THE UNIVERSITY OF WARWICK

Library Declaration and Deposit Agreement

1. **STUDENT DETAILS**

Please complete the following:

Full name: Patrick D. Thompson

University ID number: 0854553

2. **THESIS DEPOSIT**

2.1 I understand that under my registration at the University, I am required to deposit my thesis with the University in BOTH hard copy and in digital format. The digital version should normally be saved as a single pdf file.

2.2 The hard copy will be housed in the University Library. The digital version will be deposited in the University's Institutional Repository (WRAP). Unless otherwise indicated (see 2.3 below) this will be made openly accessible on the Internet and will be supplied to the British Library to be made available online via its Electronic Theses Online Service (EThOS) service.
[At present, theses submitted for a Master's degree by Research (MA, MSc, LLM, MS or MMedSci) are not being deposited in WRAP and not being made available via EthOS. This may change in future.]

2.3 In exceptional circumstances, the Chair of the Board of Graduate Studies may grant permission for an embargo to be placed on public access to the hard copy thesis for a limited period. It is also possible to apply separately for an embargo on the digital version. (Further information is available in the *Guide to Examinations for Higher Degrees by Research.*)

2.4 *If you are depositing a thesis for a Master's degree by Research, please complete section (a) below. For all other research degrees, please complete both sections (a) and (b) below:*

(a) Hard Copy

I hereby deposit a hard copy of my thesis in the University Library to be made publicly available to readers (please delete as appropriate) EITHER immediately OR after an embargo period of .... months/years as agreed by the Chair of the Board of Graduate Studies.

I agree that my thesis may be photocopied. YES / NO *(Please delete as appropriate)*

(b) Digital Copy

I hereby deposit a digital copy of my thesis to be held in WRAP and made available via EThOS.

Please choose one of the following options:

EITHER My thesis can be made publicly available online. YES / NO *(Please delete as appropriate)*

OR My thesis can be made publicly available only after.....[date] (Please give date)
YES / NO *(Please delete as appropriate)*

OR My full thesis cannot be made publicly available online but I am submitting a separately identified additional, abridged version that can be made available online.
YES / NO *(Please delete as appropriate)*

OR My thesis cannot be made publicly available online. YES / NO *(Please delete as appropriate)*

3. **GRANTING OF NON-EXCLUSIVE RIGHTS**

Whether I deposit my Work personally or through an assistant or other agent, I agree to the following:

Rights granted to the University of Warwick and the British Library and the user of the thesis through this agreement are non-exclusive. I retain all rights in the thesis in its present version or future versions. I agree that the institutional repository administrators and the British Library or their agents may, without changing content, digitise and migrate the thesis to any medium or format for the purpose of future preservation and accessibility.

4. **DECLARATIONS**

(a) I DECLARE THAT:

- I am the author and owner of the copyright in the thesis and/or I have the authority of the authors and owners of the copyright in the thesis to make this agreement. Reproduction of any part of this thesis for teaching or in academic or other forms of publication is subject to the normal limitations on the use of copyrighted materials and to the proper and full acknowledgement of its source.

- The digital version of the thesis I am supplying is the same version as the final, hard-bound copy submitted in completion of my degree, once any minor corrections have been completed.

- I have exercised reasonable care to ensure that the thesis is original, and does not to the best of my knowledge break any UK law or other Intellectual Property Right, or contain any confidential material.

- I understand that, through the medium of the Internet, files will be available to automated agents, and may be searched and copied by, for example, text mining and plagiarism detection software.

(b) IF I HAVE AGREED (in Section 2 above) TO MAKE MY THESIS PUBLICLY AVAILABLE DIGITALLY, I ALSO DECLARE THAT:

- I grant the University of Warwick and the British Library a licence to make available on the Internet the thesis in digitised format through the Institutional Repository and through the British Library via the EThOS service.

- If my thesis does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my thesis available in digital format and that this permission encompasses the rights that I have granted to the University of Warwick and to the British Library.

5. **LEGAL INFRINGEMENTS**

I understand that neither the University of Warwick nor the British Library have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights, breach of contract or of any other right, in the thesis.

Please sign this agreement and return it to the Graduate School Office when you submit your thesis.

Student's signature: Peter D. Thomas .... Date: 28 Feb 2013 ....

JHG 05/2011

Phonetic symbolism for size, shape, and motion

Patrick Douglas Thompson

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

University of Warwick, Department of Psychology

February 2013

Table of Contents

List of tables and figures

Acknowledgements

First, I would like to thank my supervisor, Dr. Zachary Estes, for all of his guidance and seemingly unending patience. Without him, this PhD would not have ever been possible. I would also like to thank the staff of the Psychology Department at the University of Warwick for allowing me the opportunity to do my PhD there. I also would like to thank my second supervisors, Dr. James Adelman and Professor Gordon D.A. Brown, and my mentor, Professor Neil Stewart, for their support.

I also would like to thank my mother for her ceaseless support and faith in my abilities, as well as my aunt, Mary O, for her support and wisdom. She is surely missed. I also would like to thank my friends back home, Jeff and Marisa for their encouragement during my PhD work, as well as all my friends I made in England.

Abstract

This thesis examines phonetic symbolism, the meaningful use of individual speech sounds to convey and infer size, shape, and motion. Chapter 1 presents a summary of the literature. Though there is evidence suggesting that phonetic symbolism exists and is pervasive, the literature presents several research opportunities. In nine experiments and one pre-test (total N = 357 participants), we use graded stimuli throughout, which is uncommon in the previous research. This use of non-dichotomous stimuli allows for the hypotheses that have arisen from a gestural model of language evolution and the Frequency Code to be more fully investigated. In the first set of experiments (Chapter 2), we demonstrate that phonetic marking for size is graded, i.e., it does not mark just very large and very small objects. In Chapter 3, the focus is on marking for size and shape, and their possible interactions. We show that marking for size and for shape are not as in line with each other as previous works might suggest. Marking for movement is the topic of Chapter 4, which includes moving stimuli, not just implied motion. We find that trait permanence is at play with the naming for motion tasks, with marking only occurring when naming the motion itself. Finally, a concluding chapter summarizes and further expounds on the results of the thesis, and how those results relate to the hypotheses suggested by gestural models and frequency code. The conclusion also includes a section of current and future research directions.

Chapter 1:

Literature Review and Introduction

A. What is sound symbolism?

Sound symbolism is the connection of sound and meaning. One of the more direct of these connections is onomatopoeia, whereby the sound is a direct imitation, e.g., the "gurgle" of a baby. A less obvious example of sound symbolism is the topic of this thesis- phonetic sound symbolism. Phonetic sound symbolism is the conveyance or inference of meaning through individual speech sounds. An example of this could be the words "cat" and "dog." Although these words do not directly imitate the sounds cats or dogs make, the comparatively high pitched /k/ and /t/ sounds in "cat" could connote smallness, while the lower pitched /d/ and /g/ in "dog" may connote largeness (Berlin, 2006).

B. "I want to see a negative before I provide you with a positive" (Scott, 1982)

Although it may be old-fashioned to directly state the null hypothesis, it can also be useful- especially when the null has also been the prevailing thought for much of the last few millennia. Here, the null hypothesis is that there is no reliable connection between sound and meaning and that any connections are rare and certainly not universal. An early example of a text that espouses the null is Aristotle's "On Interpretation." Here, Aristotle defines a noun specifically as "a sound… of which no part is significant apart from the rest" (350 BCE, p. 1). As such, to Aristotle, phonemes would not have any meaning in and of themselves and they would not be used to convey any properties of the thing they are being used to name. Aristotle goes further, stating that meaning of a noun only comes when it is paired with the (universal) mental experience of the named thing. To put it another way, speech sounds are meaningful only when they make a word (in this case, a noun) that links to a mental experience.

In the widely influential *Course in General Linguistics*, Saussure, one of the founding fathers of modern and structural linguistics, expounds upon the null hypothesis. Although he disputes Aristotle's claim that the linking of a name to a concept is a fairly simple and straightforward process, he retains the assertion that sound symbolism is not widespread and is not a common part of the language process. Unlike Aristotle, who views the linking as being between a noun and a concept, Saussure states that the linking is between a concept (or signified) and a sound-image (or signifier, the psychological sensation of the name). Of note, the meaningfulness of this connection is only at the level of the sign (the signified plus the signifier), and its meaningfulness is not the work of any systematic usage of its basic constituents (i.e., phonemes). To wit, Saussure's Principle 1 is "The bond between the signifier and the signified is arbitrary" (2006, p. 67). In his definition of "arbitrary," he goes further, saying, "I mean that it is unmotivated, i.e., arbitrary in that it actually has no natural connection with the signified" (2006, p. 69).

Saussure concludes his section on the arbitrary nature of the sign by dealing with onomatopoeia, a seemingly obvious counter to his claims. First, he argues that these cases are much less common than one would imagine. Secondly, even when there is some apparent meaningful link, it is subsumed into the conventions of the general language (that is, it is forced to follow the phonological and morphological rules of the language), stripping or at least lessening the meaningfulness of the link. For this he provides and example of the sound a dog makes- in English "bow-wow" and in French "ouaoua" (2006, p. 69). Clearly, these two words are tied to the same concept, however, they are not the same. This seems to be a bit of a straw man. Although they are not identical, they surely have obvious similarities. For one, they both are composed of similar or repeating syllables. Secondly, they both have similar phonemes, especially the vowels, which are all back or near back. It seems, then, that if an English speaker were trying to convey to a French speaker that there is a dog nearby, saying "bow-wow" would be likely understood by the French speaker as indicating a reference to a dog.

Hockett (1960) listed arbitrariness as one of his design features of language. Though his list had many flaws, including its bias towards the views that language is solely a communication system and language can only be a spoken (and not signed) system, his list was widely influential. With the rise of Generative Grammar, the viewpoint of the phoneme as not being a meaning-bearing unit in and of itself continued. With the generativists, meaning lies at the level of the lexicon, with phonemes just acting as the means to represent these concepts (Chomsky & Halle, 1991). That is, phonemes are a part of the surface structure, and not a part of the meaningful deep structure (Newmeyer, 1992). Steven Pinker, echoing the sentiments of Saussure, refers to phonetic symbolism as "a quaint curiosity" (2000, p. 162).

C. "Rather ridiculous and yet plausible" (Plato, 360 BCE, p. 19)

Despite Saussure's claim that "no one disputes the principle of the arbitrary nature of the sign" (2006, p. 68), the idea of there being a system of meaningful use of sound in language goes back at least to the Cratylus dialogue (Plato, 360 BCE). In the dialogue, Socrates puts forth the notion that a name is not just to specify one thing from another (e.g., Bill from Ted), but also to specify the characteristics of the thing. A good name, then, has a "natural fitness…[as] an instrument… of distinguishing natures," just as any good weaving tool works with its specific material (Plato, 360 BCE, p. 6). As this fitness is built upon the components of the name, even in cases whereby the name is subject to the conventions of another language, "if you analyze them, a meaning is still discernable" (Plato, 360 BCE, p. 19). So, names can (and should?) provide information about the nature of the named, and the way in which this information is conveyed would be recognizable across languages.

But how would such a natural fit work? One early explanation comes from von Hornbostel, who views words not as an abstraction, but as an event- "a happening in sound" (von Hornbostel, 1927, p. 87). These "happenings" are the incorporation of the linguistic sign with the perceptual systems that go into the word's production and reception. This should not be confused with Saussure's view of the unified sign. Rather, "happenings" are synesthetic in nature, with each of its constituents bearing its share of meaning. As von Hornbostel puts it, "even isolated single sounds still have a sense" (1927, p. 87).

Other early support for sound symbolism comes from Paget (1929). For Paget, speech is like a game, originating from pantomime (gesture). As the game became more sophisticated, the speech organs started to take part. As we use gesture to supplement our speech (i.e., paralinguistic gestures) today, Paget posits that we began vocalizations as supporting our gestural language. As such, the principles that were at play with gesture became the basis of our use of speech, i.e., speech as vocal pantomime (1929). An example of this could be that outstretched arms signifying largeness would be converted to a greater opening of the mouth when making a sound to signify the same.

In addition to contributing the idea of speech as vocal gesture, Paget notes that these vocal gestures should be an underlying universality across languages (as its origin is motivated and based on the physical) and that it could be tested via the use of synthetic words (Paget, 1929). He warns, however, that the use of sound symbolism is "imperfect" and could lead to "ambiguities, anomalies, and homophones" (1929, p. 283) .

Bolinger (1949) quite pointedly suggests that the taking of arbitrariness as a self-evident truth is the result of linguistics needing to simplify its scope of inquiry when it was a young science. Instead of taking the perspective of linguistic signs being devoid of meaning, he suggests that linguists should view language in terms of how language is actually used (including the physiology of speakers), that is "parts of the utterance correspond to parts of the event" (1949, p. 55).

D. Initial summation and some issues to consider in the experimental literature

There are two primary positions regarding sound symbolism. The first viewpoint is that sound symbolism may exist, but that it is not widespread, is not universal, and is conventionalized within a language just like any other word. Alternatively, sound symbolism may be a truly systematic usage of sound in a language that is common within (and possibly between) the world's languages. Additionally, it may be a crucial step in the evolution of our language, as well as the learning of language.

Although the previously mentioned proponents of language models that include sound symbolism may create a compelling story, without experimentally derived evidence, it is just a story. Many major questions remain. First, how common is phonetic symbolism, exactly and how much of the experimental data are the result of either anecdote or demand characteristics? What is the nature of phonetic symbolism (i.e., is it just that there is a connection between sound and meaning when the link is physically motivated, or are there other kinds of meaningful links)? What does it mean to be universal and what kind of experiments would support or reject such an assertion?

The next two sections will focus on experimental and quasi-experimental evidence regarding phonetic symbolism. Section 4 will consist mainly of works prior to the 1990's and will focus on experiments regarding the existence and ubiquity of phonetic symbolism. Section 5 will primarily focus on works that are post-1990 and have as foci what the origin of phonetic symbolism could be, what the exact nature of phonetic symbolism is, and how phonetic symbolism could fit into language evolution and language acquisition models. It will begin with a section that shows how gestural and cross-modal models have advanced beyond intuition and then go into current experiments that examine phonetic symbolism within that context.

E. Experimental and quasi-experimental evidence for widespread phonetic symbolism

i. Corpus linguistics, both within and between languages

Otto Jespersen, in two major works, finds many examples of phonetic symbolism across several languages (1922, 2010). Though he freely admits to a lack of systematicity in his studies, he is able to provide a sizable set of examples. In the paper "Symbolic Value of the Vowel i," he finds /i/ to be commonly found in words that are associated with smallness and small things (2010). He feels this arises from a natural, intuitive relationship between smallness and small cavity size of vowel, as in Paget (1929). The relationship between front vowels and smallness across languages was taken up and corroborated many years later by Ultan (1978) and Jurafsky (1996).

Though Jespersen believes that phonetic symbolism effects would be larger with young participants, presumably due to less language learning muddling the effect, he reminds us that "the influence of sound symbolism [is] not restricted to children and savages, even modern scientists and suffragists are under its spell" (2010, p. 288).

Jespersen expounds on his previous study in the "Sound Symbolism" chapter of his *Language- Its Nature, Development, and Origin* (1922). Instead of focusing only on smallness, Jespersen expands his inquiry to phonetic symbolism of movement, appearance, states of mind, and size and distance. One of the most striking points he makes regards how sound symbolism works. He suggests that in some ways, sound symbolism may be waning in language due to the pressures for words to assimilate to the language's conventions. However, it may also be that words could become more symbolic over time due to the increased survival fitness that sound symbolism brings. That is, words that have a phonetic fit with their meanings may be more useable (e.g., easily remembered), leading to more usage and more staying power over time.

Newman (1933) examined the phonemes of approximately 500 size-related words. The words were gathered from Roget's Thesaurus for entries such as "greatness" and "littleness," and were culled by 11 raters who removed all words that arose that were unrelated to size. Although Newman felt there to be little evidence of any sound symbolism within his corpus study, the findings were in the predicted direction. For example, "large" words averaged more long vowels than did "small" words, and "small" words averaged more short vowels than did "large" words (as one would predict from his experimental work, see Section iii of this chapter). With a larger corpus, it may be that these differences would have been more clear-cut.

Orr further examines vowel opposition, or as he calls it, "vowel antiphony" (Orr, 1944, p. 3). Specifically, he details the usage of front ("narrow") vowels when connoting sharpness, quickness and smallness, while back vowels tend to connote dullness, the sustained, and the large. Though Orr's work is anecdotal, Thorndike (1945) takes his lead and provides systematicity. By taking all non-proper monosyllables occurring with a frequency of at least once per million words, he was able to include almost two thousand words for rating according to size. The rating system included definite small and large, as well as probably small and large (which were included in the ratio formulas as ½ that of the definite sizes). The ratings ratios provided strong evidence supporting Orr's intuitions, with smallness being associated strongly with front vowels and largeness being strongly associated with back vowels. Thorndike, ever the visionary, also notes that a good use for this knowledge would be in the business world- a topic which has come to some attention in the past several years (e.g., Coulter, 2008; Coulter & Coulter, 2010; Lowrey & Shrum, 2007; C. Spence, 2011)

In 1966, David Heise reports an analysis of one thousand English words using the semantic differential, as laid out in, among others, Osgood (1962). Briefly, the semantic differential is a "scaling instrument which gives representation to the major dimensions along which meaningful reactions or judgments vary" (Osgood & Suci, 1955, p. 325). The major factors are Evaluative (which includes the highly-correlated pairs of good-bad, beautiful-ugly), Potency (large-small, strong-weak), and Activity (fast-slow, active-passive), hereafter referred to as E-P-A (Osgood & Suci, 1955, p. 331). Heise (1966) finds that there are many significant relationships between specific phonemes and the ratings of the words they come from, e.g., /k/ is significantly related to both activity (fastness) and potency (largeness). However, when compared to previous results derived from ratings of artificial words (specifically, from Newman, 1933 & Miron, 1961), the data from natural words do not line up at all, suggesting that there may be different mechanisms at work with phonetic symbolism in natural and artificial languages. Alternatively, phonemes in natural languages must account for more meanings than those in artificial settings (and may be more culturally-biased), which may account for the differences in the results.

In an interesting meshing of corpus studies and experimentation, Johnson (1967) asked participants to list as many words as they could think of that meant smallness and largeness. He then analyzed their responses, finding that, in terms of vowels, they produced frequencies that were in line with the subjective ratings from Newman (1933). This study suggests two things. First, magnitude appears to be highly related to vowels, but not necessarily to consonants, and second, subjective phonetic symbolism is probably more highly related to high frequency words than those that are not as common. In a follow up study using the words generated by the aforementioned participants, Huang, Pratoomraj, and Johnson (1969) found this effect to be similar with Mandarin Chinese and Thai speakers as well.

More recently, researchers have provided a good example of exhaustive and systematic corpus study. In an analysis of over half the world's languages, there appears to be higher instances of both phonetic symbolism and its similarities across languages for 40 basic terms (Wichman, Holman, & Brown, 2010). The researchers find in their "culture-neutral and stable words," e.g., "nose" and "breast," that just under 20% of the words in these languages display sound symbolism (Wichman, et al., 2010, p. 857). Crucially, this sound symbolism is the same across languages, and not just a case of each language having its own symbolism. However, given the small number of target words, the effect is still quite small as compared to findings based on other methods, suggesting that even with “culture-neutral” words, we find a strong influence of culture.

ii. The matching of words across languages

Tsuru and Fries report on anecdotal work involving 25 sets of two English and two Japanese words (1933). Hypothetical participants would be asked to match the meaning of the English word to its Japanese mate. An example set is English: 'old' and 'young' and Japanese 'wakai' and 'oitaru,' with 'oitaru' and 'old' and 'wakai' and 'young' being the correct match. Most of the word sets involve antonyms, though some involve other relationships, e.g., 'bird' and 'worm.' The researchers believe that "at least 75% of these pairs can be correctly guessed" (Tsuru & Fries, 1933, p. 283), though they report no data to back up their claims (and in Brown, Black, and Horowitzh (1955), it is reported that Tsuru and Fries' participants actually only got 69% correct- see Table 1).

Author	Language 1	Language 2	Presentation Method	% Correct
Tsuru[1]	English	Japanese	Auditory + Visual	69
Rich [2]	English	Japanese	Auditory	57.2*
Rich	English	Polish	Auditory	64.8*
Brown, et al.[3]	English	Chinese	Auditory + Visual	58.9*
Brown, et al.	English	Czech	Auditory + Visual	53.7*
Brown, et al.	English	Hindi	Auditory + Visual	59.6*
Brown, et al.	English	Chinese	Visual	61.9*
Brown, et al.	English	Czech	Visual	61.9*
Brown, et al.	English	Hindi	Visual	60.7*
Maltzman, et al.[4]	English	Croatian	Visual	54.8*
Maltzman, et al.	Japanese	Croatian	Visual	50.8
Maltzman, et al.	Croatian	Japanese	Visual	51.4
Brackbill & Little [5]	English	Japanese	Combined Conditions	50.3
Brackbill & Little	English	Hebrew	Combined Conditions	53.0*
Brackbill & Little	English	Chinese	Combined Conditions	49.9
Brackbill & Little	Chinese	Japanese	Combined Conditions	54.8*
Brackbill & Little	Chinese	Hebrew	Combined Conditions	48.1*
Brackbill & Little	Hebrew	Japanese	Combined Conditions	52.3*

Table 1. Selection of percentages of correct translations (adapted from Brown et al., 1955 & Brackbill and Little, 1957) by native speakers of English. If author specifies that the findings are statistically significant, the % Correct is marked with an asterisk. Chance levels were 50%.

[1]Tsuru & Fries, 1934; replicated by Maltzman, et al., 1956 and found to be highly significant
[2] Unpublished experiment, as reported in Brown, et al., 1955
[3] Brown, et al., 1955
[4] Maltzman, et al., 1956
[5] Brackbill and Little, 1957; percentages represent combined condition

Tsuru and Fries suggest that the supposed high success rate is due to Gestalt similarity. It would also seem likely though that similarity to words within English are also employed, i.e., similar words to the English words given as matches. For example, one could easily match ‘good’ with ‘yoi’ because ‘yoi’ is similar to the English ‘joy’ and ‘evil’ with ‘warui’ because ‘warui’ appears to be a conjuction of ‘war’ and ‘ruin.’ Put another way, it may not be that the stimuli words are organized in ways that the Gestalt principle of similarity is at work, but rather, that synonyms within the participants’ natural language have similarities to the unknown language pairs. As noted by Brown et al. (1955), as Tsuru spoke both English and Japanese fluently, such links, though likely inadvertent, were also likely unavoidable.

This paradigm was expounded on by Brown, Black, and Horowitz (1955), who made 21 pairs of words (mostly antonyms) that named sense experiences and were fairly common in frequency. This list of word pairs was then translated by blind translators into Chinese, Czech, and Hindi. This is a vital step as it makes the inadvertent loaded word choices seen in the previous studies less likely. However, the translators did know the meaning of both sets of words, and when read aloud, they could have given some tonal or other clues to which is the correct matching. Using the auditory plus written method, all three language conditions showed a significant effect. In a follow up experiment, using only the written stimuli, the participants were actually slightly more adept at choosing the correct pairings. As the authors note, this may be due to matching surface features of the words, such as word length (e.g., matching ‘many’ with ‘bahut’ and ‘one’ with ‘ek’). It could also be that participants in the written only section were given more time to complete the task (not feeling that they had to keep up with the speaker). It should also be noted that matching to words in other languages that are sound symbolic appears to be easier than when the words are not. For instance, ‘thunder’ and ‘lightning’ were matched correctly only about 30% in Chinese, but over 90% of the time in Czech. The Czech word for ‘thunder’ is ‘hrom,’ which sounds verge on the onomatopoetic.

Maltzman, Morrisett, and Brooks extended this work by having 25 stimulus words to be matched with one of the two stimulus words of another language, e.g., ‘bird’ matched with either ‘tori’ or ‘mushi’ (1956, p. 250). Their goal was to see if the order of presentation, i.e., which language was the single word and which was the pair, made any difference as compared to two pairs being presented. Upon statistical analysis, they found there to not be any significant differences in the error rates of the groups. As such, the groups were combined, and the success rate was 55.56%, which was highly significant. They extended this by adding Croatian words to be matched with English pairs, again finding a significant effect (success rate = 54.76%).

Interestingly, they then follow this up by asking a question- if there is some sort of universal similarity at work here, wouldn’t we see similar results when we give English speakers pairs of words to match from two unknown languages (Maltzman, et al., 1956)? In their third experiment, they used the same Croatian stimulus words with Japanese response pairs. The participants were not reliably more successful than chance would suggest, having just over 50% success rate, leading to $p > .60$. By reversing the order of presentation (Japanese words to be matched with Croatian pairs), they found more success, but still not enough to be significantly better than chance. By showing that a speaker's native language has to be a part of the experiment for the guesses to be better than chance, they were able to cast doubt on the idea of a universal similarity across language pairs.

Brackbill and Little (1957) point out some potential issues with the previous works. Specifically, these studies involve the inclusion of a written stimulus set, despite the fact that "the hypothesis concerns a correspondence between meaning and *sound*" (Brackbill & Little, 1957, p. 312). They also take issue with the choice of stimuli words (which seem to be only somewhat following the criteria set out by the original authors) and the choice of languages used (which are in many cases related languages). With these issues in mind, they suggest that the idea of a universal sound symbolism hasn't really been tested and that what the previous works show is that "naïve subjects can guess the English meanings of *certain* words from *some* foreign languages well above chance" – a far cry from evidence of a universal mechanism (Brackbill & Little, 1957, p. 313). From their paper on, much more attention is paid to the roles played by choice of stimuli and stimuli presentation in phonetic symbolism studies of this kind.

In their study, Brackbill and Little began by having four unrelated languages (Hebrew, Japanese, Chinese, and English) with the list of word pairs consisting of very common words. The words were translated by multiple translators to increase reliability (inter-rater reliability). Each participant completed 3 sets of 50 words, either an English – Foreign or a Foreign – Foreign set of words, with 3 levels of presentation being within subjects (audio, visual, audio and visual). In the English – Foreign word conditions, both languages and method of presentation were significant. Generally, English – Hebrew was most successful, while English – Chinese was slightly below chance. Within each language condition, there was a general trend towards better success when both audio and visual components were available, with visual-only outperforming audio-only. This finding would later be corroborated by Siegel, Silverman, and Markel in their comparison of audio vs. audio plus visual presentation (1967).

In Brackbill and Little's Foreign – Foreign conditions, there was a significant result for languages and an interaction between language and method of presentation. This interaction seems to stem from the conditions that included Chinese. In the Chinese – Japanese condition, auditory-only presentation was more successful than either of the other conditions, while in the Chinese – Hebrew, visual-only was the most successful (though it was still not above chance). This seems to suggest that when the task involves the native language of the participant, they succeed more when they have as much information to make links as is possible. When the task does not include the native language, though, similarity of sound or of transcription seems to be most beneficial. Interestingly, this appears to suggest that the only true case of sound symbolism being at play is with the Chinese – Japanese auditory-only condition.

Brown and Nuttall (1959) take another shot at foreign word pairings, this time supposedly with the critique of Brackbill and Little in mind. Unfortunately, they fail to make several major changes that were needed. First, they again inform their participants that the experiment is about sound symbolism, including a general definition of the phenomenon. Second, though Brackbill and Little specifically took experimenters to task on including written presentation in *sound* symbolism studies, Brown and Nuttall present stimuli words as both audio and written word. What they do seem to have reliably shown here, however, is that there is still an effect of having English as a component when the participants are native speakers of English. Additionally, they show that there is a major effect of knowing whether the participants know they are dealing with antonyms or synonyms, with their "same-different" condition showing almost a nine point lower success rate than the regular English- Foreign pair condition. However, in a study by Weiss (1963), English – Hindi and English – Chinese mixed pairs are guessed at almost exactly the same success rate as antonymic pairs within those conditions. This may be yet another example of the effect one's choice of stimuli can elicit.

