The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission

ALAN NIELSEN AND DREW RENDALL*

University of Lethbridge

Abstract

There exists a fundamental paradox in linguistic cognition. Experiments show consistent sound-symbolic biases in people's processing of artificial words, yet the biases are not manifest in the structure of real words. To address this paradox, we designed an experiment to test the magnitude and source of these biases. Participants were tasked with matching nonsense words to novel object forms. One group was implicitly taught a matching rule congruent with biases reported previously, while a second group was taught a rule incongruent with this bias. In test trials, participants in the congruent condition performed only modestly but significantly better than chance and better than participants in the incongruent condition who performed at chance. These outcomes indicate the processing bias is real but weak and reflects an inherent learning bias. We discuss implications for language learning and transmission, considering the functional value of non-arbitrariness in language structure and underlying neurocognitive mechanisms.

Keywords
symbol grounding, sound symbolism, artificial language learning, Bouba-Kiki, language acquisition

---

* Correspondence address: Drew Rendall, Department of Psychology, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada. Email: d.rendall@uleth.ca. Research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada and approved by the Human Subjects Research Ethics committee of the University of Lethbridge. We thank two anonymous reviewers for comments that improved the manuscript.

Language and Cognition 4–2 (2012), 115–125
DOI 10.1515/langcog-2012-0007

1866–9808/12/0004–0115

1. Introduction

A fundamental tenet of contemporary linguistics, codified by Saussure (1983), is that word meaning arises by convention because there are no clues to meaning in the structure of words themselves. In short, word form is arbitrary. This dictum survives to the present despite sporadic reports to the contrary in the form either of studies showing structural consistencies in the form of certain word classes within and between languages (e.g. Berlin 1994; Brown et al. 1955), or of experiments showing that, for artificial word material, people have strong and consistent expectations about appropriate word structure (Köhler 1947). Such findings are often lumped under the banner of Sound-Symbolism, a term that flags the possibility that language structure is not entirely arbitrary. However, they have done little to challenge the classic Saussurean dictum. Instead, such findings have generally been discounted and marginalized, at best treated as curiosities vastly outweighed by the apparently overwhelming arbitrariness of language (Newman 1933; Newmeyer 1993).

However, there has been a recent resurgence of interest in sound symbolism, one that appears to pose much greater challenges to the Saussurrean dictum. This resurgence reflects two complementary developments. First, there has been considerable systematic research into the phenomenon of synaesthesia. While initially deemed peculiar or aberrant, contemporary research demonstrates that cross-modal neurosensory associations of the sort first described for synaesthesia are about 100 times more common than previously expected (Simner et al. 2006). It has also shown that normal individuals often manifest many of the same cross-modal activity biases as synaesthetes (Ward et al. 2006; cf. Brang et al. 2011). Such findings have led some to propose that cross-modal connectivity is, in fact, a basic feature of neural organization and that synaesthetes simply manifest an exaggerated form of it (Spector and Maurer 2009).

Complementing this development has been a renewed and more sophisticated focus on the bio-cultural origins of language. This research program has specifically included experimental and modeling studies of the role of sound-symbolism in facilitating language acquisition and affecting the differential survival and learnability of word forms (e.g. Imai et al. 2008; Kirby et al. 2008; Monaghan et al. 2011).

Taken together, these two research developments have helped to illuminate a plausible mechanistic basis for sound-symbolism in the cross-modal neural connectivity that appears to be a common organizational feature of the nervous system (Rouw and Scholte 2007), and they have helped to establish a potentially important functional role of sound symbolism in language acquisition and transmission. However, this work still must deal with an outstanding paradox. On the one hand, the mechanistic and functional accounts help to explain the consistent biases observed in past experiments testing people's naïve expectations about the appropriate structure of artificial language material. On the other hand, the structural patterns implicated by people's naïve expectations seem not to be manifest in the words of real languages. More succinctly, if the biases are so consistent, why are they not manifest in the structure of real language?

A paradigm example of this paradox lies in the classic Takete-Maluma (or Bouba-Kiki) effect first described by Köhler (1947). This effect involves a bias in which people consistently map unfamiliar objects that are either jagged or rounded in form to nonsense words containing specific types of consonants or vowels (Maurer et al. 2006; Nielsen and Rendall 2011; Tarte and Barritt 1971; Tarte 1974; Westbury 2005). Across multiple studies, the matching biases are consistent and the effects are large. Given the consistency and magnitude of these effects, it is odd that there is so little tangible evidence for this effect in real languages.

