Affordance Norms for XXXX Concrete Nouns

Nicholas P. Maxwell1, Alen Hajnal2, Mark J. Huff2, Julia Blau3, Brian Day4, Kerry Marsh5, Benjamin Meagher6, Chris Pagano7, Jack Shelley-Tremblay8, & Jeff Wagman9

1Midwestern State University, 2The University of Southern Mississippi, 3Central Connecticut State University, 4Butler University, 5University of Connecticut, 6Hope College, 7Clemson University, 8The University of South Alabama, 9Illinois State University

**Author Note**

Correspondence regarding this article should be addressed to Nicholas P. Maxwell, Department of Psychology, Midwestern State University, 3410 Taft Blvd, Wichita Falls, TX, 76308, United States. Email: nicholas.maxwell@msutexas.edu. The final set of affordance norms is available for download via our OSF page [LINK]. The normed dataset can also be accessed via our interactive Shiny application [LINK]. The authors thank Morgan Ballesteros, Samantha Garcia, and Madisyn Metaxas for their assistance with cleaning the final dataset.

Abstract

Objects are commonly described based on their relations to other objects (e.g., associations, semantic similarity, etc.) or physical features (e.g., birds have wings, feathers, etc.). However, objects can also be described in terms of their actionable properties (i.e., affordances), which reflect various interactive relationships existing between actors and objects. While several normed datasets have been developed to categorize various aspects of meaning (e.g., semantic features, cue-target associations, etc.), to date, norms for affordances have not been generated. We address this limitation by developing a set of affordance norms for 3000 concrete nouns. Using an open-response format, we computed affordance strength (AFS; i.e., probability of an item eliciting a particular action response) and affordance set-size (AFSS; i.e., total number of unique action responses) for each item. Because our stimuli overlapped with Pexman et al.’s (2019) Body-Object Interaction norms, we additionally tested whether AFSS was related to BOI, as objects with more perceived action properties may be viewed as being more interactive. These analyses, however, revealed a weak relationship between AFSS and BOI, suggesting that affordance properties reflect a separate construct. Thus, [SUMMARY SENTENCE]

Word Count: XXXX

*Keywords*: Affordances; Body-Object Interaction; Word Norms; Database; *R* Shiny

Affordance Norms for XXXX Highly Concrete Objects

Investigating questions surrounding memory, language, and perception requires an understanding of what words mean as well as the context in which they are used. While a word’s meaning can be operationalized in a variety of ways, researchers commonly rely upon two broad classes of concept information when assessing a word’s meaning. First, meaning can be discussed in terms of semantic similarity. This can be assessed using a variety of methods. For example, distributional models computing similarity based on word co-occurrences words within large bodies of text. Alternatively, feature-based models assess similarity in terms of shared features between concepts. To generate feature-based similarity, feature production tasks are used in which participants list features or defining characteristics of a series of objects. For example, when asked what features are inherent to a *chair*, participants might respond with *legs*, *back*, and *seat*. By employing these tasks within large-scale norming projects, large datasets of features can be used to compute the similarity between any two measured concepts (e.g., Buchanan, Valentine, & Maxwell, 2019a; McRae, Seidenberg, Cree, & McNorgan, 2005; see Kumar, 2021, for review).

Second, meaning is often described in terms of an object’s associations with other objects. While semantic measures are concerned with the degree of similarity existing between concepts, associations reflect the likelihood of one concept activating another in memory (see Nelson et al., 2000). To empirically measure the degree of association between concepts, researchers use free association norms (e.g., De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019; Nelson, McEvoy, & Schreiber, 2004). Like semantic features, word associations can similarly be generated through norming tasks. In a typical free association task, participants are presented with a cue word (e.g., *mouse*) and must respond with the first word that comes to mind (e.g., *cat*, *house*, *cheese*, etc.). Free association norms capture a wide variety of concept information, including semantic knowledge (e.g., *mice* are chased by *cats* and eat *cheese*) as well as linguistic information (*mouse* and *house* rhyme). Thus, while feature production norms and distributional models primarily assess meaning in terms of similarity, free association norms place greater emphasis upon the context in which words are used.