Atzet and Gerard (1965) investigate the effect of having the participants' native language in foreign word pairs (Chinese or Hindi) with native speakers of Navajo as their participants. They followed the format of asking "which of the two foreign words mean _____?" (Atzet & Gerard, 1965, p. 526), with all stimuli presented verbally. Surprisingly, though the native tongue of the participants was used in the experiment, no significant results were generated. In the case of both languages, correct guesses were just barely above chance. It seems likely that two major factors are at work here. First, as with all foreign pairings, using only auditory presentation will yield lower results (and at the time of Atzet and Gerard's paper, there was no written Navajo language). Secondly, as we have seen in previous works, some languages lend themselves to matching better than others. Navajo may not be a language that is easily matched (which comes as no surprise when one remembers its use as a code during World War II).

Slobin (1968) sets out to answer two nagging questions that remain from the previous studies. First, he examines whether matching success is facilitated by the antonym pairs being sense-based, and secondly, whether this is due to the words being at opposite ends of magnitude scales. Antonym pairs were taken from each of the three semantic dimension factors, with the addition of a few pairs that had been noted to be magnitude based (1968, p. 302). These word pairs were then translated into Thai, Kanarese, and Yoruba by translators naïve to the goals of the experiment. Word pairs were presented in one of two conditions: either audio and written or written only.

All conditions were shown to be significantly above chance, with there being only small differences between the results of the two presentation methods. Crucially, pairs that involved sensation were not any more or less successfully matched than those that did not. Additionally, antonyms that lend themselves to being on magnitude scales also did not perform much differently than those that were nonmagnitude (actually, nonmagnitude success rates were slightly higher). In a comparison of the three E-P-A factors, it appears that words associated with the activity factor were most easily matched. Though potency and evaluation words trailed far behind the activity words, they were still significantly above chance.

There are some potential problems with Slobin's findings. First, the pairs that he identifies as nonsensible do not appear to be nonsensible at all, e.g., calm-excitable, clean-dirty, happy-sad, tense-relaxed. It is easy to imagine the sensations of each of these "nonsensible" words, and the link between sensation and emotion is well documented both within and without the semantic differential, e.g., even word pairs that are mostly loaded on a non-evaulative factor still commonly have some loading on the evaluative factor as well, Osgood and Suci (1955). Secondly, although he implies that only some of the word pairs exist along a magnitude scale, they do, in fact, all name scalar ends, as can be seen in the original Osgood and Suci (1955) factors. In the end, then, what Slobin has shown is that with Thai, Kanarese, and Yoruba, pairing with English antonyms can be successful and that presentation method does not appear to be a major issue with these languages. Additionally, he has shown that all three E-P-A factors are successfully matched, with activity being most successfully so.

iii. Proper names

Phonetic symbolism in proper names was discussed by Plato (360 BCE), and despite there not being a lot of early quality works on the topic, a few papers do deserve some mention. The first is by English (1916), who replicates and expounds upon a work by Kollarits (1914). (Please note, the Kollarits and Claparede articles are in French, and not being a competent speaker of French, I am relying on the descriptions within other works for its content.) While Kollarits focused only on the image that is elicited by an unknown proper name, English includes the opportunity for participants to give verbal reactions in his experiment. Though he finds there to be a lot of variance between both observers and the names they are assessing, he does agree with Claparede (as referred to in Kollarits, 1914, p. 432) that "the sound of the name has an affective tone which co-operates in the elaboration of its mental representation." There seems, however, to be little evidence of this in the data tables that he reports.

Alspach (1917) continued this idea with an experiment that involved using stimuli from English, as well as one of the participants of the English study (not to put too fine a point on it, but there was one participant and he had participated in a highly similar study already.) The single trained observer/participant, hereafter referred to as "O," reported using the sound within the name specifically to make an impression of the person with that name in about a third of the cases, e.g., with the stimulus of 'mavquawpunt,' the O "repeated the word a number of times and got the feeling for a jouncing, jumpy movement" (Alspach, 1917, p. 438).

Much more recently, researchers have made several examinations of proper names, e.g., Whissell (2001). One such study was done by Cassidy, Kelly, and Sharoni (1999), who found that phonemic typicality of gender, e.g., female names commonly having more vowels and male names are more likely to end in stops, facilitated participants' accuracy and response times when identifying the gender of the person's name. Additionally, they found that using nonwords, both children and adults were prone to matching a name with a male doll if the name ended in a stop and a female doll if the name ended in a vowel. Additionally, in an English language corpus study of names and sex, researchers found that sex was marked in similar ways to size. That is, men, who are commonly larger than women (i.e., through sexual dimorphism), were more likely to have names that were lower in pitch, while women were more likely to have names that connoted smallness (Pitcher, Mesoudi, & McElligott, 2013). Of note, these pitch differences also align with the differences in vocal pitch of men and women.

iv. Phonetic symbolism when using Nonwords

In his seminal work, Sapir provided participants with meaningless word pairs and asked them to imagine which of the words would go with a larger or smaller item, e.g., ‘mil’ and ‘mal,’ which is the smaller table (1929, p. 227). There were 60 such pairs of nonwords, with each pair only differing by one vowel. It is not entirely clear how many participants there were in the three experiments he reports on- as he puts it, “it would be quite impossible to report all the details of the experiment in this place” (1929, p. 230). However, it does appear that 500 people participated in the second experiment, with most of them being schoolchildren between the ages of 11 and 18. As such, this experiment will be our focus.

A vowel inventory of 11 vowels was used, with the researcher pronouncing each non-word pair and the participants marking their responses on paper by placing a check mark next to the 1 or 2, e.g., if the participant believed the first was larger, the 1 would be marked. Sapir found there to be a high level of agreement across items and between participants and age groups. Additionally, he found “that the symbolic discriminations run encouragingly parallel to the objective ones based on phonetic considerations” (1929, p. 233). That is, roundness and backness were related to largeness, and vice versa, in the participants’ responses.

Although these results are indeed encouraging, one must also consider the possible drawbacks of the study. First, as previously pointed out, Sapir freely admits that he did not report all the details of his experiments. As such, it is unclear what he omitted and whether the omissions were of the “what time the experiments took place” variety or the “there were several unsuccessful experiments” variety. Similarly, he only reports some of the data, and then only using descriptive statistics. Though what he reports suggests a high level of agreement, it is difficult to know how selective his reporting was. Secondly, his choice of example stimuli is worrying. ‘Mil’ and ‘mal’ are not nonsense words. Rather, ‘mil’ is remarkably close to ‘milli-‘ and ‘mal’ is a common prefix for negative words, e.g., malevolent, malcontent. It could be that he just chose poorly what set of stimuli to make as an example or it could be that the example is indicative of poorly chosen stimuli.

In a well-done follow up study, Newman (1933) reports experiments involving non-word pairs with each non-word of a pair only differing in one vowel or consonant with its mate. In experiment 1, 606 students participated, with approximately a third falling into each age group (9-13 years old, 14-15, and 16 and up) and with the focus being vowel contrasts. With the presentation of each word pair, the researcher would ask which would go with the larger/smaller object, which was fitting of an adjective for a large/small thing, or which was more like an adult/child activity. In the data, he found there to be a surprising amount of agreement, both within and between the age groups. From these data, he was able to produce a set of scale values of the subjective size of each of the vowels used (see Table 2 for average scale values). As a result, Newman suggests, much like Sapir (1929) did, that the effect arose from the more psychophysical side of language comprehension/production. That is, he cites “articulatory position of the tongue," with front being smaller; "frequency of vocalic resonance," with higher pitch being smaller; and the "size of oral cavity," with small cavity being smaller (Newman, 1933, pp. 61-62).

Vowel	Average scale value (smaller value means smaller connotation)
i	.0000
e	.6647
ε	.8375
ä	1.4852
a	1.7126
u	1.7126
o	2.1080
ɔ	2.0967

Table 2. Average scale values of vowels based on subjective size. From Newman (1933, p. 58)

In Experiment 2, again, Newman had well over 100 participants. He adds as a focus consonants and includes as a second set of dimensions light/dark. (In this section, only size will be discussed, as lightness and darkness is better suited, via metaphor, to the affect section of this review.) As with the vowels in Experiment 1 (which he replicated in Experiment 2), assessment of size of consonants appears to be based on objective factors as well, specifically, that there is a trend such that small to large seems to go in the order of dental, labial, and palatal and also in the order of voiceless to voiced (1933, p. 68). He also notes that vocalic length is highly related to size. For the scale values of the consonants in Experiment 2, see Table 3.

Consonant	Average scale value (smaller value means smaller connotation)
p	.0000
n	.0254
d	.0555
s	.0800
h	.1960
k	.2243
b	.3950
l	.4497
g	.5901
m	.6354
dj	.8023
r	.8362
gl	1.0210
gr	1.2058
br	1.2464

Table 3. Average scale values of consonants based on subjective size. From Newman (1933, p. 66)

Bentley and Varon (1933) produced an almost simultaneous follow-up with the Newman (1933) paper. Unfortunately, of the many differences between the papers, the most obvious is the choice of participants. While Newman had hundreds of untrained participants in each experiment, Bentley and Varon employ three trained observers to participate in all of their experiments (that is, the same three observers in all the experiments). Additionally, the three observers were graduate students in the authors' lab. As with Sapir (1929), Bentley and Varon willingly admit that "the positive reports form only a moderate fraction of the whole number… [and] we have omitted all unrelated reports" (Bentley & Varon, 1933, p. 83). Despite this, they conclude that they were unable to demonstrate the sound symbolism effect that Sapir reports, unless "degree in some scale was suggested and prescribed" (Bentley & Varon, 1933, p. 83). This suggests two lessons from this paper- first, demand characteristics (i.e., due to suggesting a scale) is commonly a part of the reported effects, and second, it is difficult to convincingly show even a potentially strong effect with only three trained participants.

In an interesting follow up to these studies, Eberhardt (1940) replicated the Newman experiments with 92 deaf children as participants. In each trial, the children individually were asked to read two phonetically transcribed words aloud and tell the researcher which of the two was the larger/smaller item.   Though the results suggest that the ends of the spectrum of the scale values were similar, i.e., /ɔ/ was still considered to be largest and /I/ was still considered smallest, there were some significant differences between the hearing and deaf data (Eberhardt, 1940, pp. 33-34). Most notably, deaf participants viewed /a/ and /u/ as smaller than their hearing counterparts and /i/ as larger. Eberhardt provides a possible explanation to this- it is "probable that some of the deaf children were influenced rather more by kineasthetic than by vibro-tactile sensations that would more closely correspond to the acoustic sensations of the hearing person" (1940, p. 28).

Birch and Erickson (1958) created 54 nonwords of the format consonant-vowel-consonant (CVC) using an inventory of three consonants (/f/, /v/, /n/) and six vowels (/a/, / ɛ/, /e/, / ɪ/, /o/, /ju/). Seventy-four undergraduates rated each nonword for evaluative (clean-dirty), potency (large-small), and activity (fast-slow), mostly following the semantic differential method previously discussed (Osgood & Suci, 1955). Notably, participants made forced choices between the two parts of the factors, i.e., unlike the Osgood & Suci, the values were not rated along a seven-point scale. Presentation order seems to have played some role in the results, with activity being the only scale that had strong correlations to itself in all orders. Specifically, responses to evaluative and potency items appear to have been easily biased by previous items, leading to a lack of stability within participant responses for these domains. However, vowel sounds were significant across potency and activity. Though slightly different in order, both showed a tendency towards small/fast being front vowels and big/slow being back vowels. Vowels were not significant in terms of the evaluative factor. Consonants were not as clear-cut as vowels, with both onset and ending consonant being significant in terms of activity, only ending consonant being significant with potency, and only initial consonant being significant with the evaluative factor. Surely, a larger and more varied consonant inventory could have led to a clearer and more interesting data set.

Miron (1961), too, created CVC sets of (mostly) nonwords to be evaluated by his participants. In this case, participants were both native speakers of English with no knowledge of Japanese and native speakers of Japanese with varying knowledge of English. The rating scales, again, were based on Osgood and Suci (1955), and had three composite scales of evaluation, potency, and activity. The factor loadings of the two language groups correlated highly, suggesting that although there were some major differences in individual loadings, that the overall structure was similar (1955, p. 626). This should not be particularly surprising given that all the participants (no matter their native tongue) worked at the University of Illinois and likely used English both consistently and at a high level.

Taylor and Taylor (1962) take Miron's idea several steps further in their report of nonword stimuli with participants from four distinct language groups, i.e., English speakers in the United States, Japanese speakers in Japan, Korean speakers in South Korea, and Tamil speakers in India. An inventory of 12 consonants and 6 vowels found in all the languages was employed, with the sounds forming CVC nonwords. These nonwords were transcribed into the various languages and rated by the participants for size, movement, warmth, and pleasantness (so, similar to the common E-P-A of other studies but with the addition of the warmth dimension which is normally found in the activity factor). Though significant results were common within each language, there was little correlation between the scores between each language, suggesting that although the use of phonetic symbolism may be universal, the actual phonemes used in symbolizing may not be universal. This may be due to a number of factors, specifically that of the power of culture and the tendency to incorporate sounds into one's own native tongue's rules (regardless of physiognomic factors). This is made especially apparent given that the words are common within each natural language (and, thus, may be most prone to language/cultural pressures).

v. The matching of nonwords to abstract or unfamiliar items

One of the most famous (and most replicated) phonetic symbolism experiments was done by Kohler (1947). In an anecdotal experiment, Kohler showed participants two figures, one spiky and the other rounded (see Figure 1) and asked the participants which one was “takete” and which was “maluma” (1947, p. 134).

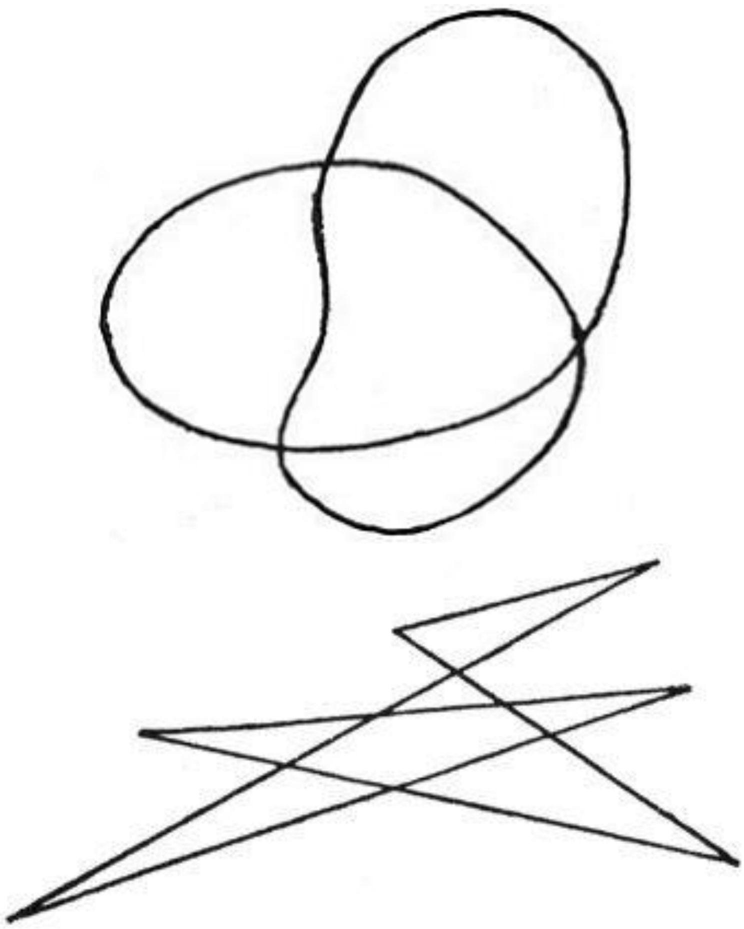


Figure 1.   *Maluma* (top) and *takete* (bottom), from Kohler (1947)

Participants very rarely failed to match 'takete' with the spiky object and 'maluma' with the rounded object. Kohler explained this as the effect of similarity across sensory domains, i.e., the roundness of the sounds in 'maluma' is matched to the roundness of the curved shape, etc. However, as there are only two items and there are no mentions of number of participants, it could easily be that something else is at play here as well. For instance, when one hears 'maluma' and 'takete,' they may also envision how they would be spelled. It could easily be that the shape of the letters in each name is what is being used to make the match, with actual sound having little or nothing to do with it at all.

Fox (1935) expands Kohler's work using several individual images, most of which have more than three possible names. Although that sounds like a step in the right direction, there are some potentially serious problems that deserve consideration. The possible names for each abstract picture were not matched in any way in terms of length or structure. Participants in all the reported experiments were the same five trained observers. Most worrying of all is what is found in the results section of the first experiment:

> "4 of the 5 Os… were somewhat confused by the instructions and found little meaning in the problem confronting them. 'It's all vague. There doesn't seem to be any method of telling which word should belong with which figure'… Very shortly, however, all the Os recovered from this initial confusion" (Fox, 1935, p. 556).

As it is unclear how much "training" was involved in this recovery, in addition to the other shortcomings of the article, it is probably best to limit our faith in Fox's findings. They do find, however, that there is much more agreement between their observers than one would expect by chance.

Irwin and Newland (1940) continue this line of inquiry with a study that involved the original Maluma/Takete names and figures, along with four other somewhat similar pairs. Unlike the maluma/takete names, Irwin and Newland see fit to use names that are unmatched in any way, such as form or length. Additionally, some of his nonwords are similar to actual words. For example, a word that should be paired with a haphazard set of slashes is "SKITZA," which reminds one of the base of "schizophrenia," while its paired item is a round swirl that should be paired with "LUN," which reminds one of "luna" (1940, p. 5). One item, "JIJ" appears to be a set of j's surrounding an empty center. The results spawning from these stimuli would likely be based less on phonetic symbolism itself and more on the ability for the participant to figure out these clues. It is of little wonder, then, that the older children who participated in his study showed a marked increase of performance as compared to the younger groups. For a non-phonetic symbolism study that uses similarly poorly-chosen stimuli, see Hall and Oldfield (1950).

Davis (1961) continued this line of research by doing a study involving two sets of drawings/word pairs, with participants that included children that are native speakers of Kitongwe and English. Kitongwe is a Bantu dialect and the African children used it at home and learned Swahili at school. As only the Kohler-type drawings are particularly related to phonetic symbolism, the data from these experiments will be the only ones listed here.

Additionally noteworthy, the non-word ‘maluma’ was changed by the researchers to ‘uloomu’ in order to avoid similarities with the Swahili word meaning ‘mother’s brother’ (Davis, 1961, p. 261). In the case of the English children, the possible names were only spoken, and the results were dramatically in the expected direction. In the case of the Swahili children, though, the results were not nearly as clear-cut. First, when the words were written down, order of presentation seemed to play a role, with only the words and figures presented in the same order showing a significant effect. In the spoken word only condition, only boys who saw and heard the stimuli in the positive order produced a significant result. Girls in both conditions did not produce a significant result, nor did boys when given the stimuli in the opposite order. Despite this, the author proclaims that “the overall effect is quite clear” (Davis, 1961, p. 264). It could be that in all actuality there is no universal tendency for the fitting of these names to these symbols. It seems more likely, however, that there could be some interference from within the Swahili culture or language that is affecting the results. It could also be that the Swahili children’s younger age and probable lesser reading skills could have led to their having less phonological awareness, and thus their not being as likely to produce the expected results.

Holland and Wertheimer (1964) used the previously discussed semantic differential (Osgood & Suci, 1955) to examine the Takete/Maluma(Baluma) phenomenon. They created a booklet consisting of four pages, with a name or a figure appearing on the top of each page. Participants were asked to rate each of the top items on ten semantic differential scales. Interestingly, 'takete' and the pointed figure had similar ratings, while 'baluma' and the rounded figure did as well. Additionally, when asked to directly rate fittingness of word to symbol, the Kohler results were reproduced and 'k' was rated as strongly fitting the 'takete' symbol and 'u' was rated as strongly fitting the 'maluma' symbol. This may suggest that there is a strong phonetic symbolism component or it may just suggest that the symbols look like k's and u's.

Tarte and Barritt (1971) presented eight participants with a pair of visual stimuli consisting of ellipses and triangles, which were matched for height and width in the inventory (but not in each trial). For each trial, the participant would see the visual pair and was asked which of the two items went with a (reportedly) nonword that followed the form of CVC, with /w/, /d/, and /k/ acting as initial consonants and /a/, /u/, and /i/ acting as vowels. The final consonant in all nonwords was /s/, which was meant to "reduce the meaningfulness of the items" (1971, p. 159), but only seemed to increase the similarity of the nonwords with real words, e.g., 'was,' 'kiss,' 'cuss.'

Tarte and Barritt (1971) found there to be the expected association between the front vowel /i/ and smallness and the relatively back vowel /a/ with largeness. Oddly, though, they found that the vowel /u/ was somewhere in between the two, and closer in subjective ratings to /i/'s pattern. The consonant data did not yield a particularly clear picture in terms of size. In terms of shapes, it appears that /a/ was not particularly preferred for either shape, while /u/ was strongly preferred for ellipses and /i/ was strongly preferred for triangles. Again, the consonant picture is not as clear-cut, but it does appear that there may be a preference for /w/ being matched with ellipses and /k/ being matched with triangles. Though these results are interesting due to the more controlled visual stimuli, it is disappointing that the auditory stimuli were so limited and confounded. Additionally, even with 252 trials per participant, it would still be a good idea to have more than 8 participants.

vi. Phonetic symbolism in the domain of affect

Though the evaluative factor in Osgood et al.'s semantic differential (1955) has been discussed in previous sections, this section will focus specifically on it, or more specifically, on the phonetic symbolism of affect. Affect is somewhat different from two other main factors, which speak more to sense and perception. However, as has been noted previously, "emotional experiences are sensory facts" Kohler (1947, p. 135). That being said, affective marking would prima facie appear to be less likely to be based on 'physiognomic' factors than potency and activity, and instead rely more heavily on something akin to magnitude scales, e.g., the clean-dirty dichotomy.

In an early experiment investigating affect done by Roblee and Washburn (1912), 15 female participants rated nonwords of the form vowel-consonant-vowel-consonant (VCVC) for pleasantness. The phonetic inventory consisted of 11 vowels and 17 consonants, semi-randomly mixed to create the stimuli, which were auditorily presented in a fixed order. They found the vowels /u/ as in 'mud' and /oi/ as in 'coin' to be least pleasant, while /e/ as in 'get' and /a/ as in 'father' were most pleasant. In terms of consonants, /g/ and /k/ were rated as most unpleasant and /n/, /m/, and /l/ were rated as most pleasant. Although there seems to be a general trend towards back vowels and plosive consonants being less pleasant and front vowels, fricatives, and nasals being more pleasant, there are some limitations that suggest that their findings may not be very generalizable. First, some of the small number of participants were "trained in introspection" and thirteen of them participated in the experiment twice. Second, although a seven-point scale was used, most scores fell fairly close to the middle. Without anything other than descriptive statistics, it is difficult to tell how much of the scores actually differed significantly.

In a parallel experiment to his size inquiry, Newman examined phonetic symbolism in the domain of lightness and darkness (1933), finding that although the symbolism followed a similar pattern as size symbolism did, there were some differences. Specifically, he found that there were three main factors involved in the judgment of size: pitch (lower pitch was associated with largeness), articulatory position (with backness being associated with largeness), and size of mouth cavity when making the sound (1933, p. 68). In terms of symbolism based on lightness/darkness, only articulation and frequency were used, with backness and lower pitch being associated with darkness. It appears that size of mouth cavity was not a factor in lightness/darkness symbolism.

E. L. Thorndike (1945) took words for pleasant and unpleasant concepts from six languages (English, German, Russian, Attic Greek, Finnish, and Hungarian) and analyzed their phoneme inventories for percentage of occurrences within pleasant words. In the analysis of English words, he found a general trend towards pleasantness being associated with front vowels (like / aɪ/ and /ɛ/) and unpleasantness being associated with vowels that were further back (like /ʌ/ and /aʊ/). In terms of consonants, there were significant results, but they are not easily simplified. That being said, it does appear that there is a trend towards stops being negative and nasals and palatals being positive (a finding supported by Demerse (1941, as reported in Johnson, Suzuki, and Olds, 1964).

Unfortunately, Thorndike does not report his findings for the other languages in the same manner as he did for English (with individual p-values), so it is somewhat difficult to glean what the data are saying to us. It does appear, however, that, again, we have examples of symbolism within languages, but not necessarily the same symbolism across languages. One interesting note is that Thorndike provides possible reasons for marking for affect to be the way it is, specifically, he notes that it may be due to sounds being pleasant because they are pleasing to the ear and they are easy to make, and vice-versa.

R. Johnson, et al. (1964) tested whether or not hearing and deaf participants would be able to reliably pair nonwords based on previous studies of affective sounds with English words. Their goal was two-fold. First, if it is sound that is most important in the effects shown in previous studies, then it should not matter if the matching is to a real word or not, just that the composition of the sounds is controlled for. Secondly, if the sound is what is driving this effect, responses from deaf participants should not be all that similar to hearing participants (as the deaf participants can only rely on the written component, not on sounding it out). Using the Demerse (1941) stimuli set as a jumping off point, they created 14 pairs of nonwords, each to be matched with an English word, with only consonants being manipulated. Their analysis shows there to be a significant effect in the expected direction for hearing participants, but no overall significant effect for deaf participants, that is, they had far lower levels of agreement amongst themselves. This study suggests that, unlike with dimensions like shape and size, that sound is vital in the success of affective phonetic symbolism studies.

vii. Tonal studies relating to cross-modality and phonetic symbolism

Tonal studies are of interest for several reasons. First, they may bypass a lot of the cultural/linguistic influences that may arise in studies that are solely language-based. Secondly, these studies may speak to one of the fundamental forces underlying phonetic symbolism, that is, the modulation of pitch.

In two early studies of this type, Solomon (1958, 1959) used samples of passive sonar recordings and elicited subjective responses about them from sonar operators using variations on the semantic differential scales. In the first study, fifty participants rated twenty sounds on fifty rating scales. Given their training, it is of no surprise that magnitude was the most loaded factor in the ratings (since the sonar operators are generally tasked with identifying what type of ship is being echolocated). Interestingly, though, common size dichotomies like “large-small” and “heavy-light” were also joined by “slow-fast” and “low-high” as highly loaded items on the magnitude factor (whereas they could commonly be on the activity factor).

In his second study, Solomon analyzed the stimuli and responses based on octave bands. Heaviness was associated with lower pitch and vice versa. Oddly, on the beautiful-ugly scale, only very deep and very high pitch sounds were considered beautiful. One could have predicted that it would have followed a trend similar to the heaviness-lightness scale, since beauty is commonly associated with youth and femininity, both of which tend to have higher pitched voices.

In a study by L. E. Marks (1974), participants were presented with grey squares of varying darkness and asked to find the appropriate pure-tone pitch that goes with the darkness level. All participants increased the pitch of the tone as the Munsell value increased (L. E. Marks, 1974). That is, as the grey got lighter, the pure tone chosen was of higher pitch.

In a more complex study, O'Boyle and Tarte (1980) employed two types of triangles, an ellipse, a circle and the Takete/Uloomu figures as the basic visual stimuli that were manipulated by shape, size, complexity, and density. The 48 pictures were then presented to participants who used an oscillator to generate the tones they felt went with each picture. A main effect was found for shape, with circles and ellipses being related to lower pitch than either triangle or the Kohler symbols. Interestingly, there was no main effect of size, e.g., larger symbols did not necessarily lead to responses of lower pitch. There was, however, an expected interaction found between size and shape, whereby larger *and* rounder were associated with lower pitch and vice versa. Of note, using this same method with native speakers of Urdu produced similar results (O'Boyle, Miller, & Rahmani, 1987).

