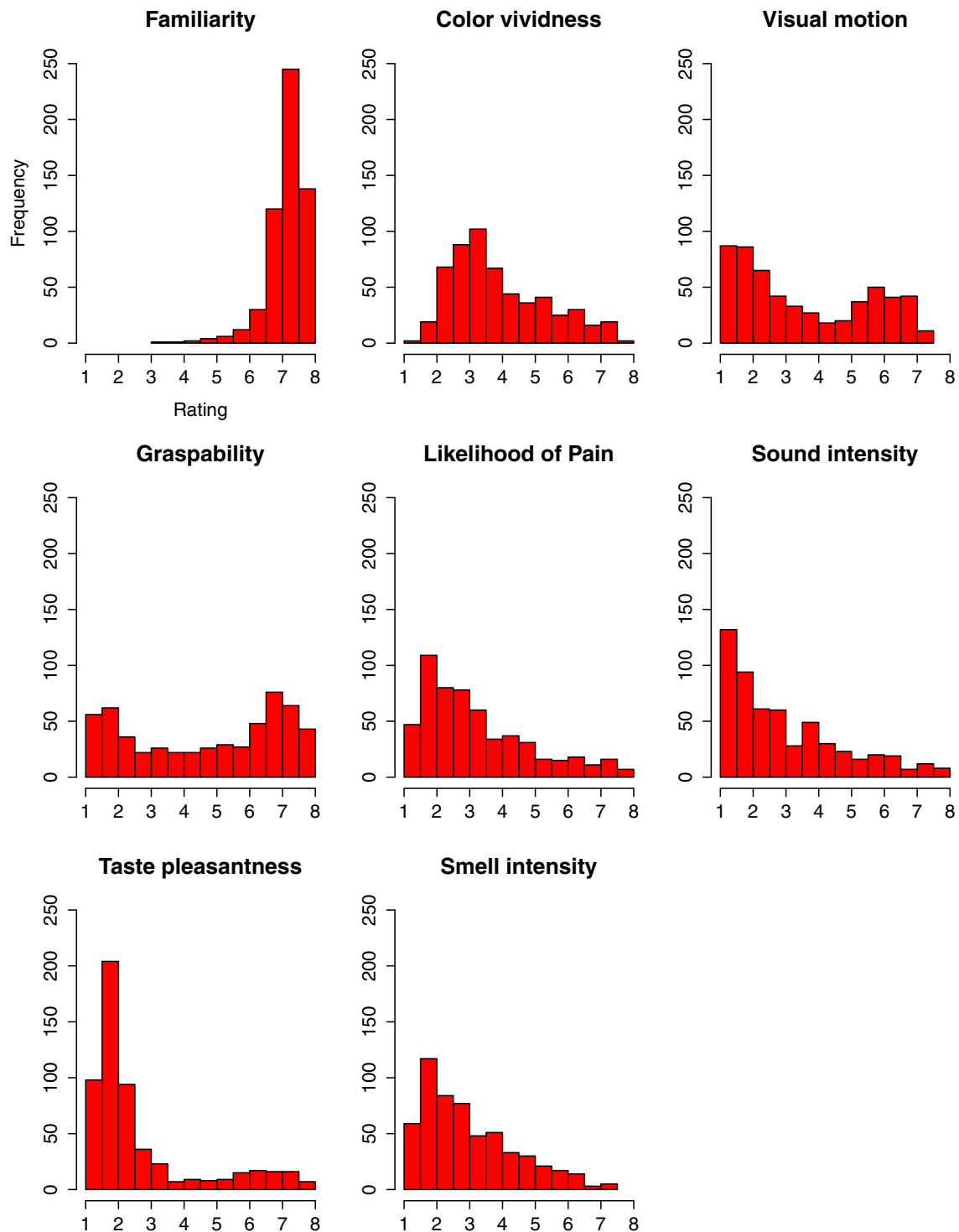


Fig. 1 Distributions of attribute ratings. By-items histograms for each attribute are depicted in order to estimate the continuous variable capturing each attribute rating. The *x*-axis depicts the full range of

the rating scale (1–8), and the *y*-axis depicts the frequency of items falling into each discrete bin. The number of bins varies according to the range of the ratings for each attribute

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.