The past two decades have seen a proliferation of relatedness norms, which largely been due to advances in computing power and the ease of data collection afforded by online methods. As a result, large sets of feature production and free association norms have been made available for a variety of languages, with recent work focusing on ensuring sufficient overlap between existing databases (see Buchanan, Valentine, & Maxwell, 2019b). However, while both feature production and free association norms are important for understanding a variety of cognitive processes (e.g., memory, comprehension, perception, etc.), exclusively relying upon these measures may exclude other facets of meaning. For example, in addition to processing an object’s features and considering its common associates, individuals also process its potential *affordances* (i.e., actionable properties; Gibson, 1977; see Glenberg, 1997). Unlike semantic and associative-based measures of meaning, affordances describe the interactive relationships existing between actors and objects, rather than reflecting a specific feature or related object. For example, based on Nelson et al.’s (2004) free association norms, *pencil* and *ink* are the strongest associates of *pencil*, while based on Buchanan et al. (2019a), commonly produced features are *ink*, *write,* and *black.* Thus, while semantic and associative information may capture some aspects of object use (e.g., that *pens* are used for *writing*), they focus instead is on a wider variety of information (e.g., features which reflect key aspects of the object, associations which often reflect related objects, etc.), providing only a limited view of potential object uses.

As previously noted, much research on meaning has focused on how individuals represent concepts via internal processes (e.g., symbolic representations, memory processes, etc.). While understanding the cognitive processes underlying meaning-based representations is undeniably important, this approach potentially underplays the importance that environmental interactions also have on acquiring new knowledge. Within the past 20 years, a growing movement within cognitive psychology has sought to understand the impact of environmental interactions on knowledge acquisition (i.e., embodied cognition; CITE, CITE, CITE; see XXX for review). According to this perspective, the way individuals physically interact with objects strongly influences later representations. Thus, to fully understand an object’s meaning, measures which capture an object’s interactive properties are required alongside more traditional measures which describe an object’s defining features and/or associations.

Previous research has sought to address this limitation by developing measures of body-object interaction (BOI), which reflect the ease with which individuals can interact with a given object. For example, Pexman and colleagues (Pexman, Muraki, Sidhu, Siakaluk, & Yap, 2019; Tillostson, Siakaluk, & Pexman, 2008) have developed a set of BOI ratings, with the most recently published dataset consisting of ratings for 9351 words (Pexman et al., 2019). Unlike feature production and free association norms, BOI ratings were elicited via a 1-7 rating scale, such that higher ratings denotated greater degrees of interactivity. Thus, BOI ratings reflect the ease with which an object can be used, rather than describing actor-object relationships (i.e., affordances). Overall, these ratings have been shown to be consistent with existing measures of semantic knowledge. For example, Pexman et al. demonstrated that BOI ratings were strong predictors of responses in semantic decision tasks. Similarly, Heard, Madan, Protzner, and Pexman (2019) found that when combined with ratings of motor dimensions (e.g., graspability, ease of pantomime, and number of actions), BOI ratings explained a greater degree of variance in semantic processing tasks than when these ratings were used alone. Taken together, these studies provide further evidence that both interactivity and sensorimotor processes are critical components of semantic processing.

While it is evident from research employing BOI ratings that considering an object’s interactive properties facilitates semantic processing, [SOMETHING]

[BOI HERE – DESCRIBE + SHORTCOMINGS]

[ALTERNATIVE APPROACH – OPEN ENDED W/ MULTIPLE RESPONSE]

**The Present Study**

[GOAL OF THE PRESENT STUDY] [FIRST DETAIL THE CREATION OF THE DATASET, THEN THE ONLINE INTERFACE, FINALLY DESCRIBE A SERIES OF VALIDATION STUDIES]

**Method**

**Participants**

We recruited 3156 participants from two general settings. First, 2399 undergraduate students were recruited from 9 universities located within the northeastern, midwestern, and southern United States. The remaining 757 participants completed the study online via Prolific (www.Prolific.co). Table 1 displays final *n*s for each testing site following cleaning. All university students completed the study in exchange for partial course credit, while Prolific participants were compensated at a rate of $3.00 per 20-minute session. Prolific participants were required to have completed at least a high-school level degree or equivalent and to be native English speakers. For completeness, demographic information is reported in the Appendix (Table Ax).

**Materials**

To generate the stimuli, we initially selected 3005 nouns from the MRC psycholinguistic database (CITE). Words were initially selected based on concreteness, such that only high concreteness words were included (*M* concreteness ≥ xx). Of the 3005 words that were generated, five were randomly selected to serve as practice items. The remaining 3000 items were once randomized before being equally split into 100 lists which each contained 30 items. Overall, the final set of 3000 words had a mean concreteness rating of XX (Brysbaert, Warriner, & Kuperman), mean SUBTLEX frequency rating of XX (Brysbaert & New, 2009), and a mean BOI rating of XX (Pexman et al., 2019). Full descriptive statistics, including the percentage of stimuli overlapping with existing lexical datasets, are displayed in Table X.