Walker and Smith (1984) report two sets of experiments, with the first being a rating by participants for pure tones of various pitches along 17 semantic scales. They found there to be a tendency to rate higher pitches as more appropriate for high v. low, light v. heavy, little v. big, sharp v. blunt, happy v. sad, fast v. slow, as well as others. They then used these tones as underlying/incidental stimuli in a modified Stroop task (instead of text color as a distractor/enabler, the word’s position on the screen acted in that role).

For example, “high” should be most easily named when it is placed at the top of the screen (and has a higher pitch underlying it) and vice versa. The results showed that the incidental tone stimuli did, indeed, act to facilitate when aligned to the word and hinder responses when they were incongruous.

viii. Section conclusion and remaining questions

"The answer to the questions of universal phonetic symbolism will not be easily obtained and when obtained is not likely to be a simple one. There may, after all, be some associations of sound and meaning that are universally known and others that are a cultural product" (Brown, 1968, p. 128). In this case, Brown is clearly referring to 'universal' as having sounds that are commonly aligned across languages, and not that there is some sort of phonetic symbolism to be found in each language. Taylor and Taylor (1965) created a diagram to assist in the assessment of universality, and it is included here as Figure 2.

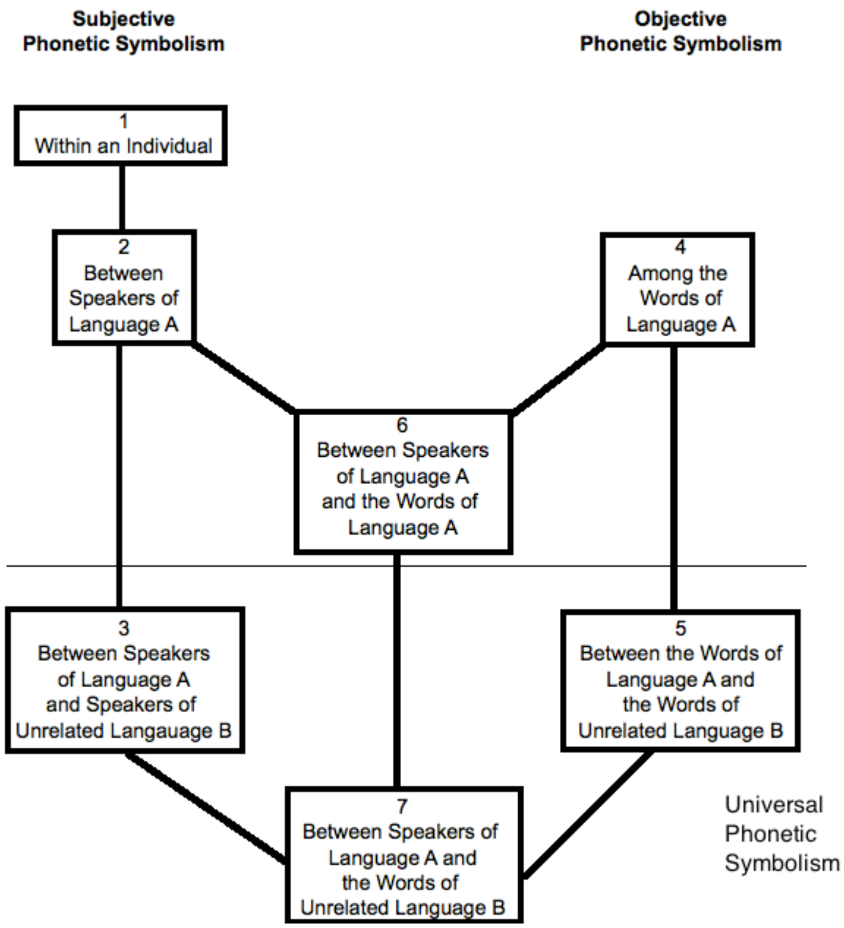


Figure 2. Seven questions of phonetic symbolism, with only questions 3, 5, and 7 suggesting universal phonetic symbolism (from Taylor & Taylor, 1965).

So in terms of a universal (cross-language similarities in sound connotation) phonetic symbolism, where do all these studies leave us? With the discussed corpus studies, there seems to be a lot of anecdotal evidence of between language phonosymbolism and at least some quasi-experimental evidence of it within English. However, as natural languages are the most likely to be affected by culture and the words within the language are most likely to be pressed by conforming pressures within language, it would seem that corpus studies are not likely to be the silver bullet in terms of showing universal phonetic symbolism. Additionally, as the computing power needed to do large-scale analysis of multiple languages did not exist until more recently, hand-picked anecdotes is the best we could hope for from this period. Even today, it is difficult to produce corpus studies that do not rely at least somewhat on experimenter choices, e.g., which of a synonymic pair should be included. Commonly, as the researcher or the translators know both languages, the sound symbolic/contrastive pair is included, e.g., large vs. tiny, while the alternate pair is not, e.g., big vs. small.

Similarly, the discussed cross-language word pair matching results have not been particularly convincing. For one, it appears that the effect derived by the presentation method may be as strong as the effect of phonosymbolism. Firstly, presenting visual stimuli may open the door to just attempting to match similarly shaped letters, leading not to sound symbolism, but rather, grapheme symbolism. Matching by surface features has been noted by several authors who suggest that word length is a primary consideration in these tasks. Secondly, by using forced choice pairs, it “causes the subject to presuppose the appropriateness of one choice or the other, when, in free association, neither might occur to him… and it forces the subject to think in terms of binary contrast” (French, 1977, p. 307). This, of course, also opens the door to the question of whether or not phonetic symbolism is based on a graded function (as one would suppose if it were based on kinesthesiology), or if it is just marking contrasts (as is suggested by Miall, 2001 and Taylor & Taylor, 1963). Finally, just as with corpus studies, how one chooses stimuli seems to be a particularly strong influence on the success of the word matching. Despite that, it does seem that there may something going on there. There are more than a few studies that show there to be an above average chance of guessing the right cross-language words. However, what it may suggest is that some languages lend themselves more to being matched and that phonetic symbolism may exist in all languages, but the sounds for each connotation may not be the same across languages. As Brown puts it, “if there are intersensory connections which are responsible for phonetic symbolism these must be common to mankind generally. One can postulate the existence of such innate connections but there is little one can offer in proof of them” (1968, p. 130).

It would appear that in terms of the discussed cross-language word pairings and corpus studies, he is absolutely right. It could also be that these studies do not do a good job of differentiating between statistical and motivated sound symbolism. As Jespersen puts it, “in some cases [phonetic symbolism] may have existed from the very first: these words sprang thus into being because that shape at once expressed the idea the speaker wished to communicate. In other cases the suggestive element is not original: these words arose in the same way as innumerable others whose sound has never carried any suggestion" (Jespersen, 1922, p. 408). By having designs that include both statistically-based symbolism (which is highly culturally dependent) and motivated sound symbolism (which should be based on universal sense and perception), one cannot be surprised to find a muddy picture of the phenomenon.

With phonetic symbolism within proper names, it may be that the symbolism is based more on gross imitation. Unlike with the original Bow-Wow theory (such as that described in Thorndike, 1943), though, this imitation need not be so, well, imitative. For instance, when creating names for a man and a woman, one could mirror the probable differences in the pitches of their voices (i.e., lower pitch for men) with the use of phonemes with appropriate pitch. (That is, a common type of sound comes from a thing, and so we will use a similar sound when making reference to the thing.)

With nonword studies, especially those that involve naming novel objects, and tonal studies, we seem to get the best evidence of a sound symbolism that could be the result of cross-modal influences, which in much of the previously discussed literature has been referred to nebulously as "physiognomic."

It is this possible source that will be the focus of the next section. Briefly, I will argue that language initially arose in the form of a gestural/mimetic language, and through cross-modal processing, developed into primarily spoken language. In between, and used as a stepping stone, were some classes of phonetic symbolism. Likely candidates are symbolism based on physical properties (size, shape) that could be easily imitated both in gesture and in speech. This underlying system is what we see being engaged when faced with novel words or the naming of novel stimuli, as well as when we ‘feel’ that a name ‘just fits.’ The statistically learned phonetic symbolism is a secondary function, and is more likely to be influenced by culture and experience. A prime example of this is that of phonesthemes, e.g., sn- words referring to the nose or mouth (snore, snort, sneeze), which have been shown to be not just found in abundance in language, but also that they are psychologically relevant (Bergen, 2004).

F. Phonetic symbolism as a stepping stone in language evolution

"A language may be likened to a cathedral. It may be Gothic to all appearances-surely externally it is and as one walks through the nave. But, if one gropes below, here are arches which are pure Roman and perhaps in some subtransept many Byzantine layered pillars" (Brown, 1968, p. 132). Brown spoke in reference to 'a language,' but it may be that his statements hold equally true for the *capacity* of language, and the idea of modern language being the summation of many adaptations (S. Pinker & Bloom, 1990), each leaving its mark in our cognition, is the one that will be argued here.

There are still many wildly differing views on how we came to have language. One of the major theories is that of language emerging via mimesis, that is, meaningful body movement that can be used for communication (among other things). Donald (1993) argues such a theory. In his model of language evolution, there is an emphasis not just of language as a communication tool, but also language as a way of representing and retrieving knowledge and memories. In his *Origins of the Modern Mind* (1991), he spells out three major cognitive transitions (though we will only concern ourselves with the first two here). The first transition is thought to have occurred with Homo Erectus, and it involved the movement away from simple episodic memory towards a system that supports symbolic thought and the retrieval of memories on demand. There are many reasons to think that such a shift occurred at that time. Homo Erectus had a much larger brain than his predecessors (and almost as large as ours), he had culture and advanced tools, and he was able to move out of Africa. Something clearly had changed, but there is evidence to suggest that it was not vocalized language, primarily that his larynx had not descended enough to support fully spoken language. Secondly, it is likely that he lacked the sufficient motor skills to produce the extremely quick and precise mouth movements needed to support speech. By attributing to him a mimetic language, we allow for the time for these developments to occur. Additionally, having a mimetic language can afford many cognitive upgrades. It gives one the ability to recall memories at will, to rehearse movement while remaining still, and to make linguistic/cultural transmissions, such as instruction on tool-making and use.

With the rise of Homo Sapien, the brain expanded once more, the larynx had descended, and the transition from mimesis to speech had begun. With these developments, words could be invented and spoken through the phonological apparatus, what Donald calls a "specialized mimetic subsystem" (1993, p. 739). This new system was built upon the old, with mimesis still being a part of our cognition, as is evidenced by our use of gesture paralinguistically and in times when our spoken language fails us (like when we are among those who do not share our language or when we have certain aphasias). Of note, Donald posits that there could have been many more smaller steps along the way (1993, p. 737). It could be that one of those steps could have been the meshing of the iconicity of gesture with the emergence of speech, resulting in a period of language that was dominated by phonetic symbolism. Even if such a period did not occur, it still seems to be the case that phonetic symbolism could be based in the “speech gestures” of our mimetic past (Berlin, 2006).

Though we have a more sophisticated gestural origin scenario, we are still faced with the question of how we moved from mimesis to speech, and from making a meaningful gesture to having a lexicon. Again, a viable answer seems to be that of cross-modal processing. Ramachandran and Hubbard suggest that, specifically, this cross-modality may be due to a crossing of auditory components of the brain with the motor cortex (2001, p. 19). As they put it, “this means that there would be a natural bias towards mapping certain sound contours onto certain vocalizations” (2001, p. 19), which is in line with the predictions one could make from a mimetic origin of language and what we see in sound symbolism studies. These natural biases, Ramachandran and Hubbard (2001) suggest, would make a move toward proto-language more feasible given the boot-strapping effect that it would have. Berlin concurs, stating “sound symbolism would be a good candidate to first drive lexical representation in spoken language” (2006, p. S38).

G. Some recent studies showing the usefulness of phonetic symbolism in language learning

In a study that included toddlers who were just beginning to speak, Maurer, Pathman, and Mondloch (2006) produce a variation on Kohler's takete/maluma experiment (1947). Crucially, they find that the toddlers (who were early in their language development) performed only slightly more poorly than the control adult group, that is, they used sound symbolism in naming at almost the same rate as the adults did. This is important because it implies that even with a limited vocabulary, these children produced the same shape-sound correspondences as those with adult vocabularies, suggesting that this is an innate tendency (Mondloch & Maurer, 2004), a finding that is corroborated by studies of preverbal infants using preferential looking, e.g., Walker, Bremner, Mason, Spring, Mattock, Slater, & Johnson (2010). Functionally, the study by Mondloch and Maurer also demonstrates that such sound-meaning correspondences could be a factor in the learning of one's natural language. Similarly, phonetic symbolism has been shown to enable learning of words in a foreign language (Nygaard, Cook, & Namy, 2009).

In addition to aiding word learning, phonetic symbolism may also support the learning of linguistic categories. Imai, Kita, Nagumo, and Okada (2008) investigate action words, the learning and generalizing of which children often find difficult. In a series of experiments, they used Japanese nonwords they showed to be sound symbolic of types of walking. The Japanese children who participated not only picked up on the sound symbolism, but were also better able to generalize the action when done by a different actor. These results were later replicated with English children, suggesting that this sort of bootstrapping may be universally useful in language learning (Kantartzis, Imai, & Kita, 2011).

Not only does there appear to be a mimetic component to learning word class, there may be a statistical one as well. In a study by Monaghan, Chater, and Christiansen (2003), they found there to be distributional differences in phonemes in nouns and verbs. That is, certain phonemes were more likely to show up in verbs than in nouns, and vice versa. Additionally, words that were more similar to those in their category were responded to more quickly in naming, lexical decision, and category tasks.

H. "Language is both arbitrary and non-arbitrary" (Bolinger, 1975, p. 22)

Though there need be no argument that arbitrariness exists in language, one may wonder why it is so prevalent given the benefits that phonetic symbolism may bring. First and foremost, phonetic symbolism seems to be most useful when the lexicon is small. As the size of the lexicon increases, the links between form and meaning start to overlap, leading to confusion and ambiguity (Gasser, 2004). Without the flexibility that arbitrariness brings, in order to keep a sizable amount of phonetic symbolism while having a large lexicon, the language would either require many more phonemes or many fewer concepts to be named. Surely, either is a detriment.

Additionally, although a nonarbitrary system seems to help in the learning of general categories (Monaghan & Christiansen, 2006), it also fosters intracategorical confusion, leading to difficulties in learning exact words (Monaghan, Christiansen, & Fitneva, 2011). Monaghan, et al. (2011) found in their simulations and corpus studies that this "division of labor" between the arbitrary and nonarbitrary produced the best situation for learning both broader categories and the words within those categories.

I. Questions that will be addressed in the experiments within this thesis

"One kind of phonetic symbolism is highly probable…that there is some kind of "appropriateness" or "inappropriateness" in new names which is common knowledge within a community" (Firth, 1935). It is this kind of appropriateness of new names that is examined in Chapters 2, 3, and 4, though we suggest that this type of fit is not just within a community, but rather is shown by data predominantly in one language community and is predicted by universal components of speech production and comprehension.

Chapter 2 addresses the issue of assuming that phonetic symbolism is based on graded functions, i.e., gesture and frequency code, yet doing so based on experiments that use pairwise/dichotomous stimuli. By using graded scales, we are able to ask whether phonetic symbolism is continuous or if it only is used to mark the ends of sensory/emotive scales. Two experiments are reported, both of which examine phonetic symbolism for size of novel objects, and both of which show there to be a linear relationship between the number of "large"/"small" phonemes in the name choice and the size of the novel object being named.

In Chapter 3, we examine the suggestion that despite being on different dimensional factors, size and shape have been shown to be marked with the same phonemes. In a series of experiments, we ask participants to rate the fit of auditorily-presented nonwords as names for novel stimuli of various shapes and sizes. Naming for size and shape are examined both independently and together, showing different trends in naming independently and an interaction between the two.

Motion has been identified as being marked- usually in tasks where the object is in fact not in motion. Is motion actually marked, and if so, what about motion is marked, e.g., speed, type of motion? In Chapter 4, we report a series of experiments that investigate the naming of novel objects in motion and one that involves naming the actual motion being made by the object.

Chapter 5 is a brief conclusion of the thesis and the experiments reported in it, as well as a set of further experiments that could be done by the researchers.

Chapter 2:

Sound Symbolic Naming of Novel Objects Is a Graded Function

(adapted from Thompson & Estes, 2011)

Although many recent studies indicate that names do indeed convey information about their referents' properties, methodological limitations common among those studies have so far constrained theoretical progress. Specifically, it remains unclear whether object names simply mark physical contrasts (e.g., small versus large), or whether they convey more finite gradations of those properties (e.g., small, medium, large). This issue is theoretically important in that evidence of a contrastive effect or a graded effect would implicate different evolutionary origins of sound symbolism, and possibly different roles that sound symbolism may play in the evolution of language. We therefore tested whether object names merely contrast small from large objects, or whether people are sensitive to degrees of largeness in object names. We then describe two alternative explanations of its occurrence, and we discuss why the prior research is unable to adjudicate between these possible explanations.

Explanations of Sound Symbolism

By a statistical version of sound symbolism, the relation between names and their referents is initially arbitrary, but with time that relation may become symbolic. That is, a given language begins with arbitrary pairings of sound and meaning, but it then evolves some phonetic systematicity. For instance, back vowels (e.g., /u/, as in "book") might initially be used to refer to objects of various sizes, but over time they may nevertheless become more common among large objects than among small objects. If so, this would constitute sound symbolism in that object names would convey information about the physical properties of their referents, and hence the relation between name and referent is non-arbitrary. However, there is no inherent connection between the name and the object. Rather, the relationship is simply statistical; these markings are only symbolic in that the phonemes are not randomly distributed among the words of the language. This explanation thus attributes sound symbolism to comparison (Berlin, 2006): Where semantic contrasts occur, they may come to be marked by phonetic contrasts. Indeed, some evidence supports this contrastive explanation. For example, Brackbill and Little (1957) presented antonymic word pairs from two different languages, and they asked participants to guess which words have the same meanings. They found that "where meaning contrasts are not as great, correct guessing becomes much more difficult" (1957, p. 318). Put another way, they found that when concepts were not at semantic extremes, e.g., thunder-lightning vs. light-dark, that correct guessing of foreign language pairs was less likely. Brown and Nuttall similarly concluded that "antonyms evolve toward phonetic contrasts appropriate to their semantic contrast" (1959, p. 444). The pressure to create such markings may extend past semantic contrasts to also include grammatical categories. Farmer, Christiansen, and Monaghan (2006) found there to be phonological differences between English nouns and verbs. They manipulated sound-category congruence, e.g., presenting a noun that has typical or atypical noun phonemes, speeding and slowing participants' reaction times (for a review of similar congruence paradigms, see Marks, 2004).

Alternatively, a cross-modal version of sound symbolism asserts that naming is directly motivated by the properties of the referent. That is, sound symbolism arises from the systematic matching of spoken sounds to physical properties in the visual or other modalities (Kovic, Plunkett, & Westermann, 2010; Maurer, et al., 2006; Ramachandran & Hubbard, 2001). Such cross-modal processing occurs when "the presentation of a stimulus in one sensory modality can be shown to exert an influence on our perception of, or ability to respond to, the stimuli presented in another sensory modality" (C Spence, Senkowski, & Roder, 2009, p. 107).

In the case of sound symbolism, the matching of auditory and visual modalities could emerge from a mimetic (gestural) system, a frequency (pitch) system, or both. Mimetic theories view symbolic sounds as speech gestures, whereby the mouth and vocal tract are used to produce a vocal "gesture" that mimics the physical properties of an object (Berlin, 1994). Frequency theories view symbolic sounds as signals of physical properties such as the dominance and size of the speaker (Berlin, 1994; Ohala, 1994). In terms of object naming, a symbolic connection is made via mouth shape (gesture) and/or sound pitch (frequency). For instance, the rounded mouth and low frequency typically used to pronounce /o/ may be directly symbolic of the roundness and largeness of the referent (Berlin, 2006). Note that this does not appear to be simple onomatopoeia, whereby a small thing is named with high pitched phonemes because it makes high pitched sounds. Rather, there is a connection between pitch and object size even when the objects are silent, as evidenced in the link between the size of fish and the pitch of sounds in their names (Berlin, 1994) and the congruence effect of pitch when judging the size of grey disks (Gallace & Spence, 2006). Thus, it is the size rather than the sound of the object that appears to influence naming.

Although the occurrence of sound symbolism is now well established, few studies have attempted to discriminate between these statistical and cross-modal versions. To date, the only extant approach to investigating these different hypotheses has been cross-language comparisons. If language evolved without sound symbolism, and sound symbolism emerged only later in the development of a given language (i.e., the statistical version), then sound symbolism should vary considerably across languages. Alternatively, if sound symbolism reflects a direct relation between name and referent (i.e., the cross-modal version), then sound symbolism for at least some domains should be universal or at least relatively constant across languages. That is, if sound symbolism is based in perception rather than language-specific contrast, then many of these symbolic sounds should be common across languages. Unfortunately, the evidence is mixed. On one hand, considerable variance in sound symbolism is observed across natural languages (Taylor & Taylor, 1965). Indeed, in some languages, sound symbolic markings are flipped, such as front vowels marking smallness in English but largeness in Bahnar (Diffloth, 1994). But on the other hand, there appear to be more similarities in sound symbolism across languages than one would expect by chance alone (Berlin, 2006; Maurer, et al., 2006; Nuckolls, 1999a). The ambiguity of the evidence may reflect the imprecision of the theoretical predictions as well as the correlational nature of the research methodology.

We propose new theoretical predictions that allow an experimental test of these alternative explanations of sound symbolism. By the contrastive explanation, sound symbolism only marks the opposite values of the given physical property, such as small and large. By the cross-modal explanation, however, sound symbolism may mark degrees of the given physical property, such as small, medium, and large. Just as physical properties like size are continuous, so too are gesture and frequency. For example, a midsized object could be indicated by a moderate hand or mouth gesture rather than a subtle or extensive gesture, and it could be indicated by a midrange pitch rather than a high or low pitch. Thus, the contrastive explanation predicts a categorical function whereby sound symbolism marks only small and large objects, but the cross-modal explanation predicts a graded function whereby sound symbolism differentiates medium from small and large objects. As described next, prior studies are not capable of discriminating between these predictions (nor were they designed to do so). We therefore report two experiments that directly tested these predictions.

Methodological Limitations of Prior Studies

Many early works in sound symbolism suffered from a variety of confounds (Taylor & Taylor, 1965; Westbury, 2005), most strikingly, strong demand characteristics. In the example of Sapir's *mil*/*mal* (1929), participants were asked to imagine two objects that only differ in size, and the chosen names have only one contrast, i.e., /i/ vs. /a/. Participants in this task are almost certainly aware of the two manipulations (i.e., object size and vowel), and there is a strong implicit demand for participants to confirm the experimenter's hypothesis. Additionally, many of the nonwords that were used as stimuli in prior studies may have been somewhat meaningful to participants. For example, *mil* is quite similar to *milli-*, a common prefix meaning "one thousandth", and *mal* is a common prefix meaning "bad or evil".

Although recent sound symbolism research is more refined, much of it uses pair-wise presentation of alternative names. Critically, this pair-wise nature does not allow one to discriminate the contrastive version of sound symbolism from the cross-modal version. For instance, Berlin (1994) examined sound symbolism by presenting pairs of names for birds and fish in the Huambisan language, asking American students to guess which name referred to a bird (or to a fish). The students' guesses were indeed more accurate than chance, thus indicating the presence of sound symbolism. Maurer, Pathman, and Mondloch (2006) tested participants' naming of rounded and spiky shapes using a small set of pair-wise choices. Both preschool children and university students showed strong preferences in matching certain sounds with specific shapes, such as using /k/ and /t/ to mark "sharp" objects. Even among studies with more items and more subtle methods, the stimuli tend to be dichotomous, such as round versus pointy shapes (Kovic, et al., 2010; Westbury, 2005) and small versus large figures (O'Boyle & Tarte, 1980). Unfortunately, these experiments are unable to determine whether sound symbolism is graded or whether it only marks opposite properties. As explained above, resolving this issue may provide evidence to sound symbolism's origin and function (i.e., contrastive versus cross-modal).

The Current Research

We report two experiments that test whether sound symbolism for size is categorical or graded. We presented a series of novel figures (i.e., Greebles) that varied in size, along with several nonwords that varied in the number of small-sounding and large-sounding phonemes, and we asked American and British undergraduates to choose the name that goes best with the given object. In contrast to most prior studies, the current experiments included many trials per participant and several possible names per trial. This methodology should increase reliability while decreasing demand characteristics. More importantly, the nature of the relationship between the size of the object and the proportion of small- or large-sounding phonemes in its preferred name will reveal whether the presumed sound symbolism is categorical or graded. The cross-modal version of sound symbolism uniquely predicts that small, medium, and large objects should be named with increasing numbers of large-sounding phonemes. We first tested this prediction with American undergraduates using written object name choices in Experiment 1, and for generality we presented British undergraduates with auditorily presented object name choices in Experiment 2.

Experiment 1

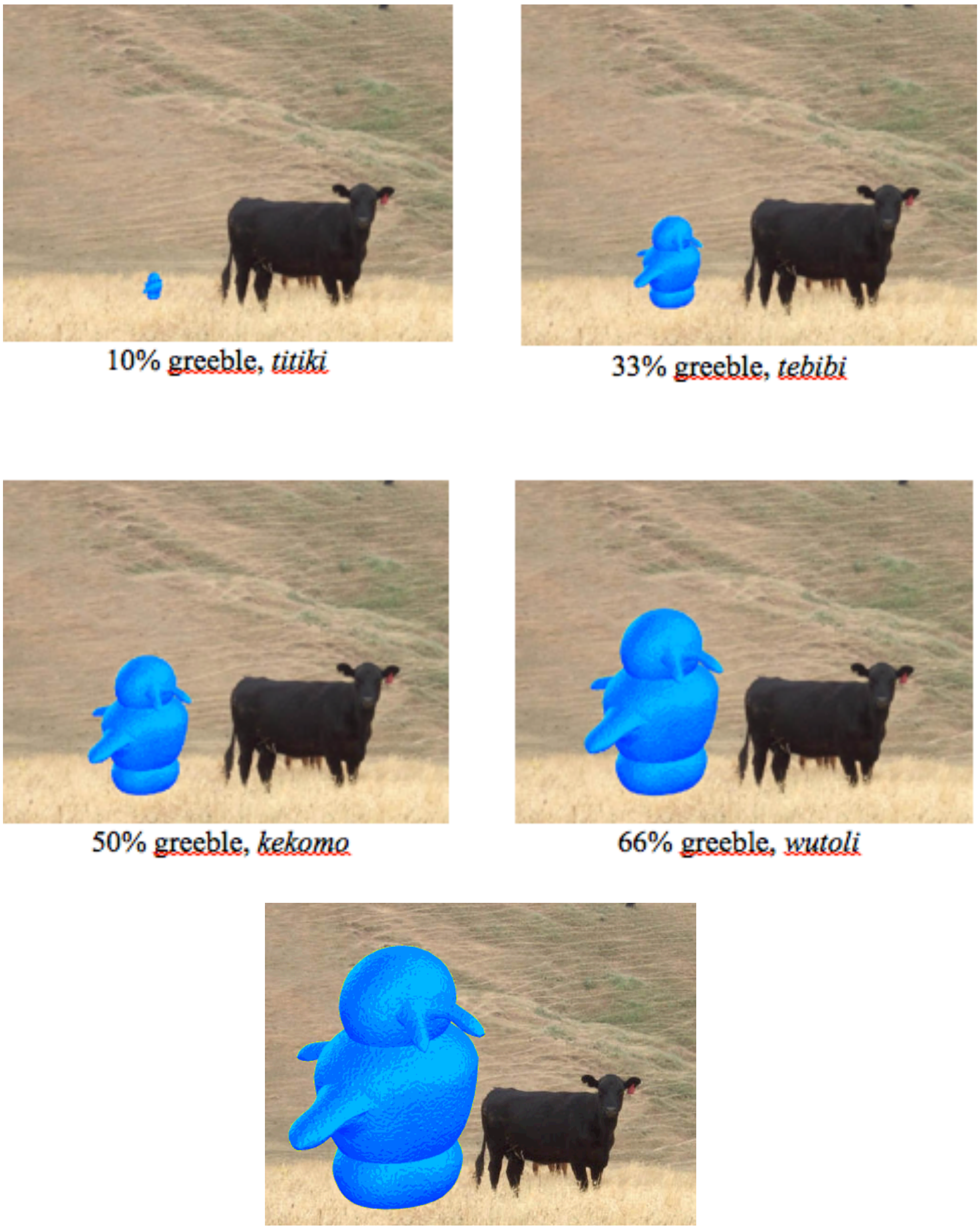
Participants viewed novel objects (Greebles) in one of five sizes. To provide relative size information, the greebles were embedded in a pastoral scene (see Figure 3). Each greeble was accompanied by a visually presented list of five possible names, which were nonwords that varied in the number of small- and large-sounding phonemes. Participants chose the name that was "most appropriate" for the given object. If sound symbolism is graded, then the size of the object should linearly predict the number of large-sounding phonemes in its preferred name.