Table 2 Descriptive statistics of by-items response latencies (in milliseconds)

	Familiarity	Pain	Smell	Color	Taste	Sound	Grasp	Motion
Mean	1,729	1,198	1,187	1,197	1,121	1,186	1,224	1,204
SD	153	239	239	228	245	252	313	262

Upon conducting a PCA with varimax rotation, we inspected the resulting scree plot, which revealed a marked decrease in the proportion of original variance explained after the second eigenvalue, thus suggesting 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 Fig. 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, and 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, and 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 locations of objects in the environment, including predators (Stebbins & Sommers, 1992). Finally, as Dawkins (2009) pointed 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* vs. *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 data sets, the correlation between our first-principal-component scores and his danger ratings was significant ($r = .67, p < .01$), as was the relationship between our second-component scores and his usefulness ratings ($r = .53, p < .01$). Examination of correlations with specific ratings provides an even more transparent explanation. The strongest associations with his danger ratings and usefulness ratings, respectively, were with our likelihood-of-pain ratings ($r = .89, p < .01$) and taste pleasantness ratings ($r = .63, p < .01$).

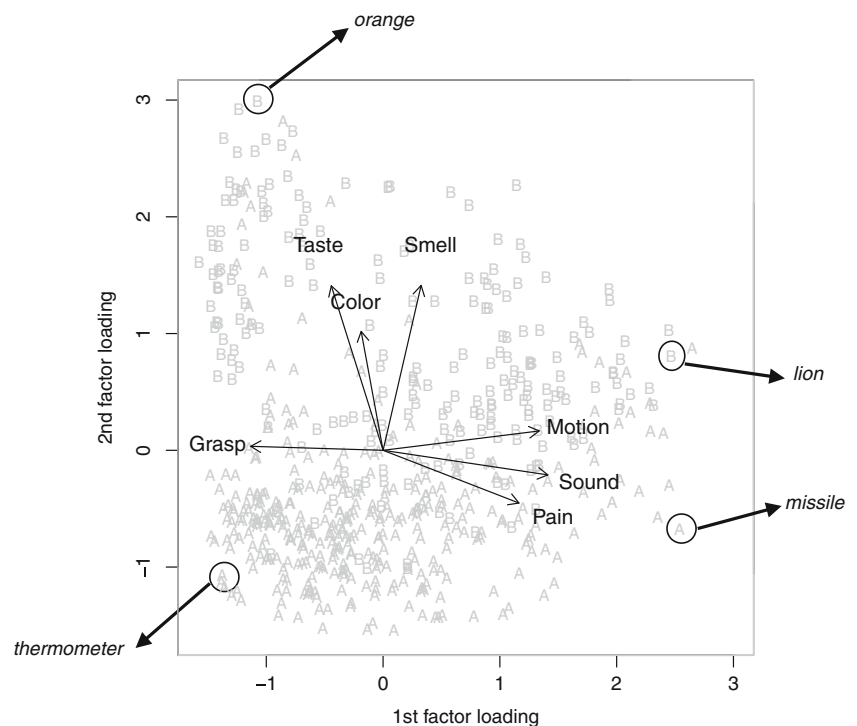
Next, we determined which of the present ratings are most strongly associated with the established concreteness and imageability variables (Coltheart, 1981). Among 358 shared

Table 3 Correlations among attributes

	Familiarity	Pain	Smell	Color	Taste	Sound	Grasp
Pain	-.17						
Smell	–	–					
Color	.17	-.21	.22				
Taste	.13	-.30	.58	.33			
Sound	-.16	.55	–	-.15	-.27		
Grasp	.31	-.23	-.11	.12	.24	-.42	
Motion	-.19	.40	.11	–	-.15	.53	-.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).