**Procedure**

For all testing sites, data collection occurred online using Collector, an open-source platform for conducting web-based psychological experiments (Garcia & Kornell, 2015). Prior to beginning the norming task, particpants were informed that they would be viewing a series of object words and that they would be required to list as many uses for each object as they could reasonably generate. Participants were reminded that a single object typically has multiple uses and were encouraged to list multiple object uses when possible. To illustrate this point, the word *ball* was provided as an example, with *throw*, *bounce*, and *step on* all provided as examples of potential affordances. The complete set of instructions is available at [OSF link].

After receiving instructions, participants then completed a set of five practice items, which familiarized them with the norming task. For each trial, a cue word was presented in the center of the screen, and participants were instructed to type possible affordances into a textbox located directly below the word. To maximize potential affordances, participants were not given specific instructions on how to format their responses (i.e., tense, single words vs. phrases, etc.) with the exception that they were instructed to separate each unique affordance with a comma. Thus, participants were allowed to respond to the cue with individual words, phrases, or full sentences. Additionally, a prompt was located directly above the cue, which reminded participants to list as many uses for each object as they could generate. After completing the practice trials, participants immediately began the full norming task, which randomly selected one of the 100 lists of 30 items. These items were presented in a randomized order, and participant responses were self-paced. Following completion of this task, participants were debriefed. The total study took approximately 20 minutes to complete.

**Data Processing**

All responses were initially screened to ensure that participants adhered to the norming task’s instructions. Data from 35 participants were omitted due to excessive blank responses or failure to list object uses (i.e., consistently responding with synonyms or associates), leading to 3121 participants included in the final dataset. The remaining data were then processed in *R* using a procedure that was adapted from Buchanan, De Deyne, & Montefinese’s (2020) guidelines for processing lexical output from feature production tasks. Below, we first detail each step used to create the final dataset before describing the calculation of two affordance measures: Affordance Strength (AFS) and Affordance Set Size (AFFS). Given both the predicted size of the final dataset and because data collection occurred in waves across multiple testing sites, the data processing steps listed below were conducted across several batches of data, ranging from approximately 25 to 500 participants each. For completeness, an *R* script detailing the full cleaning procedure is available at [OSF LINK].

***Cleaning the Raw Data.*** Figure x illustrates the data cleaning procedure. We began by removing all blank responses along with any responses suggesting that participants were unfamiliar with a specific object (e.g., “I don’t know”, “unknown”, “unsure”, “?”, etc.). Across all batches, this process removed xx% of total trials. Second, because participants generally provided multiple affordances to each cue, each row in the initial dataset generally contained multiple affordances. The *tidytext* package was used to identify and separate individual affordance responses to each cue (De Queiroz et al., 2019). This parsing process assumed that unique affordances were comma-separated, though we additionally corrected for participants who did not follow instructions (e.g., semicolons, periods, spaces, etc.). This resulted in a long-format dataset, with each individual affordance having its own row in the dataset (i.e., “to drink from, throw it, pencil holder” become “to drink from”, “throw it”, and “pencil holder”).

After extracting individual affordances for each object, we next corrected for spelling errors using the *hunspell* package (Ooms, 2022). Because participants were primarily recruited from the United States, the spell check procedure utilized the American English dictionary. For British participants recruited from Prolific, British English spellings were changed to their corresponding American English counterpart (e.g., *colour* and *socialise* become *color* and *socialize*). After using *hunspell* to generate a list of spelling errors, all errors were visually inspected to confirm whether a flagged word was indeed a misspelling or simply a word which was not available in this package’s dictionary. Following the inspection process, all confirmed misspellings were corrected by replacing each misspelled word with its corresponding *hunspell* generated correction.

Once spelling errors were corrected, affordance responses were then tokenized via *tidytext*, which split each affordance phrase into individual words. This step was included to account for two potential issues. First, as noted in the *Procedure*, participants entered their responses into a textbox, which allowed them to list multiple affordances for each cue. However, participants often included extra spacing and tabs in addition to using a comma to separate affordance entries. Thus, the tokenization process removed any additional spacing and punctuation. Second, affordance phrases often contained multiple affordances, in addition to other context specific words (e.g., nouns and adjectives) which may also contain important information regarding object use. By splitting phrases into separate lines in the dataset, we were able to compare base affordances (often represented by verb responses) while also preserving the context in which the affordance occurs. Finally, the tokenization process allowed us to remove stopwords (e.g., *the*, *of*, *but*, etc.), which were dropped using the *stopwords* package (Benoit, Muhr, & Watanabe, 2021).

