Where did Words Come from?

A Linking Theory of Sound Symbolism and Natural Language Evolution

Jamie Reilly[1], David Biun[1], H. Wind Cowles[2], & Jonathan E. Peelle[3]

[1]Departments of Psychology and Communicative Disorders
University of Florida, Gainesville, Florida

[2]Department of Linguistics
University of Florida, Gainesville, Florida

[3]Cognition and Brain Sciences Unit
Medical Research Council, Cambridge, UK

Address Correspondence to:

Jamie Reilly, Ph.D.
University of Florida
Departments of Neurology & Communicative Disorders
P.O. Box 100174
Gainesville, Florida 32610
jjreilly@phhp.ufl.edu
Fax: (352) 273-6545
Tel: (352) 273-6161

ABSTRACT

Where did words come from? The traditional view is that the relation between the sound of a word and its meaning is arbitrary. An alternative hypothesis, known as sound symbolism, holds that form-meaning correspondence is systematic. Numerous examples of sound symbolism exist across natural language phyla. Moreover, cross-linguistic similarities suggest that sound symbolism represents a language universal. For example, many unrelated languages affix an "ee" sound to words in order to emphasize size distinctions or express affection (e.g., look at the *teeny weeny baby*); other such phonetic universals are evident for object mass, color, brightness, and aggression. We hypothesize that sound symbolism reflects sensitivity to an ecological law (i.e., Hooke's Law) governing an inverse relation between object mass and acoustic resonance. In two experiments healthy adults showed high agreement in matching pure tones to color swatches and nonwords to novel objects as linear functions of frequency and luminance. These results support a degree of non-arbitrariness in integrating visual and auditory information. We discuss implications for sound symbolism as a factor underlying language evolution.

INTRODUCTION

Numerous theories exist regarding the genesis of language; however, there remains no definitive answer to the enduring question of where words came from. Contemporary linguistic theory invokes arbitrary symbolism as an explanation.[1,2] That is, words represent arbitrary strings of sounds assigned ad hoc to concepts with few constraints other than language-specific phonotactic boundaries. The relative dominance of this arbitrariness assumption is evident in virtually all modern models of language acquisition and processing.[3] However, there exists an alternative hypothesis known as *sound symbolism* that invokes non-arbitrariness as an account of the ways that cultures assign names to concepts. Sound symbolism occurs when content-free units such as phonemes convey meaning in certain linguistic contexts.[4,5] In English, perhaps the most familiar example of sound symbolism is onomatopoeia (e.g., *buzz, splash*). Yet, sound symbolism also exists in a variety of other contexts. For example, many derivationally unrelated /fl-/ initial words share a common theme of linear motion (e.g., flee, flow, fly, flutter), /kr-/ initial words often denote sudden impact (e.g., crack, crash, crush, cram), and /gl-/ initial words often denote diffusion of light (e.g., glow, gleam, glitter, glare).[4] In addition, sound symbolism is pervasive in ethnozoological nomenclature (animal names) and in brand name advertising.[4,6,7]

Ethnolinguists have demonstrated that sound symbolism is pervasive within strata of most if not all of the world's natural languages.[4,6,8] The extent to which sound symbolism acts as a driving mechanism for the assignment of new words to concepts does, however, differ across linguistic boundaries. Mayan, Navajo, and Korean, for example, exhibit higher incidence of sound symbolism than English.[9,10] Perhaps the most compelling property of sound symbolism is that it may represent a rare case of a language universal.[11,12] Many derivationally unrelated languages, for example, affix an "ee" sound to words in order to express affection or to emphasize size distinctions.[6,13] Cross-linguistic phonetic similarities also exist for words that describe physical size, predation/aggression, angularity, mass, and color.[6,10-15]

Debate regarding correspondence between word form and meaning is traceable to Plato's Cratylus dialogue (circa 360BC).[16] John Locke (1685), in *An Essay on Human Understanding,* argued against sound symbolism as a plausible account of language.[17] Locke reasoned that if sound symbolism were indeed the true source of new words, there would exist only one human language given the assumption that human visual perception is invariant. French linguist, Ferdinand de Saussure, further undermined the plausibility of sound symbolism early in the twentieth century, demonstrating that sound-meaning correspondence was indeed statistically anomalous in natural language.[2] Today interest in sound symbolism has seen a revival. This is likely due to a number of factors, including the ubiquitous appearance of sound symbolism in the worlds' languages, its emergent role in word learning, and new insights into listeners' sensitivity to aspects of word form as cues for processing word meaning and grammatical class.[18,19]