We have proposed that a potential solution lies in the transparency of the task administered in many experiments. For example, experimental designs have often used a simultaneous presentation procedure in which two unfamiliar object images and two nonsense words are displayed at the same time. This procedure allows participants to make direct comparisons between the different object shapes and the different word types which might make the associations between them obvious. As a result, such experiments produce very high concordance rates, on the order of 90%, that might overstate the magnitude of the bias and thus the expectations about its prevalence in real language. In support of this proposal, we found that, when the experiment is modified only slightly to involve sequential presentation of single images, the same matching biases are observed but the magnitude of the effects are much smaller (~60%) (Nielsen and Rendall 2011; cf. Berlin 1994; Kovic et al. 2010; Parise and Pavani 2011). This finding helps to resolve the paradox by reconciling previous apparent discrepancies between the magnitude of the matching bias and the dearth of evidence for it in real language. What remains unclear is whether the bias that this work confirms is something that participants have inherently and therefore bring to the experimental task or that they construct in the task itself simply because the word-image associations are made obvious when both categories of stimuli are presented simultaneously.

To test this possibility, we designed an experiment to teach participants matching rules that were either congruent or incongruent with the consistent matching patterns documented previously. If participants have no pre-existing matching bias, then both sets of participants should perform equally well. However, if participants bring a natural bias to the task, then it should affect their performance, facilitating performance by participants in the congruent rule condition and interfering with performance by participants in the incongruent rule condition.

2. Methods

2.1. *Participants*

Participants were 48 undergraduate students (36 female, 12 male) between the ages of 18 and 48 who were enrolled in introductory psychology courses at the University of Lethbridge and received partial course credit for their participation.

2.2. *Word and image stimuli*

The image stimuli used in this experiment were created using a radially constrained mathematical formula which created pairs of curved and jagged image forms from the same set of randomly generated calculus points. The resulting image pairs were identical to one another except in the curvature of their lines. Two example image pairs are shown in Figure 1 and full details of the image generation technique are provided in Nielsen and Rendall (2011).

This image generation procedure allowed for the wholly objective production of a large number of image sets that were thereby not influenced by implicit experimenter biases in image selection.[1] A total of 46 pairs of images were created, 40 of which (80 total images) were randomly selected for use in the experimental, or test, trials and 6 of which (12 total images) were randomly selected for use in the demonstration trials.

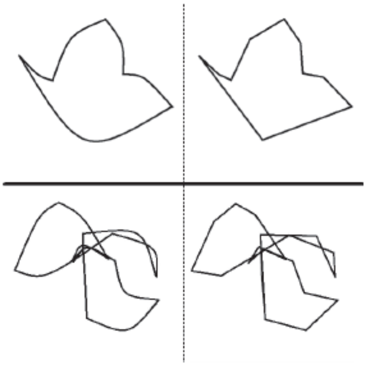

Figure 1. *Side-by-side comparisons of sample curved and jagged images generated with our randomized image construction method. Note that, within pairs, the curved (left) and jagged (right) image forms are only subtly different from each other.*

---

[1] Note that this has been a criticism of past work.

Word stimuli were created for this experiment using the plosive consonants /t/, /k/, and /p/ and the sonorant consonants /m/, /n/, and /l/ and the vowels /a/, /e/, and /u/. Words were created by combining consonants and vowels in c-v-c-v order, producing a set of four letter words that were two syllables long with the first letter capitalized. To ensure that all words were of the same syllabic length, the vowel /e/ was excluded as a possible terminal letter (to avoid words such as *pake*, which are four letters long and can be produced as a single syllable, *payk*). Words were also constructed of either entirely plosive consonants or entirely sonorant ones, with no mixing of the two consonant classes within words. This produced 54 possible permutations of each overall word form.

It is important to note that this selection of consonants and vowels minimized potential orthographic effects in which the visual form of the letters resembles the jagged or curved form of the object shapes. Such orthographic effects should not have been a factor here because all three sonorant consonants, which are predicted to be matched to curved object shapes, are actually relatively jagged in their capitalized forms (/L/, /M/, /N/) and one remains so in lower-case form (/l/). At the same time, one of the strident consonants, proposed to be matched to jagged object shapes, is patently curved in both upper- and lower-case forms (/P/, /p/). As a result, the mixed orthographic representation of the various consonants used in this experiment did not consistently align with either possible matching rule. At the same time, all three vowels (/a/, /e/, and /u/) are similarly curved or rounded in the lower-case forms used and would therefore introduce no systematic bias.