Participants

Forty-seven undergraduates at the University of Georgia participated for course credit.

Materials

Twenty greebles were randomly selected from those provided courtesy of Michael J. Tarr, Brown Unviersity, http://www.tarrlab.org. The greebles were manipulated to appear in five sizes: The original greeble (100%) was shrunk to 66, 50, 33, and 10% sizes. Because presenting the greebles in isolation would render their relative size ambiguous, we used the GNU Image Manipulation Program to embed the greebles in scenes that would suggest that the greebles were of differing size, not of differing distance to the viewer. They were placed in a pastoral scene with a cow acting as a reference for size and distance (see Figure 3). All greebles appeared in the same position and at the same picture depth as the cow. The cow's size was roughly that of the 50% greeble.



10% greeble, *titiki*

33% greeble, *tebibi*

50% greeble, *kekomo*

66% greeble, *wutoli*

100% greeble, *gobudu*

Figure 3. Experiment 1: Examples of stimuli with name matching size of greeble.

One hundred nonwords of CVCVCV form were constructed. The nonwords consisted of varying numbers of large- and small-sounding letters, which correspond to phonemes that have been found in prior research to have size associations for "large" (a, u, o; m, l, w, b, d, g) and "small" (i, e; t, k) (e.g., Berlin, 2006; Maurer, et al., 2006; Newman, 1933; Taylor & Taylor, 1962). Nonwords were constructed as randomly as possible within the constraints, while minimising similarities to real words. They were semi-randomly sorted into twenty sets of five items, so that each set contained nonwords with five different numbers of large-sounding (or small-sounding) letters. An example of a nonword set is *wodolo* (6 "large"/ 0 "small"), *tibudo* (4/2), *kuloti* (3/3), *bitiku* (2/4), and *kitete* (0/6). Although presented here in descending order, in the actual experiment the nonwords appeared in random order.

Procedure

Each participant completed 100[6] trials, which appeared in random order. On each trial, a greeble of 10, 33, 50, 67, or 100% size appeared embedded in the background scene above five possible names corresponding to five levels of large-sounding letters. Participants were instructed to choose the most appropriate name for each greeble by pressing the key corresponding to the chosen response option (name 1, 2, 3, 4, or 5). Each greeble was randomly paired with a nonword set that remained constant for each of the greeble's five presentations (once in each of the five sizes).

---

[6] Participants responded to 100 items but due to programmer error, three greeble sets (15 items) were removed from analyses.

Results and Discussion

Recall that each nonword name contained between zero and six large-sounding letters. Thus, the dependent measure was the number of large-sounding letters in the chosen name on each trial. We calculated for each participant the mean number of large-sounding letters of the chosen name for each of the five object sizes. As illustrated in Figure 4, the size of the object linearly predicted the number of large-sounding letters in its preferred name. This relationship was confirmed via repeated measures ANOVA, $F(1.83, 84.336) = 73.587$, $p<.001$, partial $\eta^2 =.615$ with significant differences between each successive object size (all $p < .01$ by paired *t*-test).

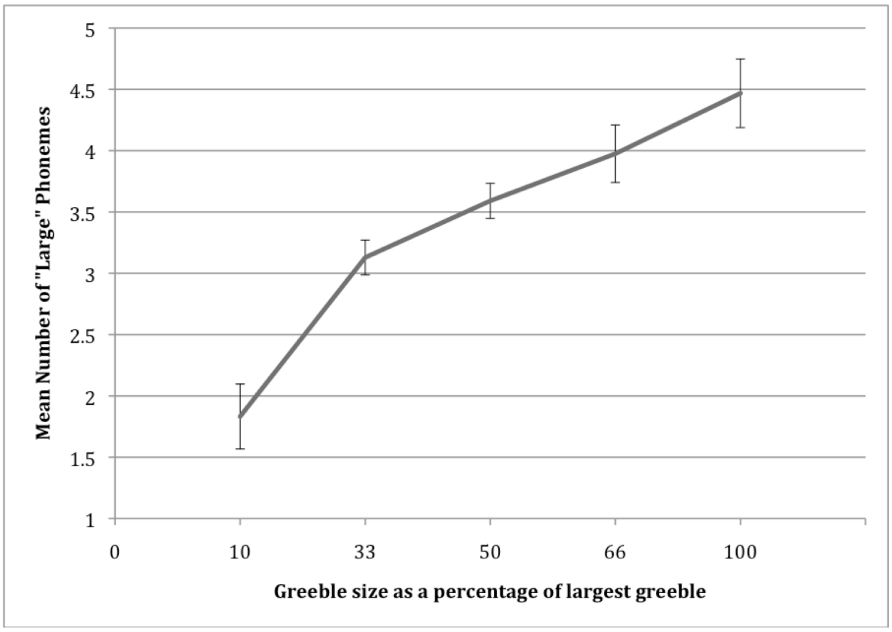


Figure 4. Experiment 1: Mean number of letters referring to “large” phonemes in the naming of different size of greebles. Error bars represent 95% Confidence Intervals.

These results suggest that participants used phonemic composition of names in a graded way to denote size. However, given the visual presentation of the names, participants might have chosen names based on visual cues from the graphemes rather than on the sounds of the names. Upon analysis, a strong correlation was found between number of large phonemes in a word and the word's width in pixels, $r(98) = .743$, $p < .001$. It is difficult to judge to what extent the width of the possible names affected participants' choices. Though a multiple regression could shed some light on the relationship between name and answer, the potential underlying problem of a study of *sound* symbolism using written words as stimuli remains. Although it has been debated in the literature, e.g., Atzet and Gerard (1965) and Taylor and Taylor (1965), whether auditory and visual presentation methods lead to substantially different results remains unclear. Though we feel that this relationship between phoneme and grapheme size is unlikely to fully explain Experiment 1's results, we nevertheless conducted an additional experiment with spoken rather than written names. Critically, the spoken names were matched for duration across conditions.

Experiment 2

Participants

Nineteen undergraduates at the University of Warwick participated for course credit. Twelve of these participants identified themselves as native speakers of English.

Materials

Participants listened to auditory stimuli on headphones. The volume of the headphones was approximately the same for each participant, and was determined by the experimenter. Participants were instructed to inform the researcher if there were any problems, including those with the level of the audio, and none chose to do so.

Auditory stimuli consisted of recordings of 42 nonwords spoken in a "standard" British accent by a male. Each nonword consisted of two CV clusters with an overall phonemic inventory of voiceless ("small") and voiced ("large") stops (/p/, /t/, /k/ and /b/, /d/, /g/) and front ("small") and back ("large") close vowels (/i/, /e/ and /u/, /o/) (Berlin, 2006; Newman, 1933). Clusters consisted of "large" and "small" CV pairs, e.g., [ki] "kee," [bo] "boh," such that a "large" name had two "large" clusters, a "medium" name had one "large" cluster, and so on. The 42 names were sorted randomly into 14 sets of choices, such that each had three levels of "large" clusters (e.g., all "large," half "large," zero "large"). The presentation order of the words was balanced. A one-way ANOVA on the duration of names (mean duration = 590 ms) found no significant difference between levels of "phonetic size," $F(3,38) = .925$, $p > .40$, thus indicating that the small, medium, and large names were matched for utterance length. Each of a trial's three name choices was denoted on the screen by a grey circle, measuring 46 pixels in diameter.

Visual presentation of greebles was also simplified, though still very similar to Experiment 1. Fourteen greebles were randomly chosen from the Tarrlab greeble set and manipulated using the GNU Image Manipulation

Program to convert the images to greyscale, resize them into small (mean = 94 x 53 pixels), medium (mean = 293 x 171 pixels), and large (mean = 491 x 287 pixels) sizes, and add them to a greyscale scene with an abstract human form (292 x 147 pixels) as reference. The medium sized greeble was approximately the same height as the human figure. The location of the greeble relative to the human (left or right) was counterbalanced.

The visual and auditory stimuli were presented via Microsoft PowerPoint, with participants marking their naming choices using pen and paper. Each of the 14 greebles was presented in each of its three sizes accompanied by the same set of name choices each time. Since this method does not lend itself to true randomisation, two presentation lists were created using the Random.org website. Figure 5 shows an example of a presentation slide with a medium sized greeble

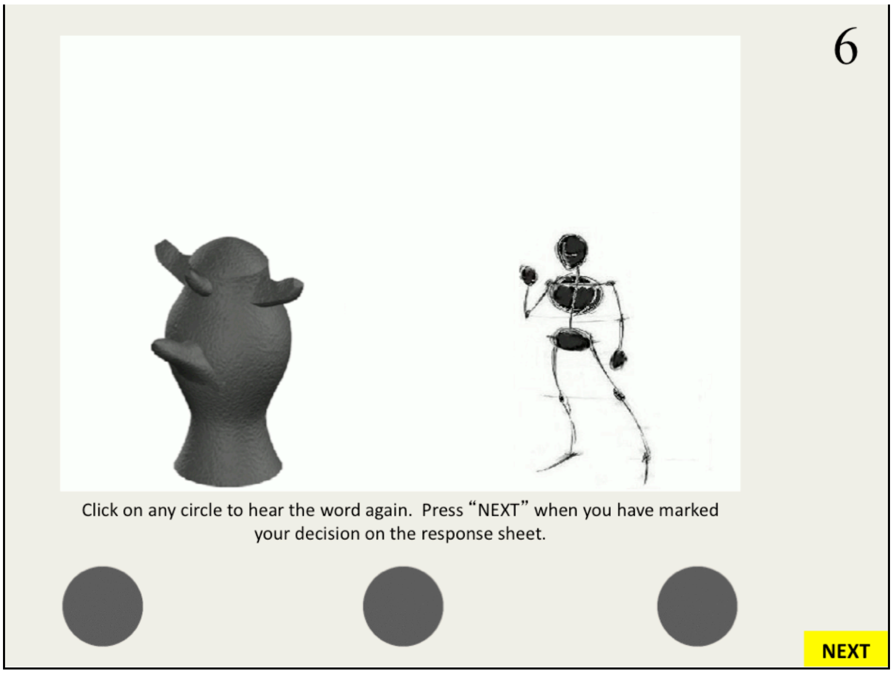


Figure 5. Experiment 2: Example of visual stimuli. Participants heard a prospective name each time a grey circle appeared.

Procedure

Each participant completed 42 trials at his or her own pace. PowerPoint slides began with a presentation of the greeble/human scene. Two seconds later, the first grey circle marker appeared, and the first possible name was presented auditorily via headphones. Two seconds after the onset of the first name and circle, the second name and circle were presented, and so on until all three circles and names were presented. Two seconds after the third set began, a box reminding the participant of the instructions and a button allowing the participant to move to the next trial appeared. Participants were encouraged to mouse click on each circle to hear the names again. Upon making a decision, participants had been instructed to clearly mark the grey circle on their response sheets that corresponded with their choices and press the button marked “NEXT” to move to the next item.

Results and Discussion

Each nonword name contained zero, two, or four “large” phonemes, and hence the mean number of “large” phonemes in the chosen names for each greeble served as the dependent measure. As illustrated in Figure 6, the size of the object being named linearly predicted the number of “large” phonemes in its preferred name. A repeated measures ANOVA showed this to be reliable, $F(1,15) = 11.779$, $p<.01$, partial $\eta^2 =.404$. All post-hoc comparisons of size were significant at the $p<.02$ level. Recall that some of the participants did not speak English as their native language. In a subsequent analysis we therefore included as an additional factor whether the participant was a native English speaker. This factor had no main effect on name choices, $F(1, 15) = .202$, $p > .65$, nor did it interact with greeble size, $F(2, 30) = .007$, $p > .99$. Thus, native and non-native English speakers exhibited similar naming preferences.

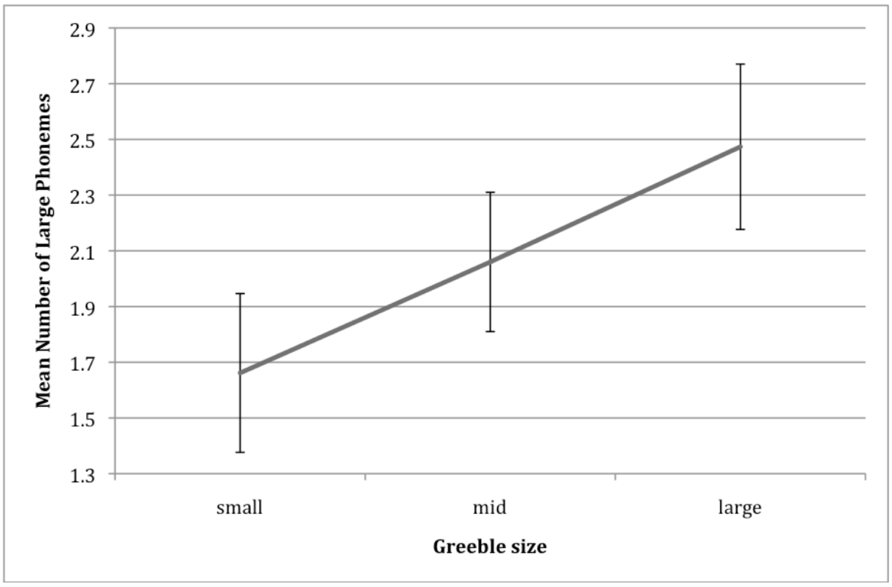


Figure 6. Experiment 2: Mean number of letters referring to "large" phonemes in the naming of different size of greebles. Error bars represent 95% Confidence Intervals.

General Discussion

The present study examined whether phonetic symbolism conveys the size of objects in a graded (i.e., with incrementally more large-sounding letters for increasingly larger objects) or dichotomous manner (i.e., marking only very large and very small objects). The results indicate that phonetic symbolism marks size in a graded manner: Participants reliably preferred names for novel objects that matched the size of each object, including an intermediate number of large-sounding letters for medium-sized objects. As the size of the greeble increased, so too did the number of large-sounding letters in its preferred name. In Experiment 1, participants reliably discriminated between five different levels of phonetic "size". For instance, participants preferred to name a 33% sized object with about three large-sounding letters (e.g., *kuloti*) and a 67% sized object with about four large-sounding letters (e.g., *tibudo*; see Fig. 3). Similarly, participants in Experiment 2 chose to name medium sized objects (M = 2.06) with names that were, in terms of number of large phonemes, almost exactly half way in between large (M = 2.47) and small (M = 1.66) objects. This is by far the most precise demonstration of sound symbolism to date.

The results are predicted by the theory that sound symbolism is the combination of gesture and the frequency code (Berlin, 2006) and originates in cross-modal processing (Maurer, et al., 2006; Ramachandran & Hubbard, 2001). As the size of the novel object increased incrementally, participants increased the proportion of letters that pointed to phonemes of lower frequency (e.g., back vowels, like /u/ and /o/) and less intensity (e.g., voiced stops, like /b/ and /d/; Berlin, 2006). This use of sounds is “naturally biased” (Maurer, et al., 2006, p. 320), such that the properties of the object (i.e., size) were matched to the phoneme in a meaningful way by participants. Put another way, the phonemes that participants used to mark for size were motivated directly by the properties of the object.

An alternate explanation to these findings could be that participants were naming based on speech sounds with this naming arising not from the properties of the objects themselves but rather from the conventions within their languages, i.e., statistical learning (Saffran, 2003). For example, a speaker of a language could become sensitive to an overrepresentation of front vowels in the words he knows for small things. When faced with a novel instance of a small item and asked to name it, he would be likely to name it following the conventions of his language (in this case, with front vowels). Given that there is variation between natural languages in terms of what sounds are attached to specific meanings, e.g., Diffloth, 1994, speakers of different languages would be exposed to different statistical trends and would name novel objects accordingly. In Experiment 2, over one third of participants (7 of 19) were not native speakers of English. If there had been a strong influence of the following of statistical trends of the participants’ native languages, we would have found there to be differences in responses of native and non-native English speaking participants. These differences were not found.

You will recall that presentation of nonwords in Experiment 1 was visual, i.e., graphemes. Although the results suggested a robust relationship between sound and object size, there was concern that the relationship could be affected by a difference between nonword groups of pixel width. Previous studies on presentation method have not been conclusive, so a second experiment was conducted using auditory presentation of nonwords. Despite Experiment 2 being simplified, the tasks in each experiment were still quite similar (the participant sees a novel object and chooses a name for it from those provided), and comparisons can be made between them. In both experiments, the interaction between size of object and number of large phonemes was highly significant, as were pairwise comparisons between each of the object sizes. The main difference appears to be in effect size- Experiment 1's partial $\eta^2 = .615$, Experiment 2's partial $\eta^2 = .404$. Given the correlation of pixel width and number of large phonemes in Experiment 1, participants could have used that as an additional cue as to what name to choose. This in conjunction with the sound symbolism that occurred when the participants read the possible names (via the internal voice) could have led to a stronger effect. Although it is likely that participants read the names silently to themselves in Experiment 1, it is unlikely that participants in Experiment 2 would have pictured the graphemes of the names they heard. Due to this lack of reinforcement, Experiment 2's effect size was somewhat smaller, though still fairly robust.

Chapter 3:

Phonetic Symbolism for Size and Shape

In most nonword and tonal studies of phonetic symbolism for size and shape, it appears that there is a large amount of overlap between the phonemes and tones used to mark both size and shape, e.g., /i/ has been shown to mark both 'small' and 'angular.' However, there are reasons to believe that these studies have not given a full picture of this aspect of this phenomenon. Commonly, previous research has used dichotomous stimuli with pairwise presentation (e.g, large v. small; Berlin, 2006). Though much of this was discussed in the last chapter, briefly, there is ambiguity as to what exactly is being marked. For example, in an experiment based on pairwise presentation, one could find that /b/ is being used intentionally to mark the large thing and /p/ is being used to intentionally mark the small thing. Alternatively, it could be that /b/ is being used to mark large things and /p/ is used to mark what is left due to forced choice, or vice versa. This may lead to a false view that both phonemes are being used for meaningful marking. As noted by French, forced-choice "causes the subject to presuppose the appropriateness of one choice or the other, when, in free association, neither might occur to him… and it forces the subject to think in terms of binary contrast" (French, 1977, p. 307).

Additionally, most studies examine marking for only one property. By only focusing on one property at a time, it may be that the other property is not being adequately controlled for. For example, many shape experiments are based on the classic Kohler (1947) takete/maluma stimuli, such as Westbury (2005) who used shaped frames to demonstrate a facilitation/interference effect on nonwords in a lexical decision task. The different shaped frames, however, also appear to be of different size, with 'spiky' frames creating a smaller amount of space for the presented word, leading to ambiguity as to whether the effect is due to the shape or the size of the frame.

This problem is not confined to shape experiments. Sapir (1929) gave participants two nonwords and asked participants to identify which of the two would be the name of a larger or smaller item. Imagine if the participant were asked to choose the name for the larger cat. Even if the imagined cats were both of the same length and height, it could be that the participant thinks of a skinny vs. overweight cat, making the actual decision bony vs. round. As such, it could be that there is conflation of size and shape that is adversely affecting the results.

There are, however, two examples of size and shape being examined together that would suggest that they would not be marked with the same phonemes. First, the semantic differential (Osgood & Suci, 1955) puts size (large-small) on the potency factor and shape (angular-rounded) on the activity factor. This would suggest that although the two are commonly related, the differences are psychologically real. Additionally, in a study by O'Boyle and Tarte (1980), participants modulated pure tones until they found the ones that fit the visual stimuli being presented. Though they found there to be a main effect for different shapes, they found no significant main effect for size.

The Current Research

These issues leave the door open to several questions. Are all levels of size and shape marked, or is it only certain levels (e.g., big)? Are size and shape marked differently, and do they interact in sound symbolism? We report three experiments that use 7-point ratings scales. First, this allows us to see if there is a graded nature to the use of phonemes in naming novel objects symbolically, as would be predicted by a speech-gesture/ Frequency code model. Second, it reduces demand characteristics by not implying to the participant that there is one right answer. Additionally, by asking the participants to rate the proposed names, it is somewhat similar to what could happen in real life, i.e., a friend saying to them, "I just found this thing. Let's call it 'budo.' What do you think?"

In Experiment 3, participants rate the appropriateness of the nonword name for greebles of varying size. Experiments 4 and 5 use the same paradigm, though the focus of Experiment 4 is shape and Experiment 5 involves the manipulation of both size and shape.

There are several likely scenarios that could emerge. First, we could find that O'Boyle and Tarte's (1980) results with pure tones extend to nonwords as well, i.e., only shape is reliably marked. Second, we could find that both size and shape are marked using the same phonemes, but they are marked in differing ways. For instance, the gestural component could be more germane to marking for shape, while Frequency Code (Ohala, 1994) may be what is being used when marking for size. Similarly, the patterns of their markings, i.e., how or if they are graded, could differ across the two domains. Given the overlap in the two domains, it may be that there is some overlap in phonemes as well. As such, we could find that both sets of properties are marked with the same phonemes in the exact same way.

General Method

In each of the following experiments, participants viewed novel objects (Greebles) of varying sizes and/or shapes. The greebles were presented against a white background and in conjunction with an abstract human figure. In each trial, participants heard a possible non-word name for the greeble, which consisted of phonemes that have been suggested to be symbolic of size and shape. Participants were asked to rate the appropriateness of the name for the greeble.

Participants

All participants reported to be native English speakers and all reported to be a part of the University of Warwick academic community. Participants were paid £6 and participated in only one experiment in this series.

As one of the criteria for participation was that the participant be a native English speaker, data from those who did not meet this requirement have been removed from analyses. Additionally, if the participant gave the same rating in more than two-thirds of his responses, the participant’s data were removed. This was an effort to only include participants who appeared to pay attention to the experiment. Several of those whose data were excluded made one large circle, straight down the page, indicating that they were not assessing each item individually. Even if phonetic symbolism is not at work in the names, some sort of varying preference for names is likely.

Materials

In each of the following experiments, the novel objects participants viewed in the assessment of possible names were symmetrical greebles that were semi-randomly selected from those provided courtesy of Michael J. Tarr, Brown Unviersity, http://www.tarrlab.org. Using the GNU Image Manipulation Program, each greeble was converted to greyscale, sized, and placed against a white background with a black-and-white abstract human figure, in a manner similar to that of Experiment 2. Further manipulation of the greebles was done for each of the following experiments and will be discussed in their sections.

The possible names were the same set of 42 non-words spoken by a man with a "standard" British accent that was used in Experiment 2. Each possible name followed the form of two consonant-vowel clusters (i.e., CVCV). Consonants in the possible names consisted of voiced and voiceless stops (/b/, /d/, /g/ and /p/, /t/, /k/, respectively). Vowels in the possible names consisted of back and front close vowel (/u/, /o/ and /i/, /e/, respectively). Voiced stops and back vowels have been shown to be related to largeness and roundness, while voiceless stops and front vowels have been shown to be related to smallness and pointiness (Berlin, 2006). CV clusters consisted of only one kind of phonemes, e.g., a voiced stop and a back vowel, such that four types of names were created- those with all phonemes related to largeness and roundness (e.g., "boo-doh"), those with all phonemes related to smallness and spikiness (e.g., “tee-pei”), and those with mixed content, either “small”-first (e.g., “tee-doh”) or “large”-first (e.g., “boo-pei”).

E-Prime was used as the experiments’ presentation software. Computers used for the experiments had 19” Sony CRT monitor with a resolution of 800x600 pixels. Auditory stimuli were presented via headphones.

Procedure

Participants were instructed that they would hear a possible name for a novel object and that their task is to rate how appropriate that name would be for the novel object. Each trial began with a written instruction to press the spacebar when ready. Upon the appropriate response, there is an interstimuli interval of 250ms, then 1000ms for the first presentation of the non-word name, both with a blank white screen. The abstract human and greeble scene is presented and after 1000ms, the possible name is repeated. After the 1000ms allotted for the second presentation of the possible name, a 7-point scale is included below the human and greeble scene. Upon an allowed response, a blank screen (ITI) appears for 500ms and the next trial begins.

Experiment 3

In Experiment 3, participants viewed greebles in three sizes and were asked to rate the appropriateness of the non-words they heard as names for the greebles.

Patricipants

Twenty-six members of the University of Warwick community were paid participants. All reported to be native speakers of English. No data were excluded in this experiment.

Materials

Forty-eight greebles were randomly selected from the greyscale greeble set. Just as in Experiment 2, greebles were manipulated into three sizes using the GNU Image Manipulation Program. Greebles in the medium size condition were approximately the same height as the human figure, though the greebles were slightly less gracile than the human figure. Greebles in the small condition were one-third the height of the human figure, while those in the large condition had a height that was an additional two-thirds to that of the human figure (see Figure 7 for examples). Left-right positioning of greeble and human figure was counterbalanced.

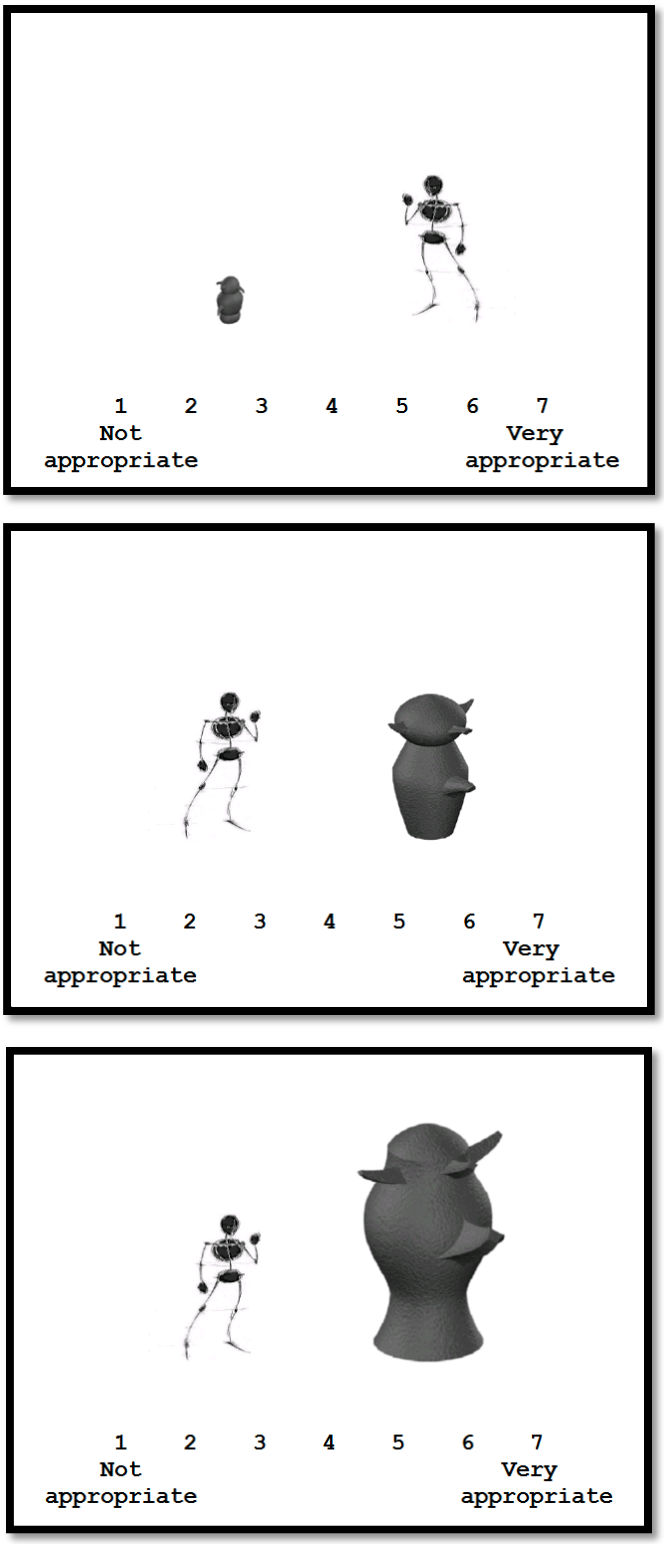


Figure 7. Experiment 3: Examples of small (top), medium (middle), and large (bottom) greeble conditions.