Finally, after tokenizing each affordance and omitting stopwords, the remaining responses were lemmatized and part of speech (POS) tagged. These steps were conducted in *R* with the *udpipe* package (CITE), which employs a trained language model to transform all tokens belonging to a particular set of lexemes (i.e., words with the same common meaning) into a shared lemma (i.e., *swim*, *swam,* and *swimming* become *swim*). We elected to use lemmatization rather than a stemming procedure since, as noted by Buchanan et al. (2020), a word’s stem may not always reflect a word existing within a particular language. Thus, our use of lemmatization ensured that all affordances in the final dataset were words existing in the English language. Finally, the model used for lemmatization was also trained to provide POS tags for wide variety of American English lemmas. However, to ensure accuracy, all tags were manually inspected. For lemmas which could potentially hold more than one tag (i.e., *fish* may be tagged as noun when referring to an animal but as a verb when referencing the lemmatized form of *fishing*), the context in which the original word was produced was used to determine the appropriate tag.

***Building the Affordance Dataset.*** Following the cleaning procedure, we inspected the dataset to ensure that all items had been normed by a sufficient number of participants. In doing so, we detected several low frequency cues which did not receive an appropriate number of responses (*n*s < 20). Eighty-five cues met this criterion and were subsequently dropped from the dataset. Additionally, we encountered several cue items that were spelling variations of the same object (e.g., *ax* and *axe*) or singular and plural forms of the same concept (e.g., *noodle* and *noodles*). We combined responses across singular and plural items such that only the singular form was used, so long as changing an object’s plurality did not substantially alter its use. After dropping low frequency cues and correcting for plurals and alternate spellings, the final affordance dataset contained responses to 2825 cues.

Next, because affordances reflect actions, the affordance measures reported below primarily reflect verb responses. As such, we filtered the final dataset to remove all adjectives, adverbs, and interjections, which accounted for XXX percent of all participants responses. Noun responses were divided into one of two categories: Those which reflected specific object uses (e.g., responding to the cue *bowl* with *hat*, *book* with *doorstop*, etc.) and those which provided contextual information as part of a phrase (i.e., for the cue *meat*, participants might respond *cut with knife*. In this case, only the verb *cut* would be considered an affordance). We note that because participants often responded with phrases rather than individual words, most responses included at least one affordance. As such, removing all non-affordance responses removed XX percent of the total dataset. For completeness, a full dataset containing all participant responses, including contextual nouns, adjectives, and adverbs is available for download on our OSF page.

After removing all non-affordance responses, we calculated the sum of each unique affordance response to a particular cue and divided it by the sum of all affordances that the cue received. In doing so, our process for computing AFS mirrored that which is used to compute FAS values following free association tasks (e.g., Nelson et al., 2004). For example, if the cue *chair* received a total of 30 responses, with 15 responses being *sit*, 10 responses being *push*, and five responses being *stand on*, the AFS values for *chair* – *sit*, *chair* – *push*, and *chair – stand on* would be .50, .33, and .17, respectively. Thus, higher levels of AFS denote a greater probability that a particular affordance would be listed as a potential action for a cue, suggesting stronger relationship between cue and affordance. Finally, we calculated affordance set-sizes (AFSS) for each cue, which was denoted the total number of unique affordance responses for each cue item. The .csv file containing the final affordance norm dataset is available at [LINK].

**Shiny Application**

While the final dataset has been made available via our OSF page, we have also developed an interactive *R* shiny application, which can be accessed at: [LINK]. Because This application provides two sets of information. First, the top table displays information regarding each cue word, including mean BOI Rating (Pexman et al., 2019), Concreteness (Brysbaert et al., 2014), SUBLTEX frequency (Brysbaert & New, 2009), age of acquisition (AoA; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), length, AFSS, and the number of participants who responded to each cue. Next, the bottom table displays AFS ratings for all cue – affordance pairs. In addition to providing mean AFS values, we also provide mean forward associative strength values (FAS; Nelson et al., 2004) and cosine similarities (COS; Buchanan et al., 2019a) when available. For both tables, users can search and filter the dataset based on the values in each common, and options are provided for downloading each table as an Excel file or .csv, including any filters which may be applied.

**Results**

[WORDS HERE]

**Research Questions**

[WORDS HERE]

**Descriptive Data**

[WORDS HERE]

**Validity**

[WORDS HERE]

**Discussion**

[WORDS HERE]

**Conclusion**

[WORDS HERE]

**Open Practices Statement**

[WORDS HERE]

**Funding Declarations**

[WORDS HERE]

**References**

[FIRST ONE HERE]

Table 1. *Final Sample Sizes for Each Testing Site*

|  |  |
| --- | --- |
| Institution | Total *n* |
| University of Southern Mississippi | 1128 |
| Prolific | 756 |
| University of South Alabama | 365 |
| Midwestern State University | 254 |
| Hope College | 215 |
| University of Connecticut | 152 |
| Central Connecticut State University | 115 |
| Illinois State University | 73 |
| Clemson University | 41 |
| Butler University | 22 |

[TABLE 2]