

Affective Arousal Links Sound to Meaning



Arash Aryani^{1,2} , Erin S. Isbilen³, and Morten H. Christiansen^{3,4,5} 

¹Department of Education and Psychology, Freie Universität Berlin; ²Center for Cognitive Neuroscience Berlin, Freie Universität Berlin; ³Department of Psychology, Cornell University; ⁴Interacting Minds Centre, Aarhus University; and ⁵School of Communication and Culture, Aarhus University

Psychological Science
 2020, Vol. 31(8) 978–986
 © The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/0956797620927967
www.psychologicalscience.org/PS



Abstract

Prior investigations have demonstrated that people tend to link pseudowords such as *bouba* to rounded shapes and *kiki* to spiky shapes, but the cognitive processes underlying this matching bias have remained controversial. Here, we present three experiments underscoring the fundamental role of emotional mediation in this sound–shape mapping. Using stimuli from key previous studies, we found that *kiki*-like pseudowords and spiky shapes, compared with *bouba*-like pseudowords and rounded shapes, consistently elicit higher levels of affective arousal, which we assessed through both subjective ratings (Experiment 1, $N = 52$) and acoustic models implemented on the basis of pseudoword material (Experiment 2, $N = 70$). Crucially, the mediating effect of arousal generalizes to novel pseudowords (Experiment 3, $N = 64$, which was preregistered). These findings highlight the role that human emotion may play in language development and evolution by grounding associations between abstract concepts (e.g., shapes) and linguistic signs (e.g., words) in the affective system.

Keywords

bouba-kiki effect, arousal, emotional mediation hypothesis, language evolution, affective iconicity, open data, open materials, preregistered

Received 7/12/19; Revision accepted 4/4/20

For most words, the relationship between sound and meaning appears arbitrary: The sound of a word does not typically tell us what it means (De Saussure, 1916). A growing body of work, however, has shown that the sounds of words can carry subtle cues about what they refer to (for reviews, see Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Perniss, Thompson, & Vigliocco, 2010). One of the first studies to demonstrate the existence of nonarbitrary mappings between sound and meaning was conducted by Köhler (1929/1947), in which participants displayed a reliable preference to match the pseudoword *takete* with a spiky angular shape and the pseudoword *maluma* with a curvy rounded shape (see Fig. 1, top). Using different stimuli pairs, numerous studies have extended this finding, showing the existence of similar preferences across speakers from diverse linguistic backgrounds (e.g., Bremner et al., 2013; Ramachandran & Hubbard, 2001) as well as in 4-month-old infants (Ozturk, Krehm, & Vouloumanos, 2013). Among these studies, the stimuli

pairs *bouba* and *kiki* used by Ramachandran and Hubbard (2001) have become so prominent that this sound–shape preference is now often referred to as the “*bouba-kiki* effect” (see Fig. 1, middle).

However, nearly a century after Köhler’s (1929/1947) initial study, and despite the increasing number of studies investigating the *bouba-kiki* effect, little is known about the cognitive processes underlying this phenomenon. Previous investigations into the source of this effect have yielded a variety of explanations, including orthographical influences (i.e., matching on the basis of aligning letter curvature and shape roundedness; Cuskley, Simner, & Kirby, 2017), the co-occurrence of certain shapes with certain sounds in the natural environment (Spence, 2011), and the possible “cross-wiring”

Corresponding Author:

Arash Aryani, Freie Universität Berlin, Department of Education and Psychology, Habelschwerdter Allee 45, D-14195 Berlin, Germany
 E-mail: Arash.Aryani@fu-berlin.de