Procedure

Each of the 48 greebles was presented three times (once in each size) with the same possible name, creating 144 total trials.

Results and Discussion

A repeated measure ANOVA of just the two mixed audio conditions and greeble size showed there to not be an interaction, $F(2, 50) = .09$, $p>.05$. As such, the mixed phoneme conditions (Mixed, Large First and Mixed, Small First) were combined. A 3 (size) x 3 (name) repeated measures ANOVA showed there to be a significant interaction between possible name and size of greeble, $F(1.74,43.48) = 5.17$, $p<.01$, partial $\eta^2 = .17$. Additionally, there was a main effect of name, $F(2,50) = 17.49$, $p<.001$, partial $\eta^2 =.41$, but no main effect for size, $F(2,50) = .50$, $p>.05$. Post hoc analyses showed there to be no significant differences between the ratings of the three name conditions when the greeble was small, $F(1.47, 36.86) = 1.21$ p>05. However, there were significant differences occurring in a linear fashion for medium-sized greebles, $F(1,25)= 19.02$, $p<.001$, partial $\eta^2 = .43$, and for large greebles, $F(1,25)= 19.83$, $p<.001$, partial $\eta^2 =.44$

The interaction seems to have been driven by the participants showing a strong preference of name for large greebles, but not much of any preference for name with small greebles. The ratings for name occurred in the predicted fashion for large greebles, i.e., ‘large’ names rated the highest, mixed names rated the second highest, and ‘small’ names rated the lowest. This would seem to be in line with any theory (such as that of Frequency Code, Ohala, 1994) that would suggest that marking for size would be based on level of threat. That is, small things, all things being equal, tend to be less likely cause harm than large things. As a consequence, marking need be only for larger items.

Oddly, there was an effect of name and medium greeble size with the same order of preference as that found with large greebles. This would suggest that participants may not have viewed the medium greebles as being in an intermediate size, but rather that they were still large. This is not particularly surprising given that the medium greebles were of approximate height of the human figure, with the human figure being less rotund than the greebles. If the size of the medium greeble were judged as one would judge a person's size, the medium greeble would be fairly large. That being said, the effect was not quite as clean as that of the large greeble condition.

Additionally, it could be that smallness should be marked (as it was expected to be), but our choice of small size was not a strong enough manipulation. Although a smaller greeble might have been able to produce the expected results in terms of size symbolism, it would also have meant that the greeble would have been too small for judgments of other features to be made, i.e., shape.

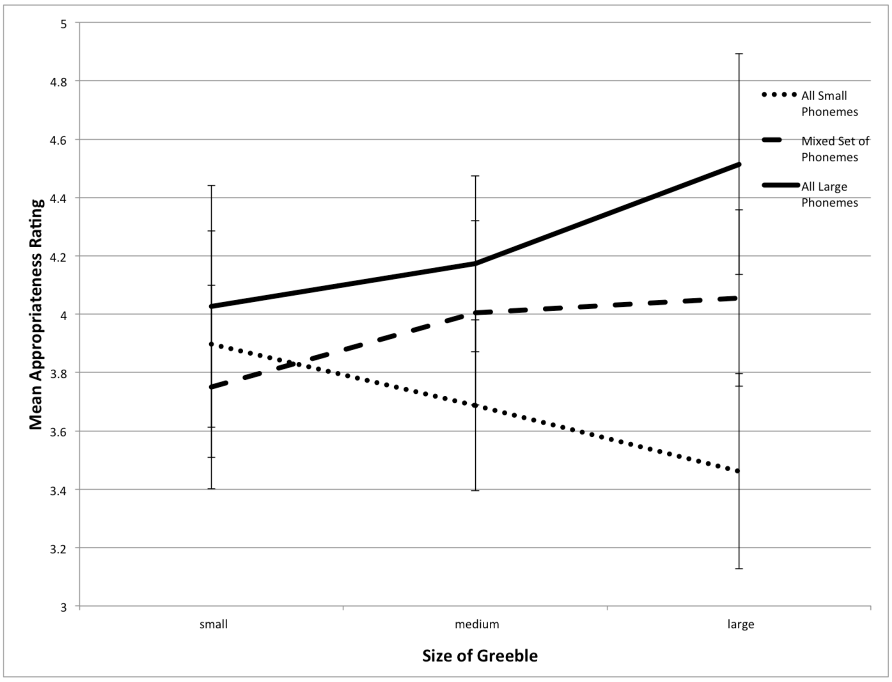


Figure 8. Experiment 3: Mean appropriateness ratings for possible names of greebles of varying size. Error bars represent 95% confidence intervals.

In the next experiment, phonetic naming for shape will be examined. The method is similar to the previous experiment on size symbolism, with the participants hearing a nonword name and rating how appropriate it is to name the greeble in the slide. In order to quantify and separate three sets of greebles based on shape, first, greebles were manipulated and a pre-test was performed.

Pre-Test

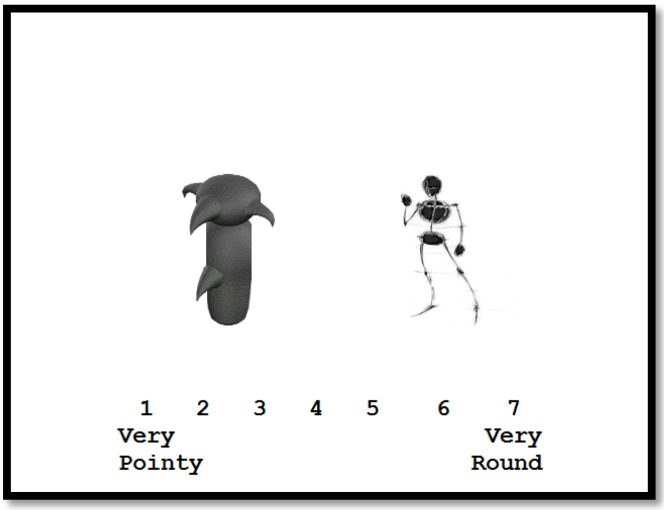
Eighty greebles were semi-randomly selected from the greyscale greeble set, all medium sized. Using the GNU Image Manipulation Program, twenty greebles were made more rounded, while twenty were made more pointed. Forty greebles were left intact, with the assumption that at least some of them would be rated to be of medium shape.

Participants

Nine members of the University of Warwick community acted as participants. Participants were either paid £6 or given partial course credit for their participation. As this is a simple, non-language-based task, no demographic information was collected.

Procedure

Participants viewed each greeble in a scene with the abstract human figure and rated, on a 7-point scale, the roundness/pointiness of the greeble (for examples, see Figure 9). The presentation of the greebles was randomized. With all participants, a rating of 1 suggested absolute pointiness, while a rating of 7 suggested total roundness.



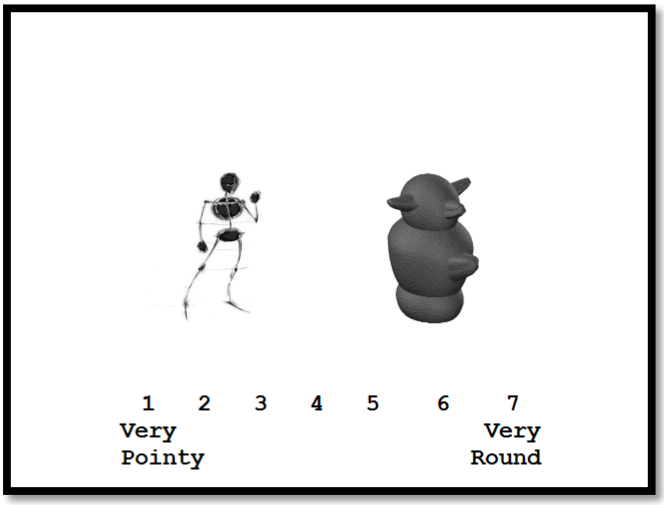


Figure 9. Experiment 4: Examples of the shape rating procedure. The greeble on top was rated as being very pointy, while the greeble at bottom was rated as being very round.

Results

Three shape groups of sixteen greebles each were created based on the participant ratings. The 16 greebles with the lowest ratings were assigned to the pointy condition (mean rating, 2.28), while the 16 greebles with the highest ratings were assigned to the round condition (mean rating, 4.89). The 16 greebles that were ranked from 33^{rd} to 48^{th} were chosen for the medium condition (mean rating, 3.61). A One-way ANOVA showed these groups to have significantly different ratings, $F(2,47) = 414.45$, $p<.001$. Additionally, each group was significantly different from the others. Using a post-hoc t-test, a significant difference was shown between the spiky and the medium groups, $t(30) = -16.55$, $p<.001$, with the spiky group having a significantly lower shape rating than the medium group. The spiky group also had a significantly lower set of ratings than did the round group, $t(30) = -25.14$, $p<.001$. The medium group was also significantly lower than the round group, $t(30) = -14.85$, $p<.001$.

Experiment 4

In Experiment 4, participants viewed greebles in three levels of roundness/spikiness and were asked to rate the appropriateness of the non-words they heard as names for the greebles.

Participants

Twenty-nine members of the University of Warwick community were paid participants in this experiment. All participants were required to be native speakers of English. As such, data from two participants who reported themselves to be non-native speakers were excluded. Additionally, data from two participants were excluded due to one rating being made for more than two-thirds of the responses. In both cases, the participants rated more than 75% of possible names a 1.

Procedure

From the shape ratings gathered in the stimuli norming, there were 16 greebles in each of three conditions: round, medium, and spiky. Each of the 48 possible names used in Experiment 3 was presented with a greeble in each shape group, making 144 total trials. That is, each of the 48 names was presented three times, each time with a greeble of a different shape group.

Results and Discussion

An analysis of the mixed name conditions showed there to be a significant interaction between name and shape, $F(2,48) = 6.02$, $p<.01$. As such, the mixed conditions were not combined. A 3 (shape) x 4 (name) repeated measures ANOVA showed there to be a significant interaction between name and greeble shape, $F(3.83, 91.84) = 10.44$, $p<.001$, partial $\eta^2 = .30$ (see Figure 10).

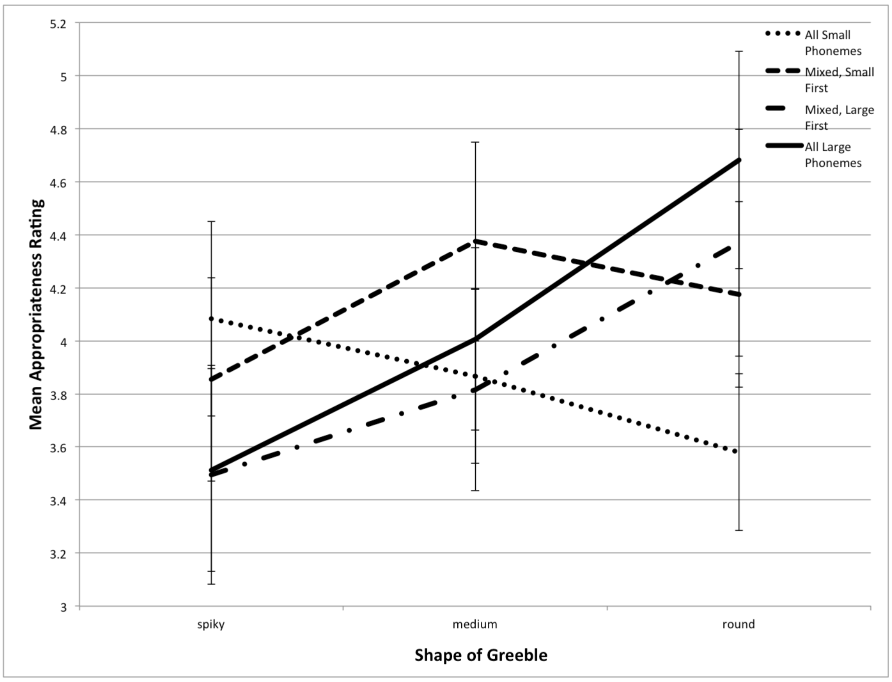


Figure 10. Experiment 4: Mean appropriateness ratings of names for different shaped greebles. Error bars represent 95% confidence intervals.

Additionally, there was a main effect of greeble shape, $F(1.45,34.73) = 9.16$, $p<.01$, with a trend towards higher ratings the rounder the greeble was. Post-hoc pairwise comparisons showed each level of shape differed from the other, with all differences significant at the $p<.05$ level. There was also a main effect of name, $F(2.23,53.41) = 4.27$, $p<.05$. Post-hoc pairwise comparisons showed there to be a significant difference in the ratings for small-first names (mean = 4.14) and the ratings for small names (mean = 3.84) and large-first names (mean = 3.89) at the $p<.05$ level. All large names were not significantly different from any other level. That is, all large phoneme names and large first phoneme names seem to be rated in similar ways for each greeble shape. Additionally, small first phoneme names appear to be most highly rated for medium shape conditions. These two findings are different from those found in the size section of this chapter, where mixed conditions did not differ from each other.

In terms of shape, One-way ANOVA's showed each to have a significant main effect of name. With the spiky shape main effect, $F(3,72) = 4.76$, $p<.01$, partial $\eta^2 = .16$, there were also significant pairwise differences between the small names (mean = 4.08) and the large-first names (mean = 3.50) and the large names (mean = 3.51). However, there was no significant difference between small names and small-first names (3.86). This may suggest that marking for spikiness may be done by the initial cluster.

With the medium shape ANOVA, main effect, $F(3,72) = 4.76$, $p<.01$, partial $\eta^2 = .17$, the small-first name (mean = 4.38) differed from the remaining conditions, small name (mean = 3.87), large-first name (mean =

3.82), and large name (mean = 4.01). There were no other differences between the groups. This shows a clear preference for the small-first name for greebles of medium shape.

Finally, there was a significant One-way ANOVA for name with round sized greebles, $F(1.97, 47.25) = 17.80$, $p<.001$, partial $\eta^2 = .43$. Pairwise comparisons showed there to be a significant difference between small names (mean = 3.58) and all other conditions, as well as between large names (mean = 4.68) and all other conditions. The mixed names, small first (mean = 4.18) and large first (mean = 4.37), did not differ from each other. In this case, there was a clear preference to name with large name, as well as a clear preference to not name with a small name.

In this experiment, we manipulated greeble shape and asked the participants to rate possible names for each greeble. Unlike with size, which had a fairly straight-forward set of findings, the data for naming based on shape appears to be a bit more complicated. Similar to Experiment 3, all large phoneme and all small phoneme names were marked in predicted ways, with preference towards rounded greebles and spiky greebles, respectively. However, mixed conditions in Experiment 4 were found to not be marked in the same ways. Specifically, it was found that large first names were marked in ways similar to all large names and small first names were marked with preference to medium-shaped greebles. This suggests that although there are some similarities between marking for size and shape, there are also some notable differences as well.

Experiment 5

In order to assess whether or not people mark objects for both shape and size, and to see if these two factors interact, in Experiment 5, both size and shape of greebles were manipulated.

Participants

Forty-two members of the University of Warwick community were paid participants in this experiment. As being a native speaker of English was a requirement for participation, data from one non-native English speaking participant were excluded. Additionally, data from one participant were presented by E-Prime in an anomalous way. As this suggested an error in the collection of that participant's data, the data have been excluded. This left a total of 40 participants whose data were included in the analysis.

Materials

Greebles that had been previously rated for shape were manipulated such that each would appear in three sizes when placed in the scene with the abstract human figure, creating a 3 (shape) X 3 (size) X 4 (name) design.

Procedure

The procedure matched that of the preceding experiments. Each participant completed 144 trials, with each possible name being repeated in 3 conditions. As the number of shape and size stimuli were not evenly divisible into the total number, multiple lists were made to accommodate these differences.

Results and Discussion

In order to examine the nature of the interactions between name, shape, and size, we first conducted an analysis of the mixed audio conditions. An analysis of mixed audio (Mixed, Large First and Mixed, Small First) conditions showed there to be no interaction with size and shape, $F(4,156) = .58$, $p>.05$. As such, these two conditions have been combined in the analyses. As list was not a significant factor, $F(5,34) = 1.96$, $p>.05$, lists were combined in further analysis.

Since mixed name conditions were not shown to be significantly different from each other and list was also shown to be a collapsible variable, these were combined such that an overall 3 (name) X 3 (shape) X 3 (size) repeated measures ANOVA could be performed. A 3 (name) x 3 (shape) x 3 (size) repeated measures ANOVA showed there to be a significant linear interaction between name, greeble shape, and greeble size, $F(1, 39) = 4.309$, $p<.05$, partial $\eta^2 = .099$. Please see Figures 11-13 for graphs of these interactions, with each of the graphs based on each of the three name conditions. Additionally, the interactions between name and size (as in Experiment 3), $F(1, 39) = 14.999$, $p<.01$, partial $\eta^2 = .278$, and name and shape (as in Experiment 4), $F(1, 39) = 13.139$, $p<.01$, partial $\eta^2 = .252$, were replicated, as can be seen in Figures 14 and 15, respectively. There was no significant interaction between size and shape, $F(1, 39) = .855$, $p>.05$.

Post hoc analyses showed there to be no significant interaction between size, and shape among small names, $F(4, 156) = .355$, $p>.05$, or between medium name, size, and shape, $F(4, 156) = .34$, $p>.05$. There was, however, a significant linear interaction between size and shape with large names, $F(1, 39) = 6.35$, $p<.02$, partial $\eta^2 = .14$. Additionally, there was a main effect of shape, $F(1.58,59.56) = 12.19$, $p<.00$, partial $\eta^2 = .24$. Pairwise comparisons within the large name data showed responses to round greebles to be significantly different than those of other shapes (both at the $p<.01$ level). There was also a main effect of size, $F(1.40,54.43) = 4.78$, $p<.05$. Responses to large greebles with large proposed names also differed from those of other sized greebles (large v. small, $p<.02$, large v. medium, $p=.055$). As such, in this experiment, it appears that size may have been the more potent variable as compared to shape, and the main interaction appears to be due to the large, round greeble.

In summary, we expected to replicate findings from the previous experiments in this chapter, where we found general trends of marking large greebles and rounded greebles with large phonemes and small and spiky greebles with small phonemes. We also expected there to be some interaction between size and shape. For instance, we predicted that we would find that a greeble that was both large and round to be more preferentially matched with large names than greebles that had just one of those properties. We generally found these effects, however, we did not find evidence that small size and shape interacted. This may have been due, though, to small greebles being too small to easily detect their shape.

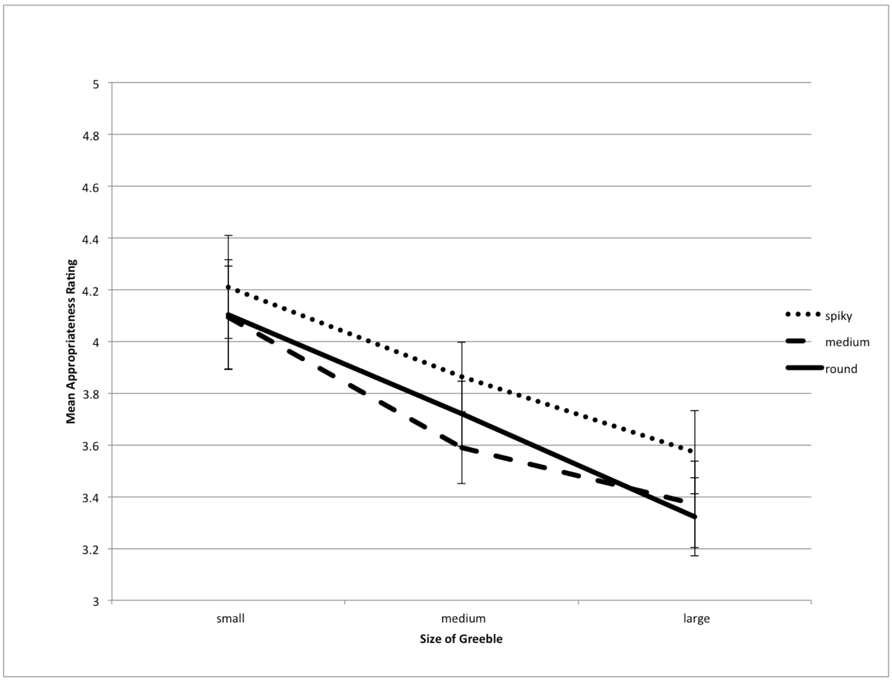


Figure 11. Experiment 5: Mean ratings for small names for greebles of varying size and shape. Error bars represent 95% confidence intervals.

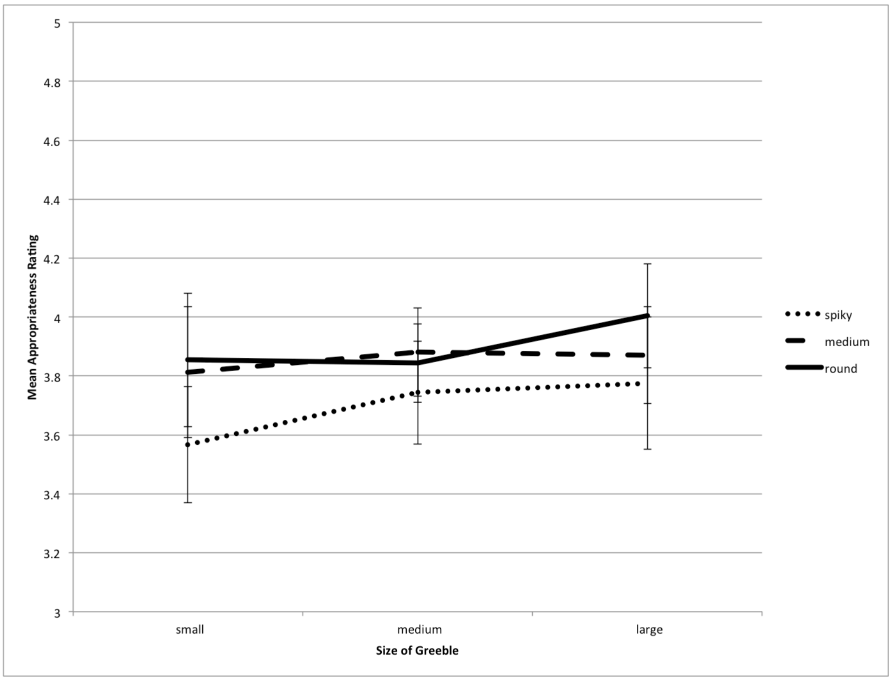


Figure 12. Experiment 5: Mean ratings for medium names for greebles of varying size and shape. Error bars represent 95% confidence intervals.

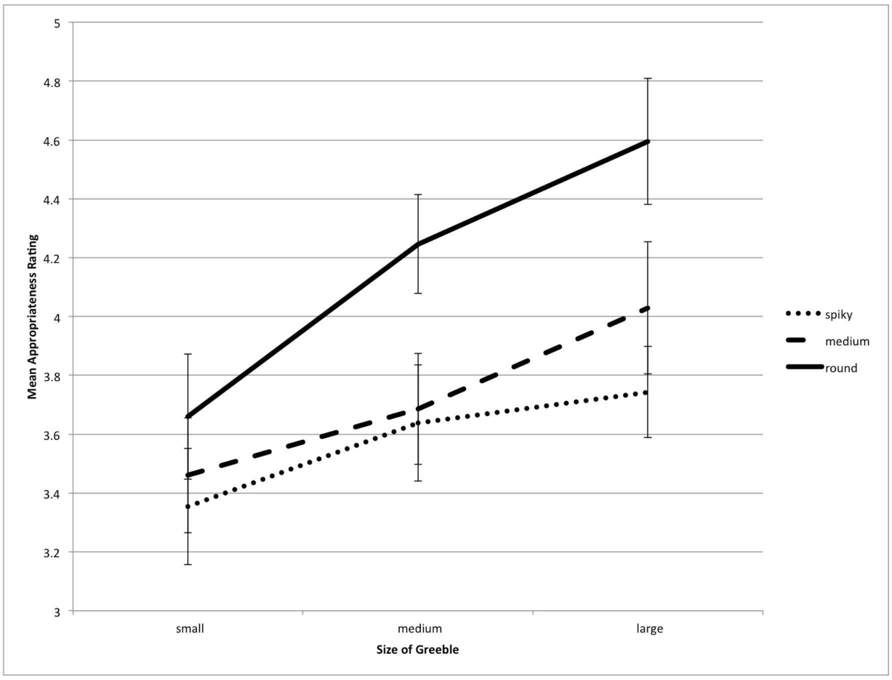


Figure 13. Experiment 5: Mean ratings for large names for greebles of varying size and shape. Error bars represent 95% confidence intervals.

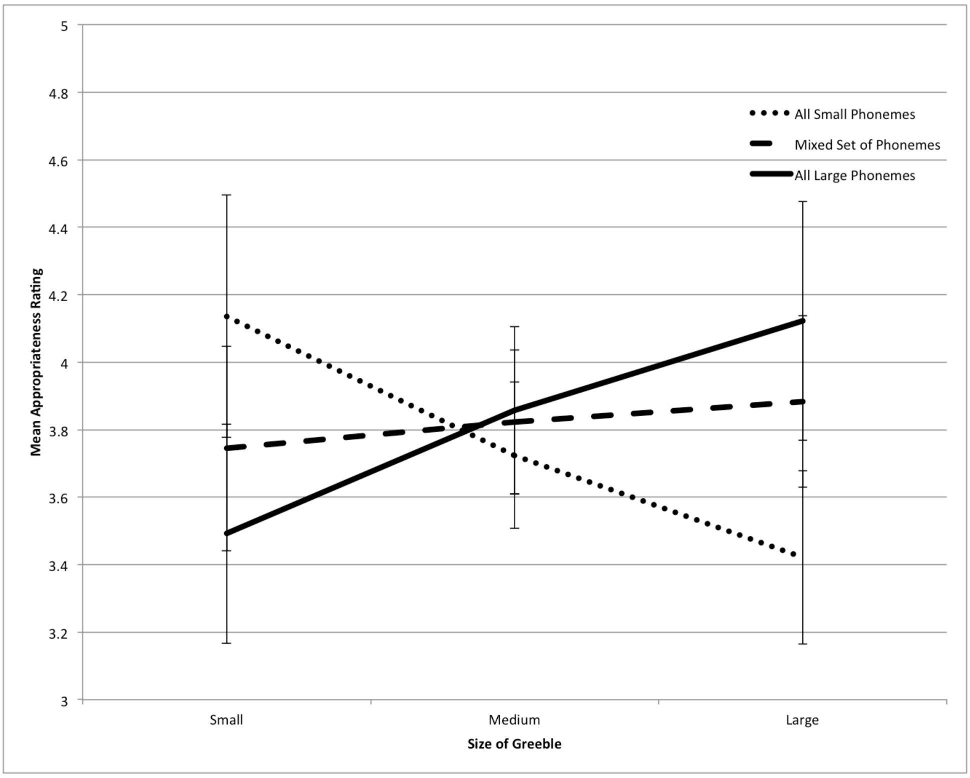


Figure 14. Experiment 5: Mean ratings for greebles of varying size, replicating findings from Experiment 3. Error bars represent 95% confidence intervals.