2.3. *Experimental design*

The experiment was conducted on computer via a graphical interface created using Runtime Revolution v. 2.8.1. Participants were split into two conditions and then shown a series of demonstration trials in which they were tasked with learning a rule for the association between unfamiliar images and nonsense words. In each demonstration trial, participants were presented with a single image with a nonsense word beneath it. After a brief delay, they were prompted with either a large green check-mark or a large red “X” below the word to inform them of whether the word-object pairing was correct or not. Participants were given no instructions on what characteristics of images and words were to be considered salient for matching. Each participant was shown twelve of these demonstration trials (six correct, six incorrect), three from each of the four possible pairings of images and word classes (three plosive words with jagged images, three plosive words with curved images, three sonorant words with jagged images, three sonorant words with curved images).

Participants in Condition 1 were presented with demonstration trials that were congruent with sound-symbolic biases observed in past studies: they were shown green check-marks for demonstration trials where curved images were paired with sonorant words and where jagged images were paired with plosive words. Participants in Condition 2 were presented with demonstration trials that were incongruent and ran counter to previous observed response patterns; they were shown green check-marks for demonstration trials where curved images were paired with plosive words and where jagged images were paired with sonorant words. To ensure that the learnability of the stimulus sets was not systematically biased between participants, all participants were exposed to the same stimulus sets in demonstration trials and in the same order, the only difference being the congruency of the rule they were taught.

Following the demonstration phase, participants received 80 experimental trials (40 of each plosive and sonorant words sampled randomly for each participant). In each trial, a single image was presented for two seconds with one of the nonsense words. Participants were then able to respond by using the keyboard to indicate whether or not the presented pairing matched the rule they had learned during the demonstration phase of the experiment. They responded by pressing the "Z" key on the keyboard for a correct match and the "/" key for an incorrect match. Order of presentation for images was randomized within and across subjects.

2.4. *Data analysis*

D′ values were calculated for each participant based on the rule they had learned in the demonstration phase of the experiment. D′ values for each condition were compared to chance using a one sample t-test and to each other using a two sample t-test.

3. Results

Participants in the congruent condition of the experiment responded correctly to the rule demonstrated for them at rates above chance ($t_{23} = 3.07$, $p < 0.01$), with an average d′ score of 0.17 corresponding to 53.3% correct. Participants in this condition also performed better than those in the incongruent condition ($t_{46} = 2.4$, $p < 0.01$). Participants in the incongruent condition performed at rates that were not different from chance ($t_{23} = 0.18$, $p = 0.43$), with an average d′ score of 0.006, corresponding to 50.4% correct. Measures of response criteria (C) for participants in the two conditions were similar and not significantly different (Congruent condition: mean C = 0.24; Incongruent condition: mean C = 0.19; $t_{46} = 0.415$, $p = 0.68$).

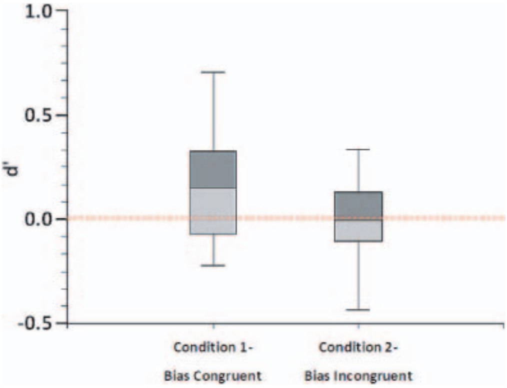

Figure 2. *d' performance by participants in Conditions 1 and 2 scored by their ability to follow the rule they learned in the demonstration phase of the experiment. Performance in Condition 1 was significantly better than chance ($p < 0.01$) and significantly different from performance in Condition 2, while performance in Condition 2 was not different from chance (see text for statistics).*

4. Discussion

Discrimination performance (d′) was low in both experimental conditions, suggesting poor discrimination, or learning. This outcome supports our previous findings (Nielsen and Rendall 2011) that the relative transparency of the experimental task in previous studies might have contributed to very high observed concordance rates. These, in turn, may have overestimated the magnitude of the matching bias and thereby inflated expectations about its prevalence in the structure of real languages. At the same time, d′ scores were influenced by experimental condition, with participants in the congruent condition performing modestly but significantly better than chance and better than participants in the incongruent condition who performed at chance levels. This pattern indicates a real but weak bias toward the rule embodied in the congruent condition that facilitated performance in this condition and interfered with performance with the opposite rule in the incongruent condition.