Perceptual Correlates of Sound Symbolism

One psychophysical explanation for sound symbolism in language is that the phenomenon reflects a direct ecological mapping between perceptual information.[11,14] This account, known as the *frequency code hypothesis,* predicts that humans show sensitivity to non-arbitrariness in the mapping of acoustic to visual information. This sensory correspondence is captured by Hooke's Law, which states that there is an inverse relation between object mass and resonant frequency. In a broad context, Hooke's Law predicts that an object with low relative mass (e.g., a bumblebee) will resonate with a high frequency, whereas an object with higher relative mass (e.g., an elephant) will resonate at a lower acoustic frequency. In nature, this translates to non-arbitrary relation between vision and audition such that bees tend to buzz and elephants tend to trumpet. Proponents of the frequency code hypothesis have argued that sound symbolism occurs when people extrapolate this mass-acoustic resonance relation onto words and object properties that covary with mass (e.g., size and predation).[14] Phonological factors that elicit sound symbolism include vowel height ("ee" vs "oh") and consonant hardness (e.g., "keek" vs. "leel").[6,11,14,20] Visual properties that elicit sound symbolism include sharpness, size, color, shape, brightness, and predation.[4]

Figure 1 illustrates sound symbolism for two novel objects. One of these objects is called a *Kiku* while the other is a *Volo*. Which is the *Kiku*?

-figure 1-

The majority of respondents in the experiments to follow claimed that the small, yellow, spiked animal was a *Kiku.* In addition, name agreement was strong despite the lack of any meaningful link between these objects and their corresponding names.

Köhler (1949) was among the first to demonstrate that participants reliably matched jagged line drawings to written nonwords composed of plosive consonants (e.g., *taketu*) while similarly matching rounded line drawings to continuant nonwords (e.g., *baluma*).[21] Westbury (2005) recently demonstrated that this matching effect also extended to speed of word recognition when written words appeared within visual frames that were themselves either spiky/jagged or rounded. This finding has lead to the question of whether sound symbolism is a purely visual phenomenon that reflects matching of letter shape to corresponding visual object properties such as angularity.[5] We know of no studies to date that have avoided this potential confound using a pure auditory-to-visual crossmodal matching format. We do so here, hypothesizing the following:

a) Healthy adults are sensitive to a non-arbitrary perceptual relationship between visual and acoustic frequency detail (i.e., Hooke's Law).

b) Sensitivity to Hooke's Law is manifested at both the sensory level (colors and tones) and at a linguistic level in mapping novel word forms onto novel objects.

c) Hooke's Law is the perceptual basis for sound symbolism.

METHODS SUMMARY

30 healthy young adults made perceptual matching judgments for pairs of sequentially presented stimuli. In the first experiment, participants first viewed a color swatch followed by auditory presentation of a pure tone. Participants keyed "yes" or "no" to the question, "Are these a good match?" We factorially varied hue and luminance (bright or dark) of the color swatch crossed with frequency of the pure tone (e.g., 200 Hz or 6000 Hz). In the second experiment, we created a series of novel objects and pseudoanimals (see figure 1) and phonological distinctiveness of nonwords (e.g., keek vs. lole). Participants made the same matching judgments for visual images paired sequentially with aurally presented nonwords.

RESULTS

Color Swatch to Pure Tone Matching

Table 1 and figure 3 illustrate robust agreement in matching high wavelength colors with low frequency tones extending linearly to short wavelength colors paired with high frequency tones. For example, 27 of 30 participants endorsed the match of a 190 Hz pure tone with a dark blue color swatch (binomial probability $p<.000003$). There was a 3-way interaction between pure tone pitch, color swatch hue, and color swatch brightness [$F(2,58)=5.86$, $p=.01$, $\eta^2=.17$] and 2-way interactions between pure tone pitch and color swatch hue [$F(2,58)=8.92$, $p<.001$, $\eta^2=.24$], tone frequency and color brightness [$F(2,58)=108.89$, $p<.001$, $\eta^2=.79$], and hue and color brightness [$F(1,29)=31.61$,

$p<.001$, $\eta^2=.52$]. In addition, participants showed main effects of pure tone frequency [$F(2,58)=22.58$, $p<.001$, $\eta^2=.27$] and color brightness [$F(1,29)=51.13$, $p<.001$, $\eta^2=.64$].