Fig. 2 Principal-components analysis: Varimax-rotated two-component solution. The words denoting the seven original rating variables are placed at their coordinates on each component and referenced by arrows originating at the zero-point of both components. The first and second components accounted for 34 % and 26 % of the original variance, respectively. The gray data points signify the 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



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 present 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 that 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. (<http://www.neuro.mcw.edu/ratings/instructions.html>) 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) BOI 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; Juhasz & Yap, 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, these were 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 second principal component. Indeed the strongest relationship here is between the second-component scores and SER ($r = .30, p < .001$), which suggests that Juhasz et al.’s SER variable may be weighted more heavily by those knowledge types most salient in the conceptual representations of edible

Table 4 Standardized component loadings

Attribute	CL1	CL2
Pain	.69	-.27
Smell	.19	.83
Color	-.11	.60
Taste	-.26	.83
Sound	.83	-.12
Grasp	-.67	.02
Motion	.79	.10

CL, component loading.

entities (cf. Cree & McRae, 2003). For instance, their five words with the highest SER ratings (among all 5,857 monosyllabic and disyllabic words) were “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 rather than words, originated from a subject pool in Atlantic Canada, and were conducted in a laboratory. Despite these procedural 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, and function features) account for unique portions of variance in, respectively, behavioral decision latencies and electrophysiological activity. 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 1 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. The number of visual color features and color vividness ratings, for example, may tap, respectively, into the salience of color information for a concept and the vividness of the color itself. For instance, “coconut,” along with two other concepts, had the highest number of visual color features (four) in the entire McRae et al. (2005) set of 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 present ratings with the measures employed by Grondin et al. 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 data sets enter into the upcoming regression equations.

We report the results of two regression analyses designed to examine the ability of the present ratings to account for variance in the 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. The word frequency (natural log of the HAL frequency), word length, and object familiarity data 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 the numbers of shared (i.e., co-occurring 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 seven attributes in the present norms. We employed an all-subsets regression followed by cross-validation to select the best model (see McLeod & Xu, 2011, for the implementation details). The best-fitting model (i.e., the 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-validation,¹ which increased the likelihood that the selected model would account for decision latencies collected on a different random sample of concrete nouns. The results from each model fit are shown in Table 5.

The 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, an increased taste pleasantness, and more encyclopedic and tactile features. The results of these reanalyses of the Grondin et al. (2009) data suggest that both feature norms and attribute ratings capture important and nonredundant 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

¹ For each candidate model, a random sample of 78 % of the by-items decision latencies were held out (i.e., the “validation set”) while regression parameters were estimated from the remaining 22 % of the by-items decision latencies (i.e., the “training set”). This ratio was determined by Eq. 4.5 in Shao (1997). The 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 differences. This procedure was repeated 1,000 times, resulting in a grand-average cross-validation error score (i.e., average *MSE*).

Table 5 Best predictors of Grondin et al. (2009) 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

The beta symbols correspond to standardized regression coefficients, and p is the p value associated with a specific term in the model. Variables are sorted by effect size (standardized regression coefficient). The variables in bold font are from the present norms, those in regular font from Grondin et al. (2009). Familiarity, word frequency, and word length were forced into each regression model, and the remaining variables competed for inclusion in an all-subsets regression analysis. The best models had the lowest mean-squared prediction error, computed from 1,000 runs of the delete- d cross-validation procedure (see the text for details).

that these types of knowledge have often been overlooked in studies of lexical and semantic processing. These results, including our analysis of results from Juhasz and colleagues (Juhasz & Yap, 2012; Juhasz et al., 2011), bolster the suggestion that a richer array of perceptually based semantic knowledge is made available during language tasks than has been 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 in both studies, typicality judgments versus pleasantness judgments and similarity-based strategies versus rule-based strategies resulted in markedly different patterns of neural activation. Similarly, Hoenig et al. (2008) found that neural activations in vision and in motion-related regions were sensitive to whether participants verified visual or action-related properties of the 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 neuroscientist could select words rated as very low or high on graspability and test whether the intensity and the time course of neural activity underlying perception of these words differ as a function of whether or not the preceding context draws the comprehender's attention to graspability.

Conclusion

We reported the results of a large-scale, Web-based object attribute rating study that included a number of informative statistical analyses, and we 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 present 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 from previous large-scale sets of 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.

Author note This research was supported by NICHD Grant 22614 to M.K. and by Center for Research in Language Postdoctoral NIDCD Fellowship T32DC000041 to B.D.A.

Appendix A: Rating instructions

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 2–9 contain the mean ratings for each attribute. Columns 10–17 contain the mean response times for each attribute. Columns 18 and 19 contain the principal-component scores for the first and second extracted components (see the text for more details).