One possible qualification to the conclusion that the matching bias involved might be weaker than previously supposed centers on the variable jaggedness of our figures. Specifically, many of the curved images contained some acute angles which gave them an element of jaggedness. This stemmed from the fact that curved and jagged image pairs were created from exactly the same set of randomly generated calculus points which was a deliberate precaution undertaken to ensure general equivalence between curved and jagged image forms and thereby eliminate inadvertent experimenter biases in the choice of image materials—a criticism of previous studies. However, in some cases, this precaution necessarily resulted in intersecting lines that yielded acute interior angles in the final form of the curved version of each image pair (Figure 1). It is possible then that this ambiguity contributed to the overall weak performance observed. While this possibility cannot be eliminated definitively, it seems unlikely given that exactly the same image materials were used in a previous study (Nielsen and Rendall 2011) where participants performed at very high rates (~80%) similar to those reported in previous studies involving simultaneous (not sequential) image presentation.

Notwithstanding this possible qualification, the results of this experiment may help to reconcile the apparent paradox arising from past work which pointed to a very robust and strong matching bias for artificial words with little or no evidence of it in the structure of real languages. The solution is possibly two-pronged. First, the magnitude of the bias might be smaller than previously believed. Second, there are other factors that affect the likelihood that the bias will be manifest in fully developed lexicons. For example, Gasser (2004) has shown that, as lexicon size increases, arbitrariness in word forms becomes critical to avoid confusion among related words. This finding is buttressed by more direct attempts to model the process of lexicalization. Monaghan et al. (2011) have shown that structural regularities (non-arbitrariness) can be important to define word class membership. However, within word classes, arbitrary distinctions among items are more stable and easier to learn. Together, these findings suggest that the greatest functional value for sound symbolic regularities probably comes during the period of language acquisition prior to lexical crystallization. Consistent with this proposal, Imai et al. (2008) have shown that sound symbolic relationships in words facilitate noun learning for both English and Japanese children.

How the sound-symbolic bias demonstrated here and in previous experiments bears on language acquisition may depend on the source of the bias. Here, there is a potential distinction between a learned bias (that results from language exposure) and a learning bias (that predates and influences language exposure). Regarding the former, it might be argued that the weak bias observed in the congruent condition of our experiment and reported in previous experiments is merely incidental. It stems from the fact that adult speakers have considerable experience with their own language which embodies the structural regularities tested. However, this argument runs afoul of several important points, namely that the bias observed is specifically reported to be lacking in real languages (e.g. Newman 1933; but see Johnson 1967) and yet the bias has nevertheless been demonstrated not only for speakers from very disparate language groups (Huang et al. 1969; Nygaard et al. 2009) but also for small children with minimal language exposure (Maurer et al. 2006).

Hence, our results seem to provide more support for the second possibility, namely of a natural learning bias that facilitated performance in the congruent condition and interfered with it in the incongruent condition. From a functional standpoint, this bias might reflect broader affective-semantic relationships embodied in the physical structure of vocal signals. As reviewed elsewhere, these relationships include the way different voiced sounds (e.g. harsh, noisy and punctuate sounds versus smoother and more harmonic ones) have naturally different hedonic values because they are habitually associated with situations that have very different social and behavioral salience and consequences for listeners (Morton 1977; Owren and Rendall 1997, 2001). Such relationships are common in humans, both in the prosodic and paralinguistic dimensions of speech (Bachorowski and Owren 1995; Murray and Arnot 1993; Owren et al. 2005; reviewed in Rendall 2003) and also in a variety of nonlinguistic vocal signals, e.g. laughter, crying (Bachorowski et al. 2001; Protopapas and Eimas 1997). They are also manifest far more widely in the vocal signals of nonhuman primates and many other animal species (Rendall et al. 2009). As a result, there are some very broad relationships between the physical structure of voiced sounds and the natural affective salience they have for listeners. These vocal affective relationships might include, at least in a limited way, the various phonemes of language that thereby inherit some natural semantic potential.

This functional level account of the observed bias dovetails with recent research on cross-modal neural processing that provides a complementary mechanistic account of the integration of cross-modal perceptions and helps to explain why some forms are more common than others and thus potentially more salient to non-synaesthetes (Nikolie et al. 2007; Simner et al. 2006; Ward et al. 2006). Concurrently, language modeling research has demonstrated that a combination of arbitrariness and sound-symbolic relationships is ideal for language stabilization and transmission (Gasser 2004; Monaghan et al. 2011). Hence, by integrating such mechanistic and functional insights, it might be possible to determine the kinds of non-arbitrary, sound-symbolic relationships that should arise in languages and why.