-Table 1, Figure 2-

Novel Object to Nonword Crossmodal Matching

Participants showed strong effects of sound symbolism in their endorsements of novel object-pseudoword pairings (e.g., 27 of 30 endorsed the match of the pseudoword, “faichee” with the yellow creature in figure 1). In addition, cross-modal pairings were completed with a rapid average reaction time of 580ms (s.d.=230ms). Results appear in table 2. There was a significant 3-way interaction between nonword phonology, object hue, and object animacy [$F(1,29)=14.20$, $p=.001$, $\eta^2=.33$] and a 2-way interaction between nonword phonology and object hue [$F(1,29)=66.54$, $p<.001$, $\eta^2=.70$]. In addition there were significant main effects of object hue [$F(1,29)=30.40$, $p=<.001$, $\eta^2=.51$], object animacy [$F(1,29)=26.03$, $p<.001$, $\eta^2=.47$], and a marginal main effect of pseudoword phonology [$F(1,29)=3.92$, $p=.057$, $\eta^2=.12$]. This interaction was such that participants tended to spontaneously map high-stop consonant pseudowords (e.g., kiku) onto bright colored novel objects and low-continuant pseudowords (e.g., volo) onto dark, rounded objects.

-Table 2-

GENERAL DISCUSSION

Participants were systematic in their cross-modal perceptual judgments. For colors and tones, this effect (i.e., pitch-color isomorphism[22]) was apparent for two perceptual properties of the visual stimuli: hue and luminance. Moreover, participants showed similar sensitivity in matching novel objects varied by color and form to pseudowords varied by their acoustic phonetic structures. Similarities across both the sensory and linguistic experimental manipulations suggest a common perceptual substrate. Hooke's Law and the frequency code hypothesis of sound symbolism offer a parsimonious account of this perceptual phenomenon.

The frequency code hypothesis predicts that acoustic features such as frequency and amplitude "feel" like particular visual properties such as color and size. Ramachandran and colleagues have argued that this type of cross-modal perception is a form of color-sound synaesthesia.[23] Ramachandran and Hubbard (2001) raised the additional possibility that synaesthesia played a substantial role in language evolution and that its effects persist in contemporary language. English, for example, is rife with synaesthetic metaphor (e.g., I sure feel *blue* today; Paris Hilton is *hot.*). Similarities in human crossmodal perception may also prove useful for explaining the universality of sound symbolism. As governed by Hooke's Law, an object with high relative mass (e.g., a large rock) makes a characteristic low frequency sound when dropped or struck. This mass-resonance relation is invariant, and ecological invariance might explain why so many derivationally unrelated languages show similar phonetic features in making size and mass distinctions.[10,12,24]

Where did words come from?

The ephemeral nature of non-written language makes its origins particularly difficult to study; we simply do not have the equivalent of a fossil record to consult. Nonetheless, there are many reasons to believe that non-arbitrariness may contribute to spoken language evolution. Supporting evidence for this idea is derived from the evolution of signed languages and home sign systems.[25] The recently formed Nicaraguan Sign Language, for example, has shifted over the last three generations of speakers from iconic (i.e., gestures mimic actual objects or events) to arbitrarily symbolic, mirroring a similar shift in well-established signed languages.[26] The spoken language analog of an iconic signed language would entail mapping acoustic properties of vocal sounds directly onto physical properties of objects. In this way, sound symbolism reflects a direct link between ecology and language.