References

- Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia, 49*, 970–983. doi:10.1016/j.neuropsychologia.2011.01.003
- Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review, 116*, 463–498.
- Bennett, S. D., Burnett, A. N., Siakaluk, P. D., & Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599

- multisyllabic nouns. *Behavior Research Methods*, 43, 1100–1109. doi:10.3758/s13428-011-0117-5
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19, 2767–2796. doi:10.1093/cercor/bhp055
- Campanella, F., D'Agostini, S., Skrap, M., & Shallice, T. (2010). Naming manipulable objects: Anatomy of a category specific effect in left temporal tumours. *Neuropsychologia*, 48, 1583–1597. doi:10.1016/j.neuropsychologia.2010.02.002
- Campanella, F., & Shallice, T. (2011). Manipulability and object recognition: Is manipulability a semantic feature? *Experimental Brain Research*, 208, 369–383.
- Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, 12, 478–484. doi:10.1006/nimg.2000.0635
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33A, 497–505. doi:10.1080/14640748108400805
- Cree, G. S., & McRae, K. (2003). 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*, 132, 163–201. doi:10.1037/0096-3445.132.2.163
- Davare, M., Kraskov, A., Rothwell, J. C., & Lemon, R. N. (2011). Interactions between areas of the cortical grasping network. *Current Opinion in Neurobiology*, 21, 565–570.
- Dawkins, R. (2009). *The greatest show on earth: The evidence for evolution*. New York, NY: Free Press.
- de Araujo, I. E. T., Rolls, E. T., Kringelbach, M. L., McGlone, F., & Phillips, N. (2003). Taste–olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. *European Journal of Neuroscience*, 18, 2059–2068.
- del Prado, M., Martin, F., Hauk, O., & Pulvermüller, F. (2006). Category specificity in the processing of color-related and form-related words: An ERP study. *NeuroImage*, 29, 29–37.
- González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Avila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, 32, 906–912. doi:10.1016/j.neuroimage.2006.03.037
- Goodale, M. A., Meenan, J. P., Bühlhoff, H. H., Nicolle, D. A., Murphy, K. J., & Racicot, C. I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology*, 4, 604–610.
- Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language*, 60, 1–19. doi:10.1016/j.jml.2008.09.001
- Grossman, M., Koenig, P., Kounios, J., McMillan, C., Work, M., & Moore, P. (2006). Category-specific effects in semantic memory: Category-task interactions suggested by fMRI. *NeuroImage*, 30, 1003–1009.
- Hoening, K., Sim, E.-J., Bochev, V., Herrnberger, B., & Kiefer, M. (2008). Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience*, 20, 1799–1814. doi:10.1162/jocn.2008.20123
- Johns, B. T., & Jones, M. N. (2011). Construction in semantic memory: Generating perceptual representations with global lexical similarity. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Expanding the space of cognitive science: Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 767–772). Austin, TX: Cognitive Science Society.
- Juhász, B. J., & Yap, M. J. (2012). *Sensory experience ratings (SERs) for over 5,000 mono- and disyllabic words*. Unpublished manuscript.
- Juhász, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology*, 64, 1683–1691. doi:10.1080/17470218.2011.605150
- Kan, I. P., Barsalou, L. W., Solomon, K. O., Minor, J. K., & Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*, 20, 525–540. doi:10.1080/02643290244000257
- Kellenbach, M. L., Brett, M., & Patterson, K. (2001). Large, colorful, or noisy? Attribute- and modality-specific activations during retrieval of perceptual attribute knowledge. *Cognitive, Affective, & Behavioral Neuroscience*, 1, 207–221. doi:10.3758/CABN.1.3.207
- Killgore, W. D. S., Young, A. D., Femia, L. A., Bogorodzki, P., Rogowska, J., & Yurgelun-Todd, D. A. (2003). Cortical and limbic activation during viewing of high- versus low-calorie foods. *NeuroImage*, 19, 1381–1394.
- Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41, 558–564. doi:10.3758/BRM.41.2.558
- Magnié, M. N., Besson, M., Poncet, M., & Dolisi, C. (2003). 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*, 25, 521–560. doi:10.1076/jcen.25.4.521.13873
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: WH Freeman and Company.
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, 58, 25–45.
- Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, 270, 102–105. doi:10.1126/science.270.5233.102
- McLeod, A. I., & Xu, C. J. (2011). *Bestglm: Best subset GLM* (R package version 0.33). Retrieved December 2, 2011, from <http://CRAN.R-project.org/package=bestglm>
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37, 547–559. doi:10.3758/BF03192726
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126, 99–130. doi:10.1037/0096-3445.126.2.99
- Medler, D. A., Arnoldussen, A., Binder, J. R., & Seidenberg, M. S. (2005). The Wisconsin Perceptual Attribute Ratings Database. Retrieved December 11, 2011, from www.neuro.mcw.edu/ratings/
- Millan, M. J. (1999). The induction of pain: An integrative review. *Progress in Neurobiology*, 57, 1–164.
- Nagasako, E. M., Oaklander, A. L., & Dworkin, R. H. (2003). Congenital insensitivity to pain: An update. *Pain*, 101, 213–219.
- Oliver, R. T., Geiger, E. J., Lewandowski, B. C., & Thompson-Schill, S. L. (2009). Remembrance of things touched: How sensorimotor experience affects the neural instantiation of object form. *Neuropsychologia*, 47, 239–247. doi:10.1016/j.neuropsychologia.2008.07.027
- Oliver, R. T., & Thompson-Schill, S. L. (2003). Dorsal stream activation during retrieval of object size and shape. *Cognitive, Affective, & Behavioral Neuroscience*, 3, 309–322.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8, 976–987. doi:10.1038/nrn2277
- Peelle, J. E., Troiani, V., & Grossman, M. (2009). Interaction between process and content in semantic memory: An fMRI study of noun feature knowledge. *Neuropsychologia*, 47, 995–1003. doi:10.1016/j.neuropsychologia.2008.10.027

- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, *15*, 161–167.
- Pexman, P. M., Holyk, G. G., & Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition*, *31*, 842–855.
- Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*, *9*, 542–549.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*, 1–15.
- Salmon, J. P., McMullen, P. A., & Filliter, J. H. (2010). Norms for two types of manipulability (graspability and functional usage), familiarity, and age of acquisition for 320 photographs of objects. *Behavior Research Methods*, *42*, 82–95. doi:10.3758/BRM.42.1.82
- Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica*, *7*, 221–262.
- Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body–object interaction effect. *Cognition*, *106*, 433–443. doi:10.1016/j.cognition.2006.12.011
- Siakaluk, P. D., Pexman, P. M., Sears, C. R., Wilson, K., Locheed, K., & Owen, W. J. (2008). The benefits of sensorimotor knowledge: Body–object interaction facilitates semantic processing. *Cognitive Science*, *32*, 591–605. doi:10.1080/03640210802035399
- Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*, 2802–2810. doi:10.1016/j.neuropsychologia.2007.05.002
- Stebbins, W. C., & Sommers, M. S. (1992). Evolution, perception, and the comparative method. In D. B. Webster, R. R. Fay, & A. N. Popper (Eds.), *The evolutionary biology of hearing* (pp. 211–227). New York, NY: Springer.
- Tillotson, S. M., Siakaluk, P. D., & Pexman, P. M. (2008). Body–object interaction ratings for 1,618 monosyllabic nouns. *Behavior Research Methods*, *40*, 1075–1078.
- Toepel, U., Knebel, J. F., Hudry, J., le Coutre, J., & Murray, M. M. (2009). The brain tracks the energetic value in food images. *NeuroImage*, *44*, 967–974. doi:10.1016/j.neuroimage.2008.10.005
- van Dantzig, S., Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: Norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods*, *43*, 145–154. doi:10.3758/s13428-010-0038-8
- Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*, 183–190. doi:10.3758/BRM.40.1.183
- Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, *64*, 956–972.
- Wu, L.-L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*, 173–189. doi:10.1016/j.actpsy.2009.02.002
- Wurm, L. H. (2007). Danger and usefulness: An alternative framework for understanding rapid evaluation effects in perception? *Psychonomic Bulletin & Review*, *14*, 1218–1225. doi:10.3758/BF03193116