References

Bachorowski, J.-A. & M. J. Owren. 1995. Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science* 6. 219–224.

Bachorowski, J.-A., M. J. Smoski & M. J. Owren. 2001. The acoustic features of human laughter. *Journal of the Acoustical Society of America* 110. 1581–1597.

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

Brang, D., R. Rouw, V. S. Ramachandran & S. Coulson. 2011. Similarly shaped letters evoke similar colors in grapheme-color synaesthesia. *Neuropsychologia* 49. 1355–1358.

Brown, R. W., A. H. Black & A. E. Horowitz. 1955. Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology* 50. 388–93.

Gasser, M. 2004. The origins of arbitrariness of language. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. 434–439.

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

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

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

Kirby, S., H. Cornish & K. Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105. 10681–10686.

Köhler, W. 1947. *Gestalt psychology.* New York, NY: Liveright.

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

Maurer, D., T. Pathman & C. J. Mondloach. 2006. The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science* 9. 316–322.

Monaghan, P., M. H. Chriastiansen & S. A. Fitneva. 2011. The Arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General* 140. 325–347.

Morton, E. S. 1977. On the occurrence and significance of motivation-structural rules in some birds and mammal sounds. American Naturalist 111. 855–869.

Murray, I. R. & J. L. Arnot. 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America* 93. 1097–1108.

Newman, S. S. 1933. Further experiments in phonetic symbolism. American Journal of *Psychology* 45. 53–75.

Newmeyer, F. J. 1992. Iconicity and generative grammar. *Language* 68. 756–796.

Nielsen, A. & D. Rendall. 2011. The Sound of round: Evaluating the role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology* 65. 115–124.

Nikolie, D., P. Lichti & W. Singer. 2007. Color opponency in synaesthetic experience. *Psychological Science* 18. 481–486.

Nygaard, L. C., A. E. Cook & L. L. Namy. 2009. Sound to meaning correspondences facilitate word learning. *Cognition* 112. 181–186.

Owren, M. J. & D. Rendall. 1997. An affect-conditioning model of nonhuman primate vocalizations. In D. W. Owings, M. D. Beecher & N. S. Thompson (eds.), *Perspectives in ethology, Volume 12: Communication*, 299–346. New York, NY: Plenum Press.

Owren, M. J. & D. Rendall. 2001. Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology* 10. 58–71.

Owren, M. J., D. Rendall & J.-A. Bachorowski. 2005. Conscious and unconscious emotion in nonlinguistic vocal communication. In L. F. Barrett, P. Niedenthal & P. Winkielman (eds.), *Emotion and consciousness*, 185–204. New York, NY: Guilford Publications.

Parise, C. V. & F. Pavani. 2011. Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research* 214. 373–380.

Protopapas, A. & P. D. Eimas. 1997. Perceptual differences in infant cries revealed by modifications of acoustic features. *Journal of the Acoustical Society of America* 102. 3723–3734.

Rendall, D. 2003. Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *Journal of the Acoustical Society of America* 113. 3390–3402.

Rendall, D., M. J. Owren & M. J. Ryan. 2009. What do animal signals mean? *Animal Behaviour* 78. 233–240.

Rouw, R. & S. Scholte. 2007. Increased structural connectivity in grapheme—color synesthesia. *Nature Neuroscience* 10. 792–797.

Saussure, F. 1983. *Course in general linguistics*. Translated by Roy Harris. La Salle, IL: Open Court.

Simner, J., C. Mulvenna, N. Sagiv, N., E. Tsakanikos, S. A. Witherby, C. Fraser, K. Scott & J. Ward. 2006. Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception* 35. 1024–1033.

Spector, F. & D. Maurer. 2009. Synesthesia: A new approach to understanding the development of perception. *Developmental Psychology* 45. 175–189.

Tarte, R. D. & L. S. Barritt. 1971. Phonetic symbolism in adult native speakers of English: Three studies. *Language and Speech* 14. 158.

Tarte, R. D. 1974. Phonetic symbolism in adult native speakers of Czech. *Language and Speech* 17. 87–94.

Ward, J., B. Huckstep & E. Tsakanikos. 2006. Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex* 42. 264–280.

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