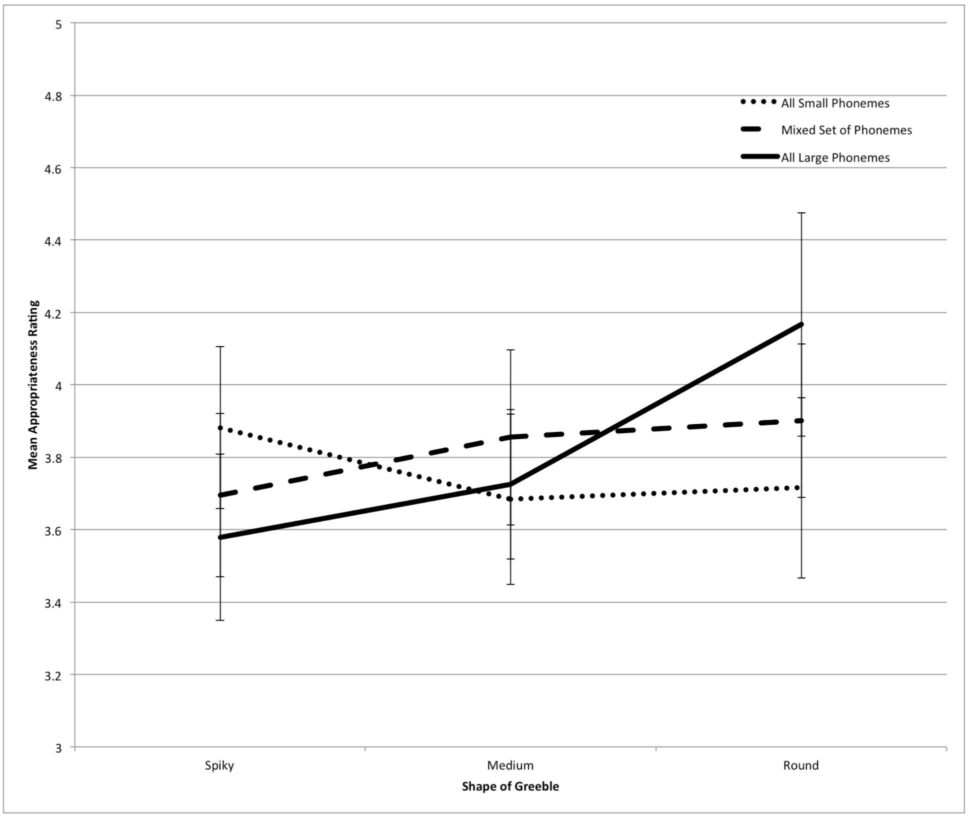


Figure 15. Experiment 5: Mean ratings for names for greebles of varying shape, replicating findings from Experiment 4. Error bars represent 95% confidence intervals.

General Discussion

In this chapter, we presented a series of three experiments to determine the nature of phonetically symbolic marking for size and shape. Previous studies had tended to use forced choice experiments and pairwise presentation, leading to possible demand characteristics. Additionally, despite claims to the contrary, there was little reason to believe that both size and shape are marked in the same way, given that most experiments involved only one of the two sets of properties.

With size, only larger objects appear to be marked, with both medium and large objects displaying significant linear trends. As the medium sized greeble is a little larger than the human figure, it is not surprising that people would possibly view it as still being large. It may be that when facing pairwise stimuli, participants are not marking small items but are using remaining phonemes after marking the large items (i.e., if a name isn't good for large items, it must be ok for items of other sizes). The findings here suggest that Frequency Code is being used in naming based on size, with the need for only large, and, by proxy, potentially dangerous, items to be marked. As noted earlier, however, it could also be the case that the small greeble manipulation was not strong enough to elicit the expected response.

In Experiment 4, the results were not as clear-cut. There was a significant overall interaction between name and shape, as well as significant one-way ANOVA's for name in terms of each shape. With spiky greebles, the marking preference seemed to be based on onset cluster, with both small and small-first being preferred over the other conditions, as well as a preference to not name with large and large first names. With medium greebles, there was a clear preference to name with small first names. With large greebles, the preference was both for naming with large names and for not naming with small names. That spiky and round had clear preferences both to name and to not name seems to be supporting the theory that shape of mouth when forming the names is a primary mechanism in sound symbolism (i.e., vocal gesture; Berlin, 2006).

Additionally, there was an interaction of size, shape, and name. This appears to have been driven primarily by two effects. First, in the condition where the greeble name was large, there was a general trend to increase in rating as the greeble size increased. However, ratings for the round greeble seemed to increase even more so. Secondly, there was the general tendency for small names to be rated more highly the smaller the greeble was. This appears to be a general trend, though spiky shaped greebles elicited the highest ratings throughout.

Of note is the question of whether there was a difference in potency of the size and shape manipulations. First, there seems to be support for the size condition being more salient than the shape condition, which is not surprising given that the differences between the sizes was much more pronounced than the differences between the shapes. Second, the medium size condition may have been too large and/or the small condition may have been too large, leading to effectively give us a set that included not very small, large, and very large greebles. This may have affected the results, and inadvertently created a quasi-dichotomous stimuli set. In addition, it may have been that the small greebles were too small for the participants to easily judge their shape. This could have lead them to have only produced ratings that suggested that large size and shape had an interaction.

There is, however, an alternate conclusion that could be drawn from the data. As the greebles have features that would be similar to an animal's and they were placed in a context with an abstract human, it could likely be that the participants viewed the greebles as abstract representations of some unknown animal, or at the very least, that they were natural kinds. As such, the participants may have inferred that shape is likely to be a permanent feature, e.g., small, pointy crabs grow up to be big, pointy crabs. It could, then, be the case that only largeness was marked with size due to the participants viewing the greebles of other sizes as still growing, with large in this case referring to anything bigger than the average human. Though there is evidence that there is marking for the diminutive, e.g., Newman (1933) and Ultan (1978), it could be that we do not have the tendency to assign meaningful symbolism to properties that are unlikely to remain stable over time. For example, the diminutive nickname "kiddo" only applies to those who are young, or at least younger than the one using the term. It also would not likely be used in the permanent naming of someone. Stability of trait in phonetic marking will play an integral part in the next chapter.

This set of experiments demonstrated that, similar to previous experimental works, e.g., Berlin (2006), size is marked via the use of vowel backness and voicing in stops in a fairly straight forward way. However, the data from Experiment 4 suggest that when size is controlled for, and not manipulated, marking for shape does not follow the same predictable pattern as marking for size does. Specifically, it appears that onset cluster plays a pivotal role in marking for size, with mixed audio conditions being used to mark shape in disparate patterns. This is in contrast to the findings of previous works, e.g., Maurer, et al. (2006) and Westbury (2005). When size and shape are both manipulated, however, we find that shape is marked in the predicted manner. This suggests that previous works may have not effectively controlled size in shape-marking experiments, leading to the simplified marking patterns we see in Experiment 5.

It appears, then, that size and shape are both marked with vowel backness and voicing of stops, though they are marked in different patterns when manipulated alone. As noted earlier, when manipulated together, we found the main effects of size and shape to have similar (and predicted) patterns. We also expected there to be an interaction of size and shape with naming. This prediction was supported by the data, but it only appears that large/round greeble naming is driving the effect. Though this may be due to only large/round being marked, it is more likely that both small/spiky and large/round are being marked, but that it was not seen in the data due to the small greebles being too small for shape to easily be detected. Despite this possible flaw, Experiment 5 does demonstrate an interaction of size and shape in naming that is novel in the current literature.

Chapter 4:

Phonetic Symbolism for Motion of Novel Objects

Activity is one of the big three factors in the Semantic Differential scales (Osgood & Suci, 1955). It also is a consideration when marking with phonetic symbolism. Birch and Erickson (1958) found a general trend towards rating nonwords with front vowels as more active (faster) than those with back vowels. Additionally, they found there to be some effect of presentation sequence (i.e., what order the dimensions were presented to the participant; e.g., Evaluative-Activity-Potency, Potency-Evaluative- Activity, etc.) on participant response, leading to a lack of stability in their responses across trials. However, "the correlations for the Activity dimension, however, tend to indicate a stability independent of responding to other dimensions" (1958, p. 293), meaning that ratings for activity appear to be more robust and consistent regarding order effects. Slobin (1968) demonstrated that word pairs relating to activity were dramatically more successfully matched (that is, matched in the way that was hypothesized) than those related to affect and potency, with a 69% success rate.

Participants in a study by Holland and Wertheimer (1964) rated the 'takete' object as more active than the 'baluma' object. Both Berlin (2006), using still life drawings of birds and fish, and Maurer, et al. (2006), using abstract pictures, point to motion as being a consideration in phonetically-symbolic naming. In all three studies, this idea of movement exists despite the use of static stimuli, i.e., motion was only implied.

Studies that have involved phonetic symbolism and objects that were moving have focused on the relationship between pitch and moving up or down. Walker, et al. (2010) found that infants attended longer to a video of a ball moving up and down when the sound (made by a slide whistle) and the direction of the ball were congruous. That is, participants attended longer to a ball moving upwards with a sound that was also increasing in pitch and vice versa. These findings are in line with studies showing that pitch and visual height congruity facilitate responding (e.g., Marks, 1987).

However, visual height and activity are not nearly the same thing. What, then, are participants inferring as activity when they are rating static images and nonwords? The most likely candidate is speed, either in terms of the possible maximum speed for the animals in the Berlin (2006) study or the speed in which the participant imagines the pen was at when making the abstract images in the Maurer, et al. (2006) study. Additionally, a study by Shintel and Nusbaum (2007) demonstrated that faster speech rate facilitated the recognition of an object when it was pictured to be in motion.

Still, speed is not the only possible factor when considering motion. Additional adjective pairs in the Osgood (1955) factor involve agitation-calmness and ferociousness-peacefulness. This may be similar to the stimuli found in Imai, et al. (2008), where speed of walking is manipulated in addition to the nature of the steps being made by the actor, e.g., “light, playful steps” vs. “very heavy steps” (2008, p. 57). It would appear that activity in terms of actual motion could possibly involve speed, continuity of motion, and the pattern of how that motion is carried out.

The Current Research

Here, we report four experiments to examine whether or not motion is actually used as a factor in phonetically symbolic naming of novel objects, and if so, what about motion (speed, continuity, pattern) is being marked. We use greeble stimuli from previous experiments and manipulate their motion across a white background. Participants hear a possible name for the greeble and rate its appropriateness. In Experiment 6, greebles move in three speeds, and begin in random positions and move back and forth. In Experiment 7, continuity and pattern are added to speed to make two gross movement conditions, fast/intermittent/spiky pattern vs. slow/continuous/rounded pattern. The last two experiments in this series use those same stimuli as those in the Experiment 7. In Experiment 8, however, the participants are explicitly instructed that the greebles are moving on their own, i.e., they are self-locomoting. Their task remained the same- that is, to assess the fit of the name for the greeble. In Experiment 9, though, the participants were tasked with assessing the name they heard to the motion being made, not to the greeble making the motion.

General Method

In each of the following experiments, participants viewed novel objects (Greebles) floating across a white background. In each trial, participants heard a possible non-word name either for the greeble (Experiments 6-8) or the motion itself (Experiment 9). The participants' task was to rate the appropriateness of the name and indicate their rating on the provided response sheet. Each experiment took 15-20 minutes to complete.

Participants

All participants reported to be a part of the University of Warwick community. No participant had been a part of previous experiments, i.e., each participant only participated in one experiment. Participants were either paid £6 or received course credit for their participation.

In order to remain in keeping with the experiments in the previous chapter, data from participants who indicated two ratings for one item or skipped an item were excluded from analyses. As in the previous chapter, data from participants who responded to more than two-thirds of the items with the same rating were excluded. One participant's data were excluded due to a fire alarm drill during testing and one participant's data were excluded due to the participant reporting poor vision.

As some of the participants were a part of the Warwick research experience pool, the criterion of being a native English speaker was not included in their recruitment. As such, data from non-native English speakers participating in these experiments are included. As scores from native and non-native speakers were not significantly different in Experiment 2, it is unlikely that this will cause a confound. Whether or not one is a native speaker is included in the demographic information we collected and will be included in the analyses.

Materials

Novel object stimuli in these experiments were the same as the 48 medium-sized, greyscale greebles in Experiment 3. The possible names were the same set of 48 spoken non-words in the previous chapter.

Presentation of auditory stimuli was done via headphones, while visual stimuli were presented using Microsoft Powerpoint, running on PC's with 19" Sony CRT monitors with resolutions of 800x600 pixels and refresh rates of 100 Hz. As PowerPoint is not meant to randomly display slides, multiple presentation orders were created for each experiment using the random function of Random.org. In each analysis, these lists were tested to ensure that there were not any interactions between order of presentation and any other factor before combining data across all lists. Participants indicated their ratings by marking a paper version of the appropriateness scale used in the previous chapter (which continued to employ a 7-point scale that ranged from "not appropriate" to "very appropriate.")

Procedure

Each experiment was auto-running, that is, once the researcher began the experiment, it ran without any further input. The participant read an introduction slide that explained the instructions. When the participant felt he was clear on the instructions, he was to ask the researcher to begin the experiment. Each slide of the Powerpoint-presented experiments was on a timer, with each trial lasting approximately 11 seconds. Each slide indicated the trial number, which corresponded to the appropriate response line on the participants' response sheets.

The first stage of each trial was the greeble in its beginning position against a white background. This stage lasted just over one second, which included the program loading the audio tracks into memory, and due to the lack of motion, the stage also acted somewhat as an inter-trail interval.

Crucially, as all the audio tracks were of very similar sizes, this stage should have a very similar duration from trial to trial. Beginning positions included the four corners of the screen as well as the far left and far right of the middle (vertically-speaking) of the screen. These starting positions were counterbalanced. Since the presentation software was PowerPoint (which does not allow for true randomization), six random lists were created and participants were randomly assigned to each list.

After the spooling-up initial stage, the greeble began its movement, which continued throughout the remaining duration of the trial. Upon the onset of movement, the possible name was auditorily presented for the first time. After 5 seconds, the name was repeated. After a total of 10 seconds of movement, the trial ended and the beginning slide for the next trail appeared. There was a pre-announced one-minute break at the mid-way point of each experiment.

Experiment 6

In Experiment 6, participants viewed greebles in three speed conditions. They were asked to rate how appropriate the non-word name was for the greeble.

Participants

Data from 31 participants were included in the analyses. Each participant self-reported to be a native speaker of English and a part of the University of Warwick community.

Additionally, five participants marked two responses for a single item and five participants skipped an item. This may have been due to a lack of sufficient attention. No matter the case, since it could have thrown off all the answers that followed, it was deemed proper to exclude the data from these participants in their entirety. Additionally, two participants responded with the same response over two-thirds of the time, and one participant was being tested when a fire alarm went off in the building. Data from each of these participants have been removed from analysis, leaving 31 participants in the analysis.

Materials

Auditory stimuli for this experiment were the same 42 spoken non-words as those used in the previous chapter. The novel objects were the medium-sized, greyscale greebles from Experiment 3. Each greeble was randomly assigned a starting position and a non-word name.

Procedure

After the researcher began the experiment, a white background with a greeble in its starting position appeared. At each corner of the slide was the trial number, corresponding to the appropriate line on the participant's response sheet. There were six starting positions for greebles (the four corners of the slide and the far left and far right of the middle of the slide.) In this experiment, each greeble traveled a straight path back and forth from its starting position to the opposite position, e.g., from the top left corner to the bottom right corner (for examples, please see Figure 16).

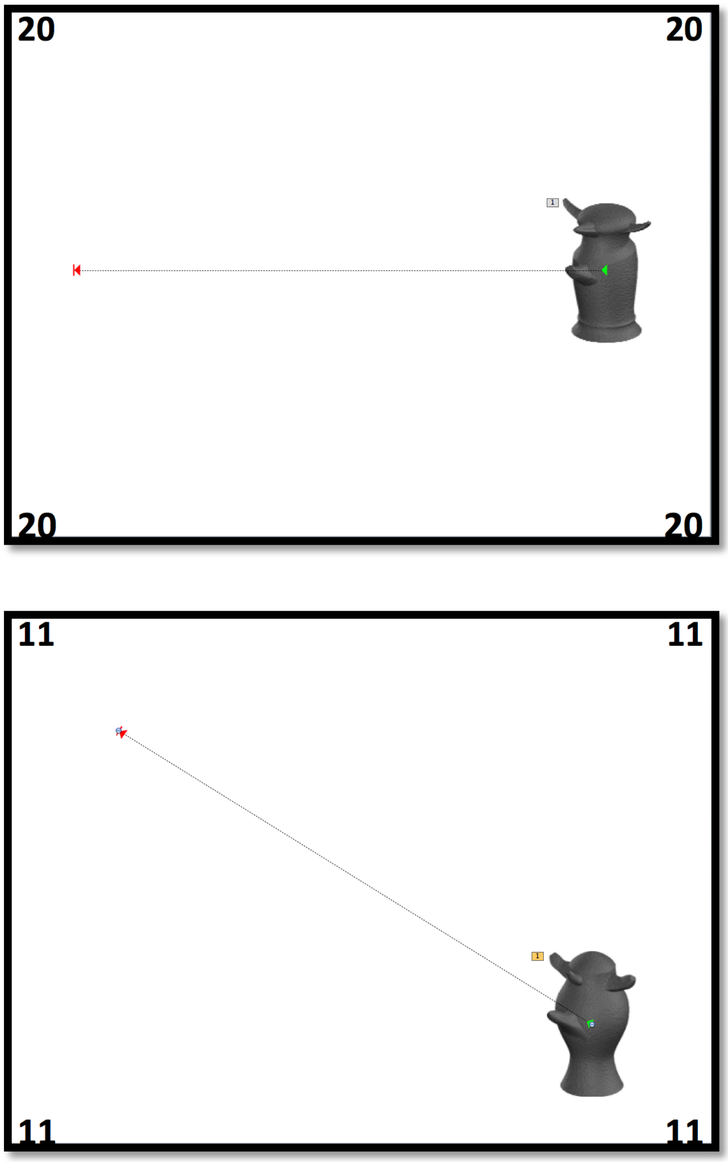


Figure 16. Experiment 6: Two examples of greebles in their starting positions with their back and forth paths indicated.

Each greeble/name combination was presented a total of three times- once for each of the three speed conditions, creating 144 trials. The three speeds were based on the amount of time it took the greeble to go from one point to the other (e.g., from a starting point at the top left corner to the bottom right corner). Greebles moving at the medium speed took 2 seconds for each leg of their path. Those in the slow condition took 3 seconds and those in the fast took 1 second. As the distance from side to side and up to down was less than any of the diagonal paths, this meant that the diagonal path trials moved more quickly. However, the difference in distance was negligible and the diagonal paths still had three clear speeds at which the greebles travelled.

Results and Discussion

A repeated measures ANOVA was performed to see if there were any significant interactions between mixed names and speed. There were not, $F(2,60) = .55$, $p>.05$, and the mixed audio conditions were combined in further analyses. A 3 (name) x 3 (speed) repeated measures ANOVA showed there to be no significant interaction between speed and name of greeble, $F(4,120) = .86$, $p>.05$. (see Figure 17).

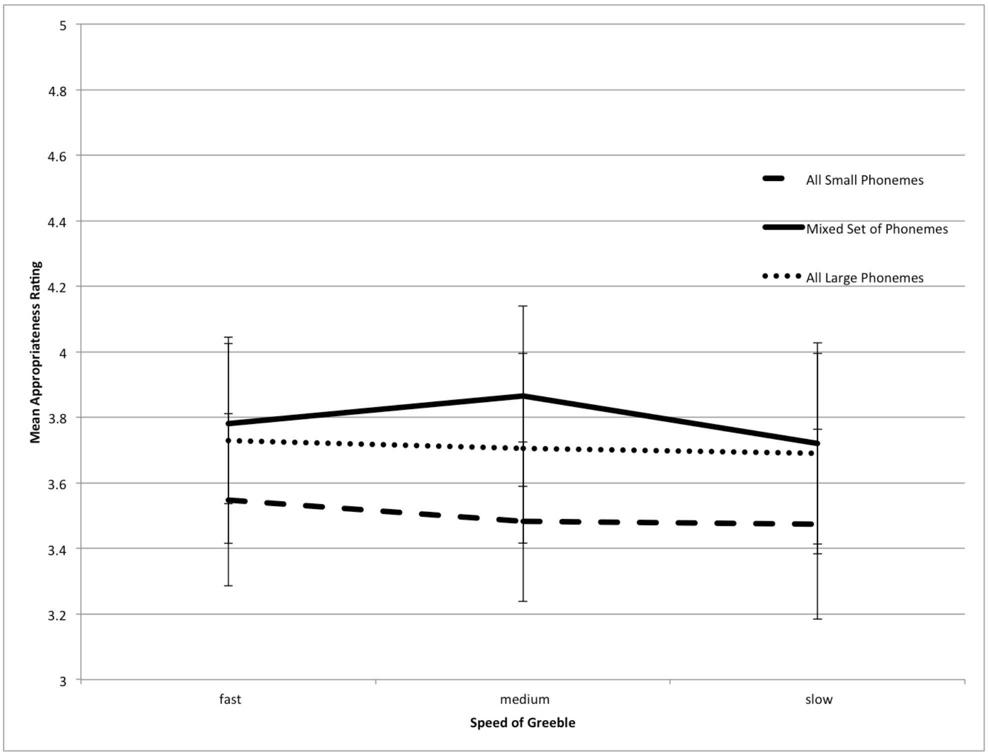


Figure 17. Experiment 6: Mean appropriateness rating for name of greebles moving at different speeds. Error bars signify 95% confidence intervals.

There was a main effect of name, $F(1.66, 49.76) = 4.75$, $p<.05$, partial $\eta^2 = .14$. Additional analyses showed there to be a significant difference between ratings for small names v. mixed names, $F(1,30) = 15.74$, $p<.001$, with small names being rated on average lower than mixed names. However, there was no significant difference between small and large names, $F(1,30) = 3.48$, $p>.05$, or between mixed and large names, $F(1,30) = .631$, $p>.05$. There was no main effect for speed, $F(1.42,42.70) = .56$, $p>.05$. Though there was a preference for mixed names over small names, there was little else going on in the data, suggesting that speed alone is not necessarily marked phonetically in a greeble's name.

Experiment 7

Speed is only one component of movement. Therefore, a null effect in terms of speed does not imply a lack of sound symbolic marking for motion in general. In Experiment 7, the net was widened and each movement had three components- speed, curviness of path, and continuity. Participants were still tasked with assessing the names they heard in terms of fit for the greeble making the motion.

Participants

Data from 35 participants were included in the analyses. Of these participants, 24 reported to be native speakers of English and all reported to be a part of the University of Warwick community.

Additionally, data from two participants were excluded from analysis due to responding to a single item with multiple answers. Data from one participant who gave the same rating for more than two-thirds of the items were also excluded.

Materials

This experiment used the same 48 greeble/name pairs as the previous experiment. Microsoft PowerPoint was used to present the stimuli. Participants wore headphones and marked their responses on the appropriate seven-point scale on a response sheet.

Procedure

There were two movement conditions, based on the conjunction of three motion characteristics. In the first condition, the greeble's motion was fast and followed a jagged pattern. At the end of each leg of its path, the greeble paused for 500ms. In the other condition, the greeble's motion was slow, following a curved pattern that included no stopping. For examples of these stimuli, please see Figure 18.

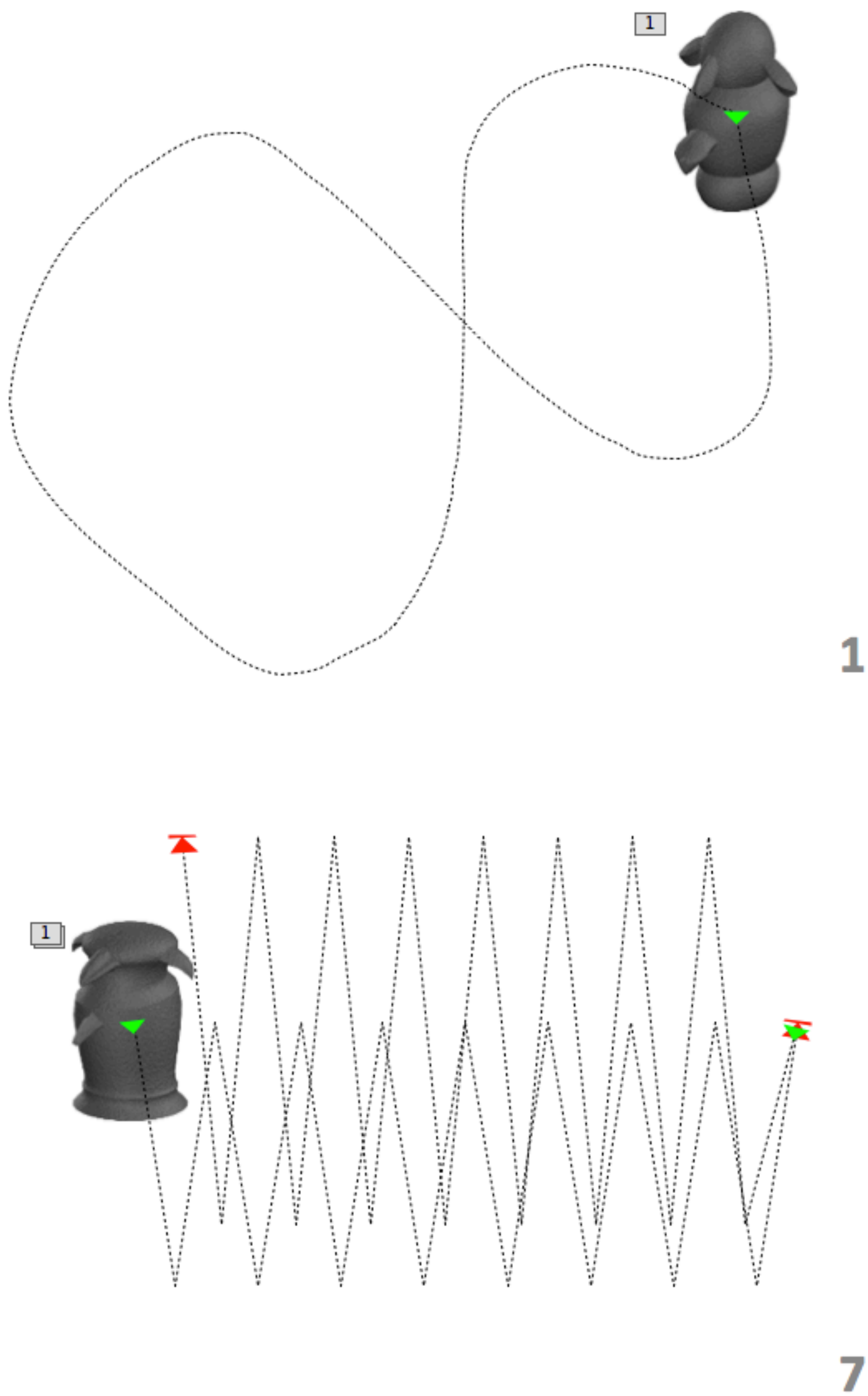


Figure 18. Experiment 7: Stimuli examples of slow/rounded/continuous (top) and fast/spiky/intermittent (bottom) movements.