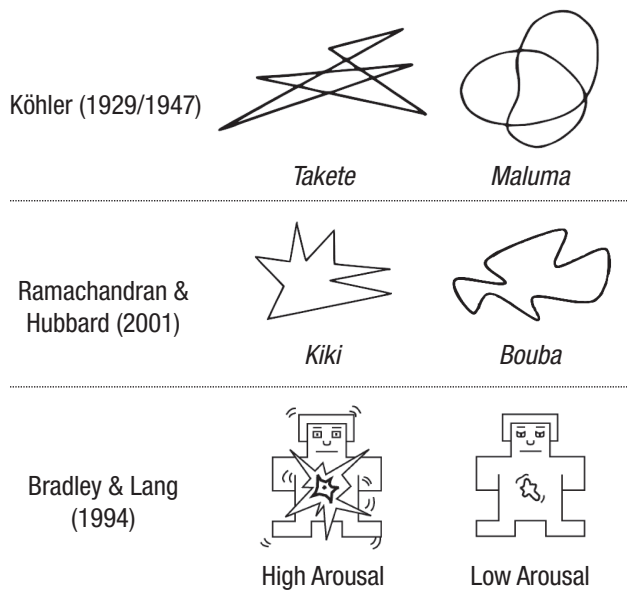


Fig. 1. Examples of mapping between shapes and other modalities. The spiky and rounded shapes used in the seminal studies by Köhler (1929/1947) and Ramachandran and Hubbard (2001) are shown in the top and middle rows, respectively. In a nonverbal method for assessing arousal, which was first introduced by Bradley and Lang (1994), the two manikins (bottom row) contain similar spiky and rounded shapes to represent the notion of high arousal versus low arousal.

between auditory and visual brain maps (Ramachandran & Hubbard, 2001).

Given that the *bouba-kiki* effect shows up early in development (Maurer, Pathman, & Mondloch, 2006; Ozturk et al., 2013), exists in radically different cultures (Bremner et al., 2013), is implicit (Westbury, 2005; but see Westbury, 2018), and occurs even prior to conscious awareness of the visual stimuli (Heyman, Maerten, Vankrunkelsven, Voorspoels, & Moors, 2019; Hung, Styles, & Hsieh, 2017), it is plausible that the mechanism underlying this phenomenon may be fundamental to human experience. Here, we propose that emotional congruence—the similarity in the arousal elicited by auditory and visual stimuli—may mediate the association between shapes and words, with *kiki*-like words¹ and spiky shapes invoking a higher level of arousal than *bouba*-like words and rounded shapes.

Emotion has been implicated in orchestrating and regulating stimulus–response mappings. Accordingly, emotional states need to connect with different types of stimuli across different modalities, suggesting a possible role for emotion in cross-modal associations. The mediating role of affective meaning has already been observed in cross-sensory correspondences: among auditory, visual, and tactile stimuli (Marks, 1975; Walker, Walker, & Francis, 2012); between music and color (Isbilen & Krumhansl, 2016); and even between more abstract concepts such as vowels and size (Auracher,

2017; Hoshi, Kwon, Akita, & Auracher, 2019). Furthermore, when participants are asked to rate the meaning of concepts using scales between bipolar adjectives, such as rounded–angular and soft–hard, most of the variance in ratings could be accounted for by the affective dimensions of valence and arousal (Osgood, 1952).

Within existing affective dimensions (Osgood, 1952), arousal seems to be the most well-suited candidate for mediating the *bouba-kiki* effect, given its dominant role in vocal communication and iconic mappings (Aryani, Conrad, Schmidtke, & Jacobs, 2018). At the auditory level of processing, the group of *kiki*-like words used in previous studies has often involved voiceless plosives (e.g., /k/ and /t/) and short vowels (e.g., /i/), which make words feel more harsh and arousing, as shown at both the behavioral-processing level (Aryani, Conrad, et al., 2018) and neural-processing level (Aryani, Hsu, & Jacobs, 2018, 2019). Similarly, noisy and punctuated sounds as well as sounds with sharp (abrupt) onsets (e.g., *kiki*, *takete*) are generally associated with high arousal, hostility, and aggression (Nielsen & Rendall, 2011). At the level of visual processing, people generally associate sharp-angled visual objects with negative valence and high arousal (Bar & Neta, 2006), which can potentially stem from a similar feeling of threat