LeCron Foster (1978) proposed an expansive theory of sound symbolism's role in the evolution of protolanguage, modeling sound-meaning correspondence into a comprehensive lexicon. LeCron Foster ultimately concluded that evolutionary pressures on linguistic diversity forced a schism between word form and meaning that occurred relatively recently in human evolution, 50,000 to 75,000 years ago.[8] As lexical and semantic demands increased, language clearly evolved systems of arbitrary assignment of names to objects. Without the flexibility of an arbitrary phonological system, human communication would be limited to only the most salient concrete words. In contrast, arbitrariness affords the possibility of a potentially infinite lexicon.

Despite the adaptive evolutionary advantage of linguistic arbitrariness, there are many reasons to believe that sound symbolism is also contributes to language processing. Recent work has demonstrated that listeners and readers are sensitive to non-arbitrary aspects of language (e.g., word length, syllable stress placement) and that they exploit phonological regularities to speed the efficiency of language processing.[19,27] Sound symbolism is also likely to play a role in early language development as the exaggeration of pitch contours (e.g., look at the BIG BAD WOLF) is a common cross-linguistic property of infant-directed speech.[28]

One potentially useful way of viewing language is that of co-existence between arbitrariness and sound symbolism. Here we have demonstrated the cross-modal perceptual basis for sound symbolism in terms of a physical law that likely explains the universality of the effect. The next logical step is to examine the sensitivity to this perceptual phenomenon in non-native English speakers.

METHOD

Participants made yes/no perceptual matching judgments for pairs of sequentially presented stimuli. Presentation was standardized using E-Prime 1.2 stimulus delivery software.[29] Visual stimuli appeared as 250 × 250 pixel bitmap images centered on a 19" monitor; auditory stimuli were presented in wavefile format over noise canceling headphones.

Participants

Participants included 30 young adults (mean age=19.8 years) recruited from the University of Florida. All were monolingual native speakers of American English. We screened for color blindness by having participants name 10 color swatches. We screened for hearing impairment using a standard tone detection paradigm for tones pulsed binaurally at 25 dB SPL at frequencies 500Hz, 1kHz, and 4kHz.[30]

Experiment 1: Color-Tone Matching Procedure

Participants viewed a color swatch followed by auditory presentation of a pure tone. Participants then made a yes/no judgment endorsing the fit of the tone-color match. We employed a 2 × 2 × 3 factorial design, crossing color swatches varied by wavelength (short/long) and brightness (bright/dark) with tones varied by frequency (low-medium-high). This design contained 10 items per cell for a total of 120 tone-color pairings. Tone-color pairings were pseudo-randomized. For example, the cell that crossed short wavelength, bright colors with low frequency pure tones consisted of 12 color swatches that spanned the color family of "blue" paired with pure tones that spanned the 100 to 260 Hz frequency range. Fixed color-tone pairs were then presented in randomized order.

Color Characteristics

The visible light spectrum ranges from approximately 400-650nm, ordered in increasing wavelength from: violet → indigo→ blue → green → yellow → orange → red. We operationally defined "short" wavelength hues as purple and blues, whereas "long" wavelength hues were variations of red and yellow. We created fifteen different color swatches spanning each of these color families in approximately 5nm increments (e.g., 410nm, 415nm, etc.) scaled to the 360° color wheel of Adobe Photoshop. Hues were also manipulated in terms of luminance (bright versus dark). We operationally defined "bright" as 100% brightness, whereas "dark" was 40% brightness via the Adobe Photoshop slider bar. Thus, the identical short wavelength hue (e.g., violet, 410nm) was presented in two conditions: bright and dark.

Tone Characteristics

Pure tones were blocked by three frequency levels (low-medium-high). The low frequency range was from 100-250 Hz in 10 Hz increments (100, 110, etc.). The middle frequency range was from 1000-1500 Hz in 50 Hz increments. The high frequency range was from 4500-5600 Hz in 100 Hz increments. All tones were identical duration (810 ms) and matched in intensity to their root mean square amplitude.