Each of the greeble/name pairs was presented twice- once in each of the motion conditions, creating 96 trials. Each trial followed the same timing and order as in the previous experiment. Again, the participants were tasked with rating the appropriateness of the name for the greeble.

Results and Discussion

In order to see if the mixed audio conditions differed from each other, a repeated measures ANOVA was performed. There was no interaction between name and motion, $F(1,34) = .001$, $p>.05$. As such, mixed audio conditions were combined. Additionally, an analysis was performed to see if there was any interaction between native v. non-native English and name and motion. There was no interaction between native-speaking, name, and motion, $F(2,62) = .27$, $p>.05$, and the conditions were combined.

A 3 (name) x 2 (motion) repeated measures ANOVA showed there to be no significant interaction between motion and name, $F(1.59,54.00) = .43$, $p>.05$ (see Figure 19). There was no main effect of name, $F(1.61, 54.87) = .242$, $p>.05$. There was also no main effect of motion, $F(1,34) = 3.02$, $p>.05$. As such, it appears that participants were not marking greeble names based on the movements that they were in.

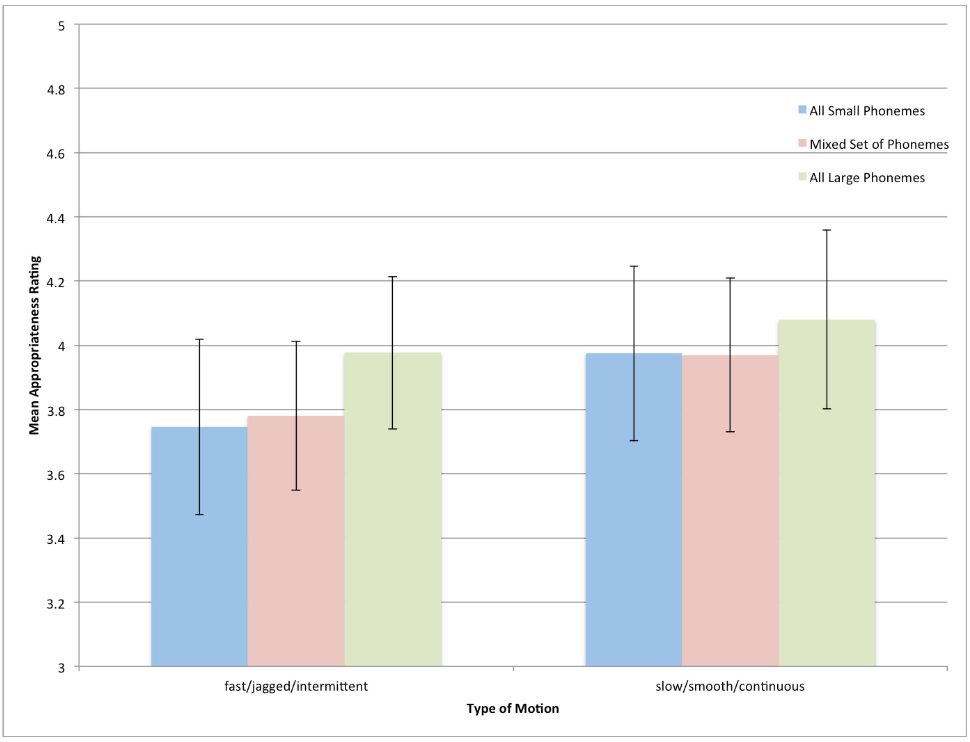


Figure 19. Experiment 7: Mean appropriateness of name for greebles with different motions. Error bars represent 95% confidence intervals.

One possible conclusion from these data is that the participants do not believe the object is moving itself, i.e., self-locomoting, but, rather, is being moved about by an external force. It has been shown in visual attention tasks that animate motion elicited more attention and facilitated faster responses (Pratt, Radulescu, Guo, & Abrams, 2010). Additionally, one of the adjective pairs strongly associated with activity is active-passive (Osgood & Suci, 1955). As such, the next experiment includes explicit instructions that the object is moving itself, with the prediction that participants will see the motion as a property of the greeble (and not that the greeble was being manipulated by some outside force) and, therefore, use the property of motion in the naming of the greeble.

Experiment 8

In Experiment 8, participants again viewed greebles in two gross motion conditions with the task of rating the appropriateness of the accompanying name for the greeble. Unlike the previous experiment, however, participants were explicitly instructed that the greebles were self-locomoting.

Participants

Data from 33 participants were included for analysis. Each participant reported to be a native speaker of English and a member of the University of Warwick community.

Additionally, data from three participants were excluded due to responding to the same item twice. One participant's data were excluded due to skipping an item. One participant was not a native speaker of English, so these data were excluded. Finally, two participants provided the same response to more than two-thirds of the items and their data were excluded.

Materials

All materials in this experiment were identical to those used in the previous experiment. Participants viewed and heard 48 greeble/name pairs, with greeble motion continuing to be in two gross categories (fast/jagged/intermittent vs. slow/smooth/continuous).

Procedure

The procedure matched that of the previous experiment except that the participants were explicitly instructed that the greebles were moving themselves, specifically, that the participants' task is to rate the names they hear for the novel objects as they move themselves around. Again, there were 96 trials and the participants were to rate the appropriateness of the name for the greeble.

Results and Discussion

An analysis was performed to see if the mixed audio conditions interacted with name and motion. They did not, $F(1,32) = .17$, $p>.05$, and were combined for further analyses. A 3 (name) x 2 (motion) repeated measures ANOVA showed there to be no significant interaction between greeble motion and name, $F(1.48,47.36) = .92$, $p>.05$ (see Figure 20). There was no significant main effect of motion, $F(1,32) = 1.11$, $p>.05$, but there was a significant main effect of name, $F(2,64) = 3.92$, $p<.05$, partial $\eta^2 = .11$. Further analyses showed there to be a significant difference between ratings of small and mixed names, $F(1,32) = 5.55$, $p<.05$, partial $\eta^2 = .15$, and between small and large names, $F(1,32) = 5.03$, $p<.05$, partial $\eta^2 = .14$, in both cases, with small names being rated significantly lower than the other name. However, mixed and large names did not differ significantly, $F(1,32) = .31$, $p>.05$.

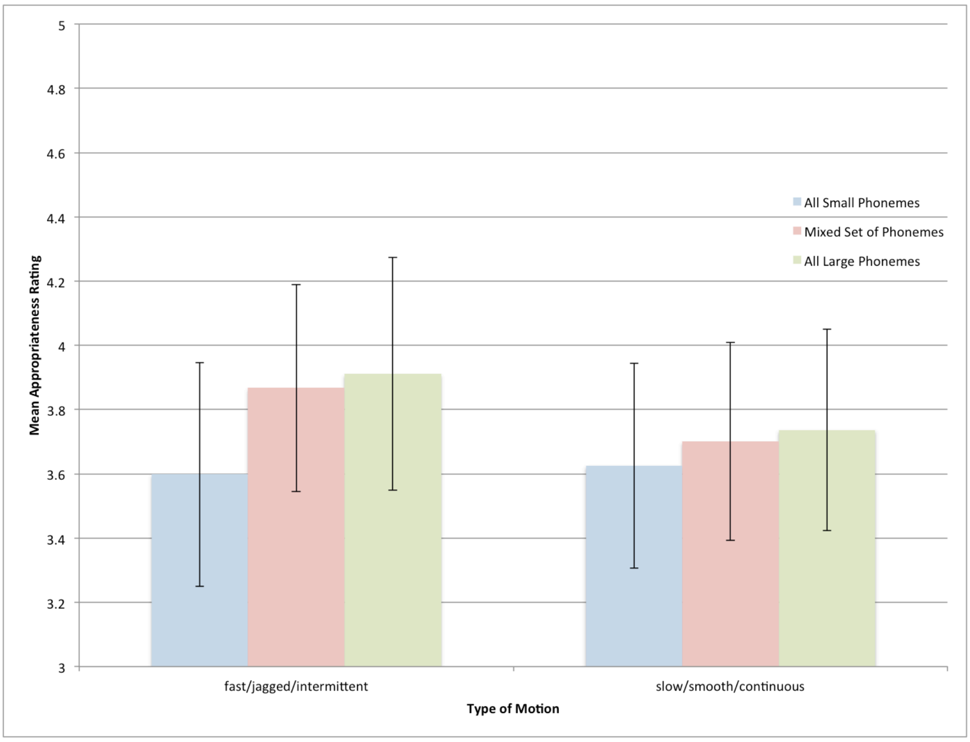


Figure 20. Experiment 8: Mean appropriateness ratings for names of self-locomoting greebles. Error bars represent 95% confidence intervals.

This seems to imply that when naming an object, its common type of motion is not a primary consideration. In order to further test this, data from Experiment 7 and 8 were compared, with experiment acting as a between subjects independent variable. First, mixed names were analyzed to see if they interacted with name and motion. They did not, $F(1,67) = .05$, $p>.05$, and the two mixed conditions were combined. A 2 (experiment) x 3 (name) x 2 (motion) repeated measures ANOVA was performed to see if there was an interaction between experiment and name and motion. There was not, $F(2,132) = .25$, $p>.05$. As such, the data sets were combined. A 3 (name) x 2 (motion) repeated measures ANOVA was performed to see if there was an interaction between name and motion. There was not, $F(1.53,102.57) = 1.20$, $p>.05$. Additionally, there was no main effect of motion, $F(1,67) = .29$, $p>.05$. However, there was a main effect of name, $F(1.71, 114.37) = 5.24$, $p<.01$, partial $\eta^2 = .07$. Additional analyses showed there to be no difference between small and mixed name ratings, $F(1,67) = 2.58$, $p>.05$. Small names were significantly rated as lower than large names, $F(1,67) = 7.56$, $p<.01$, partial $\eta^2 = .10$. Large names were also rated significantly higher than mixed names, $F(1,67) = 4.18$, $p<.05$, partial $\eta^2 = .06$. For a graph of these combined data, please see Figure 21.

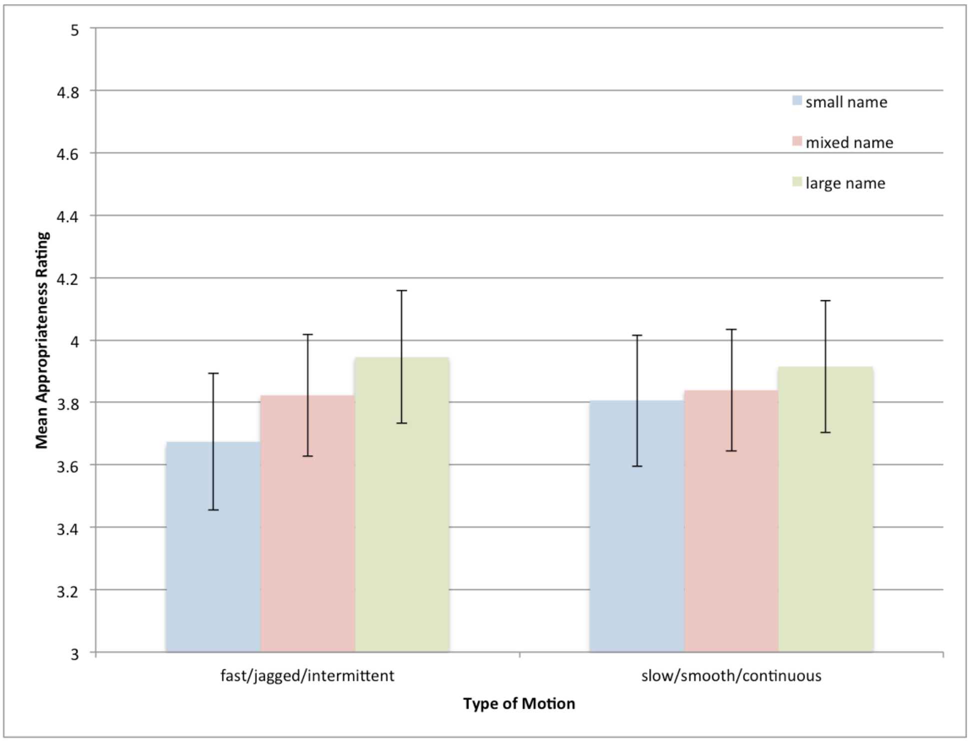


Figure 21. Combined data from Experiments 7 and 8. Error bars represent 95% confidence intervals.

So, even with the increased power brought by combining the data from Experiments 7 and 8, there was no evidence of any interaction between motion being made and the name for the thing doing the motion. This runs contrary to the prediction that motion would be a trait that would be used in phonetic symbolic naming, as has been suggested in previous works. This may be due to motion not being a stable trait, that is, just because it moves that way now does not mean that it always moves that way and cannot move in any other way. There may be the desire, as we may have seen in the previous chapter, to only use phonetic symbolism in the marking of stable traits. As such, the next experiment focused on naming the motion itself, as the motion is in fact a stable trait of itself.

Experiment 9

In Experiment 9, participants heard and saw the same stimuli as in the previous experiment. The participants' task in this experiment was not to rate the appropriateness of the name for the greeble, but rather, for the motion the greeble was making. We predicted that the participants may not be reliably matching motion of object with object name because the greebles' motion may not be stable. However, they should be more likely to name the motion itself sound symbolically, as the motion is a stable trait of itself.

Participants

Data from 46 participants were included in the analyses. Each reported to be a native speaker of English and a member of the University of Warwick community.

Data from two participants were excluded due to multiple responses for a single item. Eight participants skipped items, and their data were excluded. Two participants' data were excluded due to the participants not being native speakers of English. One participant's data were excluded due to the participant having a visual impairment.

Materials

All materials in this experiment were identical to those used in the previous experiment. Participants viewed and heard 48 greeble/name pairs, with greeble motion continuing to be in two gross categories (fast/jagged/intermittent vs. slow/smooth/continuous).

Procedure

The procedure matched that of the previous experiment except that the participants were instructed to assess the appropriateness of the name for the motion of the greeble (and not for the greeble itself). That is, their task was to determine the appropriateness of the possible name for the motion being made by the greeble.

Results and Discussion

A comparison of the mixed audio stimuli was performed to see if there was an interaction between mixed audio conditions and name and movement. There was such an interaction, $F(1,45) = 4.66$, $p<.05$, partial $\eta^2 = .09$. As such, the mixed audio conditions were not combined. A 4 (name) x 2 (movement) repeated measures ANOVA showed there to be a significant interaction between motion and name, $F(2.21, 99.46) = 3.24$, $p<.05$, partial $\eta^2 = .07$. There was no main effect of movement, $F(1,45) = 2.93$, $p>.05$.

There was, however, a main effect of name, $F(3,135) = 11.00$, $p<.001$, partial $\eta^2 = .20$. Within the data from the fast/jagged/intermittent condition, the appropriateness rating of the mixed set, small first names (mean = 3.74) was significantly less than the ratings for small names (mean = 4.06), $t(45) = 2.52$, $p<.05$; for large first names (mean = 4.15), $t(45) = 3.56$, $p<.01$; and for large names (mean = 4.12), $t(45) = 3.39$, $p<.01$. None of the other names differed from each other significantly. In the slow/smooth/continuous data, the all large names (mean = 4.18) were rated significantly higher than small names (mean = 3.65), $t(45) = 3.88$, $p<.001$; than small first names (mean = 3.82), $t(45) = 4.00$, $p<.001$; and than large first names (mean = 3.84), $t(45) = 3.03$, $p<.01$. Again, the other names in that condition did not differ from each other.

In comparing names across motion conditions, we find that ratings for the all small names, $t(45) = 2.36$, $p<.05$, and the large first names, $t(45) = 2.40$, $p<.05$,are both significantly higher when the greeble is making a fast/jagged/intermittent motion Pairwise comparisons across the two motion conditions for both the mixed, small first and the all large names show there to be no significant differences. For a graph of these data, please see Figure 22.

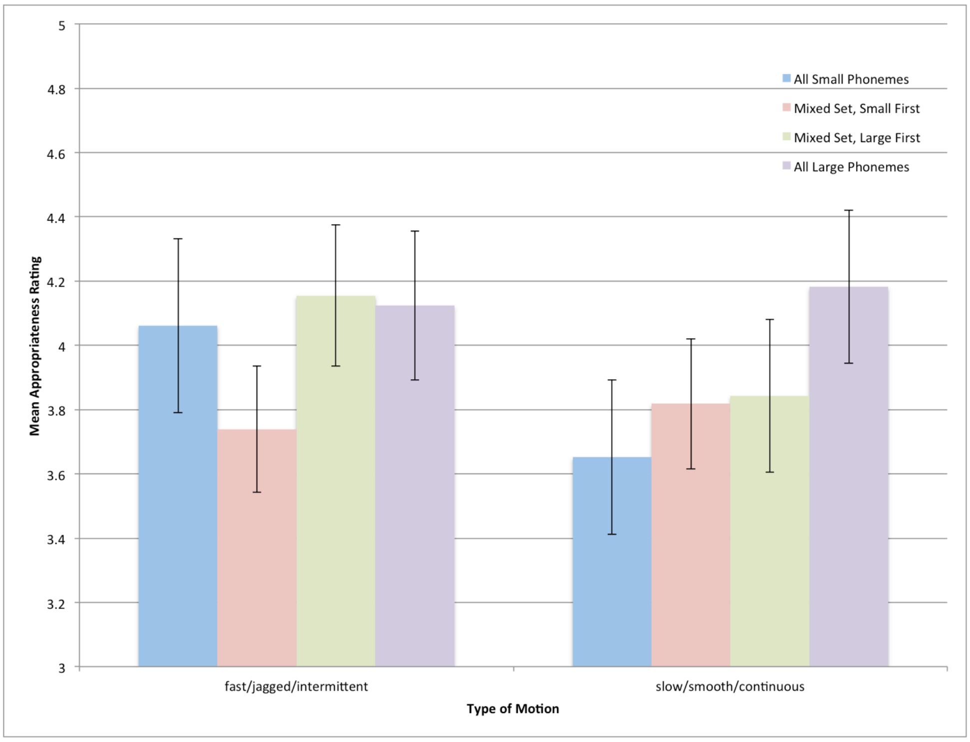


Figure 22. Experiment 9. Appropriateness of names for different motions made by the greebles, with naming for the motion itself. Error bars represent 95% confidence intervals.

Additionally, an analysis was done using data from Experiments 8 and 9. A 4 (name) x 2 (movement) x 2 (experiment) repeated measures ANOVA showed there to be a significant interaction between name, movement, and experiment, $F(2.35,180.69) = 2.689$, $p<.05$, partial $\eta^2 = .03$. This strongly suggests that the naming for motion was significantly different than naming for the object in motion, even when the participants were instructed that the object was self-locomoting. There was no interaction between name and experiment, $F(3,231) = 2.41$, $p>.05$, nor between movement and experiment, $F(1,77) = .04$, $p>.05$. Also, there was no significant interaction between name and movement, $F(3,231) = 1.24$, $p>.05$.

There was, however, a main effect of movement, $F(1,77) = 4.21$, $p<.05$, partial $\eta^2 = .05$, with ratings being higher when fast/jagged/intermittent movements were being presented. There was also a main effect of name, $F(3, 231) = 9.92$, $p<.05$, partial $\eta^2 = .11$. Small and small first names were not significantly different from each other, $F(1,77) = .17$, $p>.05$, though small and large first names were, $F(1,77) = 11.90$, $p<.01$, partial $\eta^2 = .13$, with large first names being rated as higher than small names. Large names were also rated as significantly higher than small names, $F(1,77) = 16.39$, $p<.001$, partial $\eta^2 = .18$. Large first names had significantly higher ratings than did small first names, $F(1,77) = 8.64$, $p<.01$, partial $\eta^2 = .10$; as did large names, $F(1,77) = 17.76$, $p<.001$, partial $\eta^2 = .19$. Large first and large names did not significantly differ from one another, $F(1,77) = 1.87$, $p>.05$.

General Discussion

In this chapter, we reported four experiments involving the naming of motion using phonetic symbolism. In Experiment 6, greebles of varying speed were presented with nonword names and participants were asked to rate the nonword for its appropriateness for the greeble. There was no significant interaction found between name of greeble and speed. In Experiment 7, continuity and path shape were added to speed to create two gross motion conditions. The task was the same as in Experiment 6, as were the results, i.e., no significant interaction between motion and name. For Experiment 8, we considered that it may be that the participants are not viewing the motion made by the greebles as their own doing, that is, that they were just objects being moved around by an outside force. As such, in Experiment 8 the participants were explicitly instructed that the greebles were self-locomoting. Unfortunately, this also led to a non-significant result.

Participants in Experiment 9 received the same stimuli as those in Experiment 8, but had the instructions to judge the name for its appropriateness for the motion, not for the greeble making the motion. In this experiment, there was a significant interaction between motion and name for the motion. It appears that much of this interaction is driven by the naming of the fast/spiky/intermittent motion condition, with both small and large first names being rated significantly higher in this condition Additionally, it looks as though the marking is being done with the last cluster. That is, nonword names that end with 'small,' 'spiky' clusters like /ki/ and /pe/ are preferred in the naming of the faster gross motion. Though it may appear at first glance that marking in the slow/smooth/continuous condition is being done by the large name, it may be that the preference for the large name is fairly stable across conditions and the names ending in small clusters are being suppressed in this condition.

As with the results of those in the previous chapter, this is possibly a case of participants only using phonetic symbolism when naming for a trait that is stable. In Experiments 6, 7, and 8, the greeble was what was being named. The lack of significant results in these experiments could be due to the participants' understanding that the motions being made were not necessarily constant, i.e., just because it moves that way now does not mean that its only movement is like that. When naming for the motion itself, however, a significant effect was found. This is in line with the idea of naming only for permanent traits, as motion being made is a permanent trait of itself.

Chapter 5:

Conclusion

The focus of this thesis has been the marking for size, shape, and motion via phonetic symbolism, the meaningful use of individual speech sounds. In this conclusion, I will summarize and expound upon the previous four chapters, with the goal of placing the experimental work of the thesis in the context of the previous literature and the models of gestural origins (e.g., Donald, 1993) and Frequency Code (e.g., Ohala, 1994). After that, I will briefly describe some ongoing experiments and future directions for research.

A. Previous literature

Since people began thinking about the nature of language, most have made the assumption that there is no reliable link between the sound of a word and the meaning of the word. And why shouldn't they think so? No one argues that most words have arbitrarily designated sounds, and no one argues against the flexibility that having arbitrariness in language brings. As such, many of the most influential voices in philosophy and linguistics have come out against the idea of there being any systematicity in the mating of sound and meaning, including Aristotle, to whom meaning of a word only comes with the pairing of the sound to the mental experience (350 BCE).

Saussure continues this line of thinking when he proclaimed that "the sign is arbitrary" (2006, p.67). That is, meaning comes at the conjunction of the concept and the psychological sensation of the name, and that bond has "no natural connection" (2006, p.69). He admits there are some isolated cases of sound symbolism, such as onomatopoeia, but even these are integrated into the rules of the language they come from, and, as such, have no direct, natural connection to their referent. This notion of absolute arbitrariness in language was picked up by Hockett (1960), who included it in his design features of language. This led the way for Chomsky and the Generativists' idea that phonemes are merely the surface structure of the meaning-bearing lexicon (which is part of the deep structure).

However, there is growing evidence that phonetic symbolism does exist, that it has a purpose, and that its existence may give us insight into our linguistic past. (As these topics were covered in some length in Chapter 1, they will only be briefly discussed here.) The strongest evidence appears to come from experiments that involve the use of nonwords, either abstractly or in the naming of novel objects. For instance, in experiments done by Sapir (1929) and Newman (1933), participants were given a nonword pair that differed only by one phoneme and asked which of the two would go with an object that had x property, e.g., was the larger horse. They both found there to be a general trend towards naming for size based on backness of vowels, i.e., bigger is marked with back vowels. Additionally, Newman found there to be similarly objective underpinnings for consonant marking, e.g., voiced is larger than voiceless. Specifically, he points to tongue position (front is smaller than back), vocal pitch (higher pitch is smaller than lower pitch), and size of mouth when making the sound (smaller being smaller). Many recent studies, e.g., Berlin (2006), have supported these findings.

Similarly, there has been a lot of evidence arising from studies that involve naming novel objects. The most famous of these is the Kohler (1947) takete/maluma example (see Figure 1), whereby people are asked which of the two words goes to the rounded or spiky of the two figures. With few exceptions, e.g., Davis, 1961, this has been replicated across many cultures with great success. More modern interpretations, such as those done by Maurer, Mondloch, and Pathman (2006), have shown that this type of matching occurs not only with adults, but also with young children.

Additionally, experiments have shown there to be an increased ability for language learning when phonetic symbolism is employed. Mondloch and Maurer (2004) suggest that the innate tendency to match shape and sound could be used as a stepping-stone to learning of first vocabularies. Nygaard, Cook, and Namy (2009) demonstrated that phonetic symbolism also helps in the learning of foreign language vocabulary. Moreover, experiments involving action words suggest that sound symbolism is useful in the learning of linguistic categories (Imai, Kita, Nagumo, Okada, 2008).

Finally, phonetic symbolism may have been an intermediary step in our language evolution. If, as many argue, e.g., Donald (1993), language arose from a mimetic origin, then it could be that phonetic symbolism could have been a step in between the highly symbolic mimesis and our current, mostly arbitrary language of today. Such a transition from gesture could have taken place via cross-modal connections between the language centers of the brain and the motor control system, with a more fully sound symbolic language being the beginning result of that crossing. This cross-modal processing also gives rise to Frequency Code (Ohala, 1994), whereby symbolism for things such as size is created by the modulation of vocal pitch. Of note, both gesture and frequency code are graded functions, i.e., they are not used categorically, but rather are continuous.

Some questions are still unanswered regarding phonetic symbolism. First, how universal is it, and why do some languages seem more apt to be comparable to other languages, for instance, Japanese and English seem to pair well in cross-language studies, but Swahili and English do not seem to. Surely, the influence of culture is at work here, but to what extent does culture influence phonetic symbolism?

It appears that phonetic symbolism exists on two levels. First, there is the physiologically-motivated level. That is, phonetic symbolism that matches the meaning of a word and the sound of the word via vocal-gesture. For example, when naming a large object, there could be the tendency to name it based on the property of size. In order to do this, one would replicate the largeness of the object's size by using sounds that involve making a larger mouth cavity when speaking, e.g., back vowels. In this way, the speaker is using a naturally-motivated matching of sound and referent.

Additionally, there could be the coupling of sound and meaning that is not naturally-motivated, but is, rather, just statistically driven. Imagine, if you will, that the first thing made of wood was a chair. If enough things made of wood started to be named with /ch/, then /ch/ would start to be phonetically symbolic of 'things made of wood,' despite there being no natural link between the way /ch/ is made and the nature of wood. Of course, these two types of sound symbolism are not mutually exclusive, e.g., naturally-motivated symbolism would create a statistical component, so all naturally-motivated symbolism is also statistical, but not all statistical symbolism is naturally-motivated.

Additionally, if naturally-motivated symbolism is indeed naturally motivated, i.e., motivated by our use of our physiology in naming, then there should be a lot of overlap across languages in terms of traits of this type of sound symbolism. For example, the mouth as a speech apparatus is basically the same across peoples, so shouldn't the symbolism they employ be as well? Good candidates for this sort of symbolism would be those that lend themselves to being easily symbolized through mouth shape and articulation (vocal gesture) and vocal pitch (frequency code), such as size, shape, and motion.

Such universality (or near-universality) would not be the case with statistical sound symbolism. Though the symbolic sounds could be psychologically felt, e.g., phonesthemes (Bergen, 2004), the statistically-motivated phonetic symbolism would be highly culture specific. As such, all languages may have both naturally-motivated and statistically-motivated sound symbolism, but naturally-motivated symbolism would tend to be universal, while statistical symbolism would tend to vary greatly across languages.