Experiment 2: Nonword-Novel Object Matching Procedure

Participants made matching judgments (n=96) for pronounceable but meaningless pseudowords paired with novel objects. Visual stimuli included bitmap images of a combination of 48 unfamiliar items obtained from online photo and microscopy libraries or created using a 3-dimensional computer graphics program (Cosmic Blobs: http://www.cosmicblobs.com). Stimuli included 24 inanimate blobs and 24 animate pseudoanimals. Animate items include cartoon images of deep sea creatures, microorganisms, and pseudoanimals. Inanimate blobs included cells and textured geometric shapes all distorted to the point of unfamiliarity (e.g., a warped and inverted pepperoni pizza). Half the stimuli were depicted as small, had predominantly sharp edges, and were composed of bright, long wavelength hues (e.g., yellow). The remainder were large and were composed of dark, short wavelength hues (e.g., purple).

We created two overarching phonological categories of nonwords. “High-Stop” nonwords were composed of voiceless stop consonants (e.g., p, k, t) with high front vowels (e.g., “ee” as in keet; “i” as in kitch). “Low-Continuant” nonwords (n=48) had low back vowels (e.g., “ah” as in hall, “o” as in boat) and voiced consonants such as nasals (e.g., m, n, ng); voiceless affricates (e.g., j), voiced alveolars (e.g., d) and liquids (e.g., r,l). Half the stimuli were monosyllabic with a consonant-vowel-consonant (CVC) phonological syllable structure (e.g., keek); the remainder were disyllabic with a CVCV structure (e.g., volo). We digitally recorded an adult female, who real aloud the 96 nonwords. We then spliced the recording into 96 individual wavefiles.

Each blob and pseudoanimal was paired pseudo-randomly on two different trials with a *High-Stop* and a *Low-Continuant* nonword. That is, participants saw each blob and psuedoanimal twice throughout the experiment. However, two variables differed upon repeated presentation. Visual orientation of the blob or pseudoanimal was left-right flipped, and the phonological “category” of the pseudoword differed. For example, participants verified the match of a yellow, left-facing pseudoanimal with the pseudoword, “kiku”. On a different trial, the participant verified the match for the identical pseudoanimal, this time right-facing and paired with the pseudoword, “volo”.

REFERENCES

[1] Christiansen, M. H. and Kirby, S., *Trends in Cognitive Sciences* **7** (7), 300 (2003).

[2] Saussure, F. d., *Cours de linguistique generale (1907).* (Philosophical Library, New York, 1916).

[3] Levelt, W. J. M., Roelofs, A., and Meyer, A. S., *Behavioral and Brain Sciences* **22** (1), 1 (1999).

[4] Parault, S. J. and Schwanenflugel, P. J., *Journal of Psycholinguistic Research* **35** (4), 329 (2006).

[5] Westbury, C., *Brain and Language* **93** (1), 10 (2005).

[6] Fischer-Jorgensen, E., *Studies in Linguistics* **32**, 80 (1978).

[7] Berlin, B., in *Sound symbolism.* , edited by L. Hinton, J. Nichols, and J. J. Ohala (Cambridge University Press, Melbourne, AU, 1994).

[8] LeCron Foster, M., in *Human evolution biosocial perepctives*, edited by S. L. Washburn and E. R. McCown (Benjamin Cummins Publishing Company, Merlo Park, CA, 1978), pp. 77.

[9] Brown, R. W., Black, A. H., and Horowitz, A. E., *The Journal of Abnormal and Social Psychology* **50** (3), 388 (1955).

[10] Hinton, L., Nichols, J., and Ohala, J. J. eds., *Sound symbolism* (Cambridge University Press., Melbourne, AU, 1994).

[11] Nuckolls, J. B., *Annual Review of Anthropology* **28**, 225 (1999).

[12] Taylor, I. K. and Taylor, M. M., *Canadian Journal of Psychology/Revue Canadienne de Psychologie* **16** (4), 344 (1962).

[13] Ladefoged, P. and Broadbent, D., *Journal of the Acoustical Society of America* **29**, 98 (1957).

[14] Ohala, J., in *Sound symbolism*, edited by L. Hinton, J. Nichols, and J. J. E. Ohala (Cambridge University Press., Melbourne, AU, 1994).

[15] Sapir, E., *Journal of Experimental Psychology* **12** (3), 225 (1929).

[16] Plato, (circa 360 BC).

[17] Locke, J., (1685).

[18] Kelly, M. H., *Psychological Review* **99** (2), 349 (1992); Saffran, J. R., *Current Directions in Psychological Science* **12** (4), 110 (2003).