B. Size, Experiments 1 and 2

"If it is in fact possible to make any claims for universality in sound symbolism, then those claims will have to begin with magnitude sound symbolism" (Nuckolls, 1999b, p. 229)- and this is what we did. One of the primary questions left from the literature was that despite gesture and frequency code being continuous functions, much of the previous research had used dichotomous stimuli, e.g., Berlin, 2006. As such, their findings did not speak specifically to the hypotheses that arose from gesture and frequency code as being the underlying factors in phonetic symbolism. Since the natures of gesture and frequency are graded, then shouldn't we expect to find sound symbolic naming that is based on these functions to also be graded? If we did not find that to be the case, then we should expect that gesture and frequency code were not underlying sound symbolic marking for size, suggesting instead that statistical (non-motivated) marking may be what is actually going on. Additionally, the use of graded stimuli would help to lessen the effect of strong demand characteristics, such as those found in Sapir (1929).

In Experiment 1, we presented participants with greebles in various sizes in a pastoral scene, and gave them the task of choosing which of the five names (which varied in number of 'large' and 'small' phonemes) listed below the picture was the most appropriate name for the greeble. An analysis based on the mean number of large phonemes per greeble size showed there to be a significant linear effect. That is, as the greeble's size increased, so too did the tendency to use more large phonemes in its name.

However, there were possible confounds arising from our use of written names instead of spoken names. First, and most obviously, the study was meant to be about sound symbolism, not grapheme symbolism. Second, the names with more large phonemes also tended to be physically longer, i.e., in pixels, than those with more small phonemes, leading to the question of whether or not participants were marking based on speech sounds in the name or on the physical length of the name. As such, we did a follow-up experiment (Experiment 2), which used simplified stimuli that were presented auditorily. On each trial, participants saw a greeble in one of three sizes, standing next to an abstract human figure. They then heard three possible names for the greeble, which varied in their phonetic content, including all large, half large, and no large phonemes. Their task was to choose which of the names best fit the greeble. Again, we found there to be a significant linear trend towards naming the greebles, i.e., as the greeble got larger, so too did the mean number of large phonemes in its chosen name. The effect size in Experiment 2 was slightly less than that of Experiment 1, suggesting that the participants may have used the nature of the written words as a reinforcement of the speech sounds. In both experiments, however, there was the tendency to name objects of increasing size with an increasing number of large phonemes.

C. Size and shape, Experiments 3-5

Though Chapter 2 presented two experiments that were an improvement over two-item forced-choice experiments, they still were forced choice, which may have caused the participants to answer in ways that they would not have done so had they not had the limited number of choices presented to them. Additionally, like many previous studies, e.g., Westbury, the experiments focused on only one property- size. Though an effort was made to make sure the sizes were all appropriate, there was little attention paid to the shape of the greebles being presented. As in other experiments that only examine size or shape, these two properties may be unintentionally confounded. It also calls into question the results of such studies that suggest that the same phonemes being used to mark size are also being used to mark shape, e.g., smallness and spikiness being marked with the same phonemes. Additionally, it could be that only one of the properties is actually being marked for, with the other property just being highly related, as was found in O'Boyle and Tarte, 1980.

As such, Chapter 3 included experiments that used the same experimental method and similar stimuli to examine first size, then shape, and finally, size and shape. The research procedure involved the participant viewing a greeble, manipulated for size and/or shape, next to an abstract human figure. The participant would then hear a possible name for the greeble and rate the appropriateness of that name on a 7-point scale, with 7 being most appropriate. This lessens demand characteristics, as there is no forced choice, and the graded nature of the stimuli and responses allows for the continued testing of the gestural and frequency code hypotheses. Additionally, since we are using a similar set of stimuli and an identical procedure for each experiment in this series, it allows for a direct comparison of the marking for size and for shape.

In Experiment 3, participants showed a strong tendency to mark large greebles, and to a lesser extent, medium greebles. However, they showed no preference for name with small greebles. We have two concerns regarding these findings. First, as the medium greeble was approximately the same height as the human figure, it may have been viewed as being fairly large, leading to the results that we see. Second, it is likely that the greebles were viewed as natural kinds. If viewed in this light, it is not unreasonable to think that the participants felt the small greeble was still growing, the medium greeble was fully grown, and the large greeble was particularly large. Or, it could have been that they felt they were viewing a baby (small), a mother (medium), and a father (large), and were marking them for gender, e.g., Cassidy, Kelly, and Sharoni, 1999. Since the gender of the baby was not known, it was left unmarked. In either case, it appears that a trait needs to be stable before it will be used for marking. It is worth mentioning that these results were at odds with the results from Chapter 2. In Chapter 2, it appears that all levels of size were marked, with phonemes hypothesized to be small marking small items and phonemes hypothesized to be large marking large items. However, the results from this experiment was not so clear, suggesting that even though we had more possibilities in Experiments

1 and 2 than one normally finds in a forced-choice experiment, participants were still being swayed by being in a forced-choice situation.

In Expriment 4, participants heard the same stimuli and viewed greebles that were manipulated for shape (and had been rated during a pre-test). The results suggest that both medium shape (with small first names) and round shape (with large names) were marked, though, spikiness did not appear to be particularly marked for. Large first names were used in almost the same way as all large names were, suggesting that roundness may be marked with the onset cluster, while medium shape may be marked with a small onset cluster and a large ending cluster. It is worth noting that had we only had spiky and round conditions, it would have appeared that there was a simple, clean interaction with both spikiness and roundness being marked with the expected phonemes.

In Experiment 5, we used greebles that were manipulated for both size and shape, with the participants again rating the appropriateness of the names they heard. With small names, there was no real tendency to mark for shape. However, size seemed to play a dominant role. Very little seemed to be going on in terms of preference of name when the name was mixed. However, when the name was large, a strong preference for both roundness and the larger two conditions emerged. This may suggest that there is an interaction for size and shape with phonetic marking, mainly though, when the greeble is round and fairly large. This makes some sense, as roundness is generally a stable trait, it would be marked no matter the size of the greeble, whereas only largeness would be marked as the small greebles may have been seen as immature. Generally, it can be said that marking for size appears to be only when the object is fairly large, while marking for shape appears to be both for spiky and for round.

D. Motion, Experiments 6-9

Although motion has been cited as a property that is phonetically marked, most of the studies that make this suggestion have used static stimuli, e.g., Berlin, 2006 and Maurer, et al., 2006. As such, it is unclear what exactly about motion is being marked, i.e., what property of motion is being marked? In Experiment 6, we investigated the most likely of these properties – speed. Experiment 6 involved the same auditory stimuli as in the previous set of experiments and medium-sized greebles, which moved across the screen in counter-balanced directions in three different speeds. Again, the participants' task was to rate how appropriate each name was for the greeble they saw in motion. We found there to be no significant interaction between name and speed.

As speed is only one part of movement, Experiment 7 involved greebles that moved in two gross conditions. In one condition, greebles moved rapidly, in a jagged pattern, and stopped and started repeatedly. In the other condition, greebles moved slowly in a circular pattern with no stopping. Again, there was no interaction of name and motion. Since whether or not the object is taken to be moving on its own has been shown to be relevant (Pratt et al., 2010), we included in Experiment 8's instructions that the object in motion was moving itself across the screen. The rest of the task was identical to the previous experiment. Unfortunately, so were the results – there was no significant interaction between name and movement, even when participants were told that the greeble was moving itself. It could be that participants did not find the objects' self-locomotion to be particularly relevant. It could also be that the instructions were not bold enough to have made an impact on the participants. That is, the wording of the instructions may have been too weak to have made any impact on the participants. This suspicion is supported by an analysis of the data from Experiments 7 and 8, which showed there to be no interaction between experiment, name, and motion. This suggests that the participants did not alter their ratings preferences based on the instructions given to them. This may have been due to weak instructions or to their desire to not mark an object based on its motion.

In Experiment 9, we asked the participants to judge the name for the motion being made by the greeble, and not for the greeble itself. This time, there was a significant interaction between name and motion, though the effect size was quite small (partial $\eta^2 = .07$). The data suggest that there really is no preference when naming fast/jagged/intermittent motions, other than that small first names are not particularly appropriate. However, it appears that slow/smooth/continuous motions are marked very clearly with large names. It appears though, that the data do not suggest such a simple picture. Instead of large names being only preferable in naming slow/round/continuous motions, they also seem to be equally appropriate when naming fast/jagged/intermittent motions. What really appears to be the case is that the ratings for small and large first names are depressed when the motion is slow/round/continuous, leading to large names being most appropriate for slow/round/continuous. So, in terms of marking for motion, marking is only done for the motion itself, and not for the object doing the marking.

E. Theoretical Contributions

The biggest single theoretical contribution of this thesis has come in the demonstration that phonetic symbolic marking is not as clear-cut as previous studies have suggested. By our use of graded stimuli and responses, we are able to see that had we employed a pairwise dichotomous forced choice paradigm, we would likely have seen a different, and misleadingly incomplete, picture of the phenomenon. Not only would we have increased demand characteristics, which have been shown to be an issue in previous research, but we also would have only seen what marking at the ends of the spectra look like. Using such paradigms seems even the more inappropriate when the hypotheses being tested are based on the graded functions of gesture and frequency code.

Additionally, in Chapter 2, we are able to show that sound symbolism for size appears to be a graded function. Previous research had generally used dichotomous stimuli and were unable to show this, despite their claims that it supported the theory that phonetic symbolism arises from graded functions. These experiments directly support the hypotheses arising from gesture and frequency code. Additionally, by our use of written stimuli in Experiment 1 and auditory stimuli in Experiment 2, we were able to lend support to the idea that written stimuli allows for reinforcement of the speech sounds, leading to larger effect sizes and more successful experiments. As many recent experiments still employ written stimuli, this supporting finding of previous works is not trivial.

Previous works have tended to either examine size or shape, but not both within the same paradigm. When they have, it has tended to either be using pure tones or dichotomous stimuli, with the latter commonly suggesting that the same phonemes are being used to mark both size and shape. In Chapter 3, we examine the two both separately and together, using the same paradigm and graded stimuli and responses. In doing so, we find that marking for size appears to be mostly when the stimuli are fairly large, i.e., at least the size of a human. There appears to be no preference of name when the greeble is small, but marking for large follows the expected pattern, i.e., large names are strongly preferred, mixed names are in the middle, and small names are not at all preferred.

With shape, we found there to again be an interaction, but with a somewhat different pattern. Small names and small first names were preferred for spiky greebles, i.e., based on onset cluster, the medium shape was strongly marked with small first names, and the round greebles were marked with large and large first names. It is worth noting that had we used dichotomous stimuli, the pattern would have appeared to suggest that only spiky and round were marked and that the interaction was clear-cut.

Additionally, there is an interaction between size and shape with phonetic marking. With small names, the appropriateness ratings suggest that shape is not of primary concern. Rather, it appears that greeble size is what is leading to the ratings, with a general tendency to rate the small names as less appropriate the larger the greeble gets. In the case of the large names, there is still a general trend towards marking based on size.

However, there is a separation of preference based on shape, especially in terms of the larger round greebles.

In Chapter 4, we examined the question of whether motion was a consideration in sound symbolic naming. Motion had been suggested as one of the major factors considered when naming an object. In Experiment 3, however, we found that speed alone did not elicit any preferences for name when naming the object in motion, and neither did the gross movement conditions that included speed, pattern of movement, and movement continuity. Even when participants were explicitly told that the greeble was self-locomoting, they still did not display any phonetic symbolic marking based on motion. Only when we asked the participants to assess the name for the motion being displayed by the greeble did we see an interaction of name and motion. This suggests that marking an object for the motion that it makes is unlikely. Instead, the marking for the object, if it is indeed done, comes from the relationship it has to certain movements, which would be marked phonetically.

Additionally, we are able to add as a possible prerequisite for marking that of trait stability. In Experiment 3 (size) and especially in the motion experiments, it became clear that participants are not willing to make any marking decisions unless the trait they are marking for is at least semi-permanent. This has not been specifically discussed in other works, and may only show up when there is some autonomy in the decision making process, i.e., the experiment is not based on forced choice. If the stimuli are artifacts, then there is no expectation of change over time, but with natural kind-like stimuli, i.e., greebles, size is only marked in a linear way when there is some sort of forced choice involved.

F. Limitations

There are several important limitations to note in these experiments. First, the choice of auditory stimuli used in Experiments 2-9 included a limited number of phonemes (six consonants and four vowels). Though these phonemes were chosen because they had been shown to be particularly potent in similar experiments, it may be that certain types of symbolic marking are done with other phonemes. This could lead us to conclude that no marking has been done, when in reality, it is being done with other phonemes. However, the phonemes that we chose were those that showed up time and again in the literature (front v. back vowels, voiced v. voiceless stops). They gave us a good amount of flexibility in the making of nonwords and their contrasts were easily understood and accounted for. Had we employed more phonemes, it would likely have just muddled the picture in terms of how marking was being done.

Additionally, the choice of greebles as the visual stimuli may have led us to conflate size with maturity and/or sex. Greebles were a good choice for stimuli that could easily be manipulated for size and shape. Additionally, they had been shown to have recognizable attributes, like those in objects we see daily, but with the added advantage of not actually looking like anything we normally would be familiar with. However, by putting them next to abstract human figures, we may have opened the door for participants to think that they were natural kinds, and let their assessments be swayed by thoughts of the greeble life cycle, their sexual dimorphism, and possibly their vocalizations.

The choice of methodology from Experiments 3-9 may have been a bit flawed as well. It may have been a better paradigm to ask the participants to name the items/movements themselves, possibly using a set inventory of phonemes. Although this may have been more like the actual task of naming an unknown object for the first time, it too has its issues. For one, having any constraints on the participant in terms of phonemic inventory leads to a lessening of the ecological validity. Secondly, having fewer constraints leads to a lack of ability to keep number of variables in check. For example, in the stimuli we used, the sounds were in CV clusters. How does one effectively analyze data that include CCV, CVV, CCCV clusters without adding a nearly indefinite number of conditions to their analysis? Additionally, any task that involves the creation of names will undoubtedly take longer than one that is just assessing names. As previous research had suffered from a lack of stimuli items, we decided to go with a paradigm that would allow for the maximum number of data points within a reasonable amount of time, i.e., an amount of time that would not cause undo fatigue for the participants.

That being said, the nonword names were well controlled for and easily analyzed. The greebles still seem to be the best choice for these tasks, especially as we suggest in one experiment that they are self-locomoting. The choice of paradigm was also defendable. By changing the question from "how would you name this?" to "how good is this name?," we were able to get a lot more trials per participant and have better control over the responses.

G. Current and Future Directions

One of the current research projects that we are working on involves the phonetic marking for affect. We have completed a corpus analysis of 5188 English words rated for both valence and arousal. Through the use of python programming, we ran the words through an online pronunciation dictionary. We then took an inventory for each phoneme and used them as predictors in a regression analysis (with the predicted values being valence and arousal). Phonemes with a significant p value were then used to make nonwords to be used in a lexical decision task. In the experimental phase of the study, participants were presented with a prime (either negative or positive nonword) for 200ms, an ISI for 250ms, and then a target (either a negative or positive nonword or a neutral word). Our hypothesis is that participants will be slower to respond to the real words when primed with a negative nonword. Final analyses are still ongoing.

An additional ongoing project in the initial stages involves seeing if people will be willing to pay more for a box of chocolates if the box has a larger name. Using a paradigm that involves telling participants that they will have to use real money in a randomly-drawn trial, we ask participants to view a box of chocolates that may contain a small, medium, or large chocolate figurine. We are hoping to show that participants pick up on the phonetic symbolism and show willingness to pay more for a box with a larger name (as it may imply that it contains a larger piece of chocolate).

References

Alspach, E. (1917). On the psychological response to unknown proper names. *The American Journal of Psychology, 28*(3), 436-443.

Aristotle. (350 BCE). On interpretation.

Atzet, J., & Gerard, H. (1965). A study of phonetic symbolism among native Navajo speakers. *Journal of Personal and Social Psychology, 1*, 524–527.

Bentley, M., & Varon, E. (1933). An accessory study of phonetic symbolism. *The American Journal of Psychology*, 76-86.

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, 290-311.

Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols & J. J. Ohala (Eds.), *Sound Symbolism* (pp. 76-93). Cambridge: Cambridge University Press.

Berlin, B. (2006). The first congress of ethnozoological nomenclature. *The Journal of the Royal Anthropological Institute, 12*(s1), 23-44.

Birch, D., & Erickson, M. (1958). Phonetic symbolism with respect to three dimensions from the semantic differential. *The Journal of general psychology, 58*(2), 291.

Bolinger, D. (1949). The sign is not arbitrary. *Boletín del Instituto Caro y Cuervo, 5*, 52-62.

Bolinger, D. (1975). *Aspects of Language*. Atlanta: Harcourt Brace Jovanovich, Inc.

Brackbill, Y., & Little, K. (1957). Factors determining the guessing of meanings of foreign words. *Journal of abnormal psychology, 54*(3), 312.

Brown, R. (1968). *Words and Things*. New York: Free Press.

Brown, R., Black, A., & Horowitz, A. (1955). Phonetic symbolism in natural languages. *Journal of Abnormal Psychology, 50*(3), 388-393.

Brown, R., & Nuttall, R. (1959). Methods in phonetic symbolism experiments. *Journal of abnormal and social psychology, 59*, 441.

Cassidy, K. W., Kelly, M. H., & Sharoni, L. (1999). Inferring gender from name phonology. *Journal of Experimental Psychology: General, 128*(3), 362-381.

Chomsky, N., & Halle, M. (1991). *The sound pattern of English*. Boston: MIT Press.

Coulter, K. S. (2008). *The effects of phonetic symbolism on comparative price perceptions*. Paper presented at the Association for Consumer Research.

Coulter, K. S., & Coulter, R. A. (2010). Small sounds, big deals: Phonetic symbolism effects in pricing. *Journal of Consumer Research, 37*, 315-328.

Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology, 52*(3), 259-268.

Demerse, B. J. (1941). *The scaling of two-syllable adjectives in terms of their sound and of their connotation.* University of Iowa, Unpublished Master's Thesis.

Diffloth, G. (1994). i: big, a: small. In L. Hinton, J. Nichols & J. Ohala (Eds.), *Sound Symbolism* (pp. 107-114). Cambridge: Cambridge University Press.

Donald, M. (1991). *Origins of the modern mind : Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.

Donald, M. (1993). Precis of Origins of the modern mind: Three stages in the evolution of culture and cognition. *Behavioral and Brain Sciences, 16*, 737-791.

Eberhardt, M. (1940). A study of phonetic symbolism of deaf children. *Psychological Monographs, 52*, 23-41.

English, G. (1916). On the psychological response to unknown proper names. *The American Journal of Psychology, 27*(3), 430-434.

Farmer, T., Christiansen, M., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*(32), 12203.

Firth, J. R. (1935). The use and distribution of certain English sounds. *English Studies, 17*, 8-18.

Fox, C. (1935). An experimental study of naming. *The American Journal of Psychology, 47*(4), 545-579.

French, P. (1977). Toward an explanation of phonetic symbolsim. *Word, 28*, 305-322.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68*(7), 1191.

Gasser, M. (2004). *The origins of arbitrariness in language.* Paper presented at the Annual Conference of the Cognitive Science Society.

Hall, K., & Oldfield, R. (1950). An experimental study on the fitness of signs to words. *The Quarterly Journal of Experimental Psychology, 2*(2), 60-70.

Heise, D. (1966). Sound-meaning correlations among 1,000 English words. *Language and Speech, 9*, 14-27.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*, 88-96.

Holland, M., & Wertheimer, M. (1964). Some physiognomic aspects of naming, or, maluma and takete revisited. *Perceptual and Motor Skills, 19*(1), 111-117.

Huang, Y. H., Pratoomraj, S., & Johnson, R. (1969). Universal magnitude symbolism. *Journal of Verbal Learning and Verbal Behavior, 8*(1), 155-156.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54-65.

Irwin, F., & Newland, E. (1940). A genetic study of the naming of visual figures. *Journal of Psychology, 9*, 3-16.

Jespersen, O. (1922). *Language- its nature development and origin*. New York: Henry Holt.

Jespersen, O. (2010). Symbolic value of the vowel i *Selected Wrtitings of Otto Jespersen* (pp. 287-300). New York: Routledge.

Johnson, R. (1967). Magnitude symbolism of English words. *Journal of Verbal Learning & Verbal Behavior, 6*(4), 508-511.

Johnson, R., Suzuki, N., & Olds, W. (1964). Phonetic symbolism in an artificial language. *Journal of Abnormal and Social Psychology, 69*(2), 233-236.

Jurafsky, D. (1996). Universal tendencies in the semantics of the diminutive. *Language, 72*(3), 533-578.

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound symbolism facilitates word learning in English speaking children. *Cognitive Science, 35*(3), 575-586.

Kohler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright Publishing Corporation.

Kollarits, J. (1914). Observations de psychologie quoidienne. *Archive de Psychologie, 14*, 225-240.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition, 114*(1), 19-28.

Lowrey, T., & Shrum, L. (2007). Phonetic symbolism and brand name preference. *Journal of Consumer Research, 34*, 406-414.

Maltzman, I., Morrisett, L., & Brooks, L. (1956). An investigation of phonetic symbolism. *Journal of Abnormal and Social Psychology, 53*(2), 249-251.

Marks, L. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 384-394.

Marks, L. (2004). Cross-modal interactions in speeded classification. In G. Calvert, C. Spence & B. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 85-106). Cambridge, MA: MIT Press.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology, 87*(1/2), 173-188.

Maurer, D., Pathman, T., & Mondloch, C. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science, 9*(3), 316-322.

Miall, D. (2001). Sounds of contrast: An empirical approach to phonemic iconicity. *Poetics, 29*, 55-70.

Miron, M. (1961). A cross-linguistic investigation of phonetic symbolism. *Journal of abnormal and social psychology, 62*, 623.

Monaghan, P., Chater, N., & Christiansen, M. (2003). *Inequality between the classes: Phonological and distributional typicality as predictors of lexical processing.* Paper presented at the Proceedings of the 25th annual conference of the cognitive science society.

Monaghan, P., & Christiansen, M. H. (2006). Why form-meaning mappings are not entirely arbitrary in language. *Proceedings of the 28th annual conference of the cognitive science society*, 1838-1843.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General, 140*(3), 325-347.

Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience, 4*(2), 133-136.

Newman, S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology, 45*(1), 53-75.

Newmeyer, F. J. (1992). Iconicity and Generative Grammar. *Language, 68*(4), 756-796.

Nuckolls, J. (1999a). The case for sound symbolism. *Annual Review of Anthropology, 28*, 225-252.

Nuckolls, J. (1999b). The case for sound symbolism. *Annual Reviews of Anthropology, 28*, 225-252.

Nygaard, L., Cook, A., & Namy, L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112*, 181-186.

O'Boyle, M., Miller, D., & Rahmani, F. (1987). Sound-meaning relationships in speakers of Urdu and English: Evidence for cross-cultural phonetic symbolsim. *Journal of Psycholinguistic Research, 16*(3), 273-288.

O'Boyle, M., & Tarte, R. (1980). Implications for phonetic symbolism: The relationship between pure tones and geometric figures. *Journal of Psycholinguistic Research, 9*(6), 535-544.

Ohala, J. J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols & J. J. Ohala (Eds.), *Sound Symbolism* (pp. 325-347). Cambridge: Cambridge University Press.

Orr, J. (1944). On some sound values in English. *British Journal of Psychology. General Section, 35*(1), 1-8.

Osgood, C. (1962). Studies on the generality of affective meaning systems. *American Psychologist, 17*(1), 10-28.

Osgood, C., & Suci, G. (1955). Factor analysis of meaning. *Journal of Experimental Psychology, 50*(5), 325-338.

Paget, R. (1929). Human speech. *Nature, 123*, 281-284.

Pinker, S. (2000). *The language instinct: How the mind creates language*: Harper Perennial Modern Classics.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13*(4), 707-784.

Pitcher, B., Mesoudi, A., & McElligott, A. (2013). Sex-Biased Sound Symbolism in English-Language First Names. *PLoS ONE, 8*(6), e64825.

Plato. (360 BCE). Cratylus.

Pratt, J., Radulescu, P., Guo, R. M., & Abrams, R. A. (2010). It's alive! Animate motion captures visual attention. *Psychological Science, 21*, 1724-1730.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia -- A window into perception, thought and language. *Journal of Consciousness Studies, 8*, 3-34.

Roblee, L., & Washburn, M. F. (1912). The Affective Values of Articulate Sounds. *The American Journal of Psychology, 23*(4), 579-583.

Saffran, J. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*(4), 110-114.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*(3), 225–239.

Saussure, F. (2006). *Course in General Linguistics*. Peru, IL: Open Court.

Scott, R. (Writer). (1982). Blade Runner [Film]. In M. Deeley (Producer). Los Angeles: Warner Brothers.

Shintel, H., & Nusbaum, H. C. (2007). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition, 105*, 681-690.

Siegel, A., Silverman, I., & Markel, N. N. (1967). On the effects of mode of presentation on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 171-173.

Slobin, D. (1968). Antonymic phonetic symbolism in three natural languages. *Journal of Personality and Social Psychology, 10*(3), 301-305.

Solomon, L. N. (1958). Semantic approach to the perception of complex sounds. *The Journal of the Acoustical Society of America, 30*, 421.

Solomon, L. N. (1959). Search for physical correlates to psychological dimensions of sounds. *The Journal of the Acoustical Society of America, 31*, 492.

Spence, C. (2011). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology, 22*, 37-54.

Spence, C., Senkowski, D., & Roder, B. (2009). Crossmodal processing. *Experimental Brain Research, 198*(2), 107-111.

Tarte, R., & Barritt, L. (1971). Phonetic symbolsim in adult native speakers of English: Three studies. *Language and Speech, 14*(2), 158-168.

Taylor, I. (1963). Phonetic symbolism re-examined. *Psychological Bulletin, 60*, 200-209.

Taylor, I., & Taylor, M. (1962). Phonetic symbolism in four unrelated languages. *Canadian journal of psychology, 16*, 344-356.

Taylor, I., & Taylor, M. (1965). Another look at phonetic symbolism. *Psychological Bulletin, 64*(6), 413.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology, 64*(12), 2392-2404.

Thorndike, E. (1945). On Orr's hypotheses conserning the front and back vowels. *British Journal of Psychology. General Section, 36*(1), 10-13.

Thorndike, E. L. (1943). The origin of language. *Science, 98*, 1-6.

Thorndike, E. L. (1945). The association of certain sounds with pleasant and unpleasant meanings. *Psychological Review, 52*(3), 143-149.

Tsuru, S., & Fries, H. S. (1933). A problem in meaning. *The Journal of General Psychology, 8*(1), 281-284.

Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg (Ed.), *Universals of human language* (Vol. 2, pp. 525-568). Stanford, CA: Stanford University Press.

von Hornbostel, E. (1927). The unity of the senses. *Psyche, 7*(28), 83-89.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., et al. (2010). Preverbal infants' sensitivity to synaestheic cross-modality correspondences. *Psychological Science, 21*(1), 21-25.

Walker, P., & Smith, S. (1984). Stroop interference based on teh synaesthetic qualities of auditory pitch. *Perception, 13*(1), 75-81.

Weiss, J. H. (1963). Role of" meaningfulness" versus meaning dimensions in guessing the meanings of foreign words. *The Journal of Abnormal and Social Psychology, 66*(6), 541-546.

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language, 93*(1), 10-19.

Whissell, C. (2001). Sound and emotion in given names. *Names: A Journal of Onomastics, 49*, 97-120.

Wichman, S., Holman, E., & Brown, C. (2010). Sound symbolism in basic vocabulary. *Entropy, 12*, 844-858.