[19] Reilly, J. and Kean, J., *Cognitive Science* **31** (1), 1 (2007).

[20] Maurer, D., Pathman, T., and Mondloch, C. J., *Developmental Science* **9** (3), 316 (2006).

[21] Kohler, W., *Gestalt Psychology. 2nd ed.* (Liveright, Oxford, England, 1947).

[22] de Thornley Head, P., *Cortex* **42** (2), 164 (2006); Ward, J., Huckstep, B., and Tsakanikos, E., *Cortex* **42** (2), 264 (2006).

[23] Ramachandran, V. S. and Hubbard, E. M., *Journal of Consciousness Studies* **8** (12), 3 (2001); Ramachandran, V. S. and Hubbard, E. M., *van Hemmen, J. Leo* **23**, 432 (2006).

[24] Taylor, I. K. and Taylor, M. M., *Psychological Bulletin* **64** (6), 413 (1965).

[25] Botha, R., *Language and Communication* **27**, 41 (2007).

[26] Senghas, A., Kita, S., and Azrek, A., *Science* **305** (5691), 1779 (2004).

[27] Kelly, M. H., Morris, J., and Verrekia, L., *Memory & Cognition* **26** (4), 822 (1998); Saffran, J. R., *Cognition* **81** (2), 149 (2001); Saffran, J. R. and Thiessen, E. D., *Developmental Psychology* **39** (3), 484 (2003).

[28] Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., and Cassidy, K. W., *Journal of Child Language* **16** (1), 55 (1989).

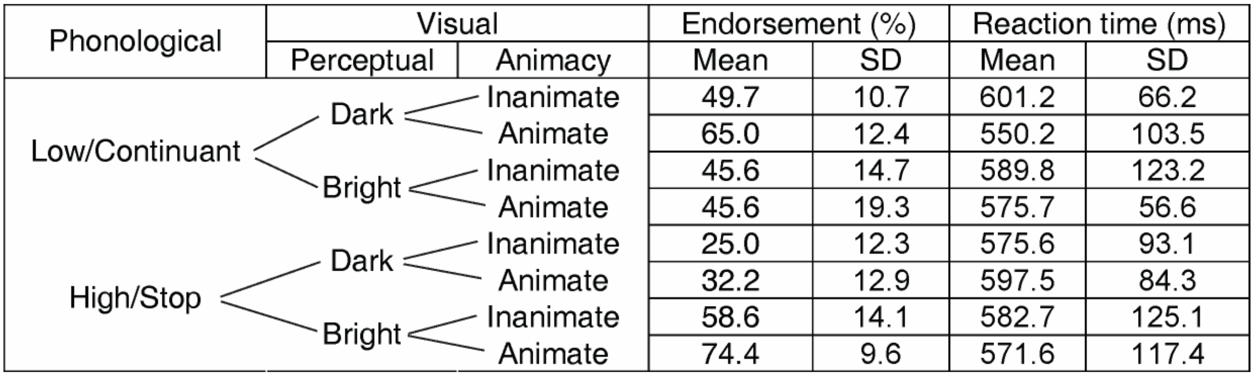
[29] Schneider, W., Eschman, A., and Zuccolotto, A., *E-Prime User’s Guide*. (Psychology Software Tools Inc., Pittsburgh, PA, 2002).

[30] Reilly, J., Troiani, V., Grossman, M., and Wingfield, A., *Behavior Research Methods* **39** (3), 667 (2007).

Table 1. Color-Tone Cross-Modal Matching Results

Auditory	Visual		Endorsement (%)		Reaction Time (ms)	
Pure Tone	Hue	Brightness	Mean	SD	Mean	SD
Low	Short	Bright	84.0	4.9	512.0	81.1
		Dark	33.7	16.6	608.0	169.7
	Long	Bright	80.0	10.9	468.2	87.2
		Dark	25.3	11.2	517.5	101.2
Medium	Short	Bright	31.7	10.6	566.6	103.9
		Dark	69.7	11.0	558.0	62.4
	Long	Bright	22.7	13.6	534.3	51.6
		Dark	74.7	6.5	551.8	63.9
High	Short	Bright	6.3	4.6	390.5	64.2
		Dark	43.0	6.7	507.0	82.5
	Long	Bright	5.3	5.9	428.1	119.6
		Dark	71.3	9.1	497.9	81.1

Table 2. Blob-Pseudoword Cross-Modal Matching Results



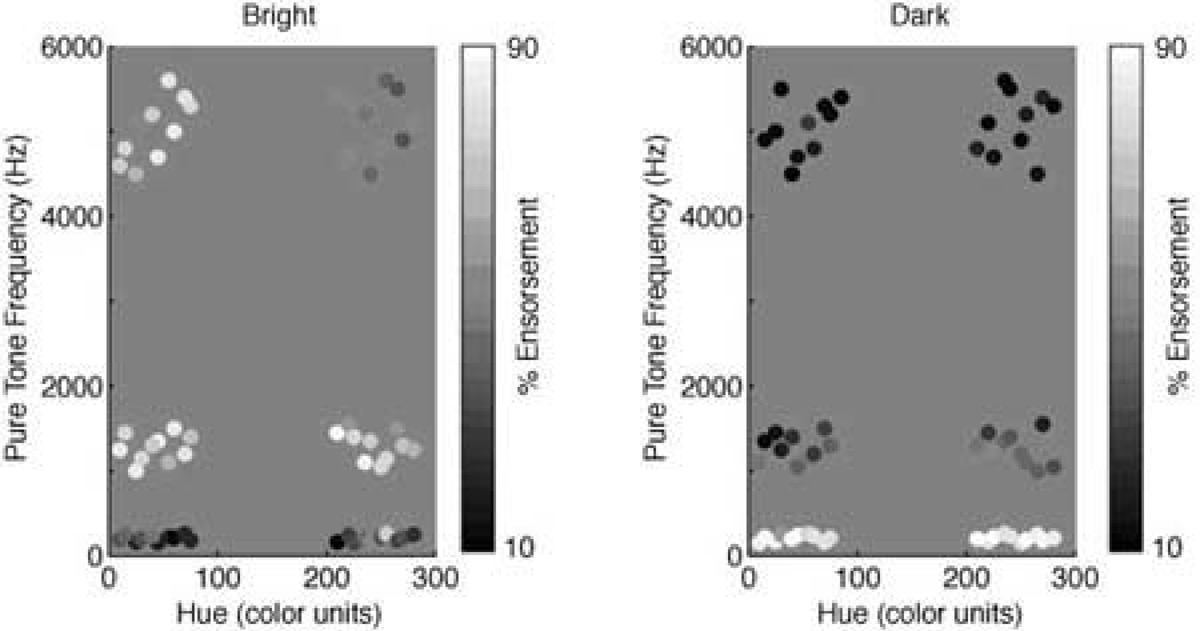
Phonological	Visual		Endorsement (%)		Reaction time (ms)	
	Perceptual	Animacy	Mean	SD	Mean	SD
Low/Continuant	Dark	Inanimate	49.7	10.7	601.2	66.2
		Animate	65.0	12.4	550.2	103.5
	Bright	Inanimate	45.6	14.7	589.8	123.2
		Animate	45.6	19.3	575.7	56.6
High/Stop	Dark	Inanimate	25.0	12.3	575.6	93.1
		Animate	32.2	12.9	597.5	84.3
	Bright	Inanimate	58.6	14.1	582.7	125.1
		Animate	74.4	9.6	571.6	117.4

Figure Titles & Captions

Figure 1. Name Assignment for Two Novel Animals

Figure 2. Endorsement for Color-Tone Crossmodal Matching

Note: Participants showed stronger agreement as a function of luminance/brightness over hue. Panel (a) reflects color/tone pairings with bright luminance colors, (b) color/tone pairings with dark luminance colors. Color units reflect an approximate linear mapping of the visible color spectrum to a 0-360 range (i.e., color wheel). Brighter points indicate higher levels of subject endorsement on a forced-choice pairing of color and tone.


Bright
Dark
Pure Tone Frequency (Hz)
6000
4000
2000
0
Hue (color units)
0
100
200
300
% Ensorsement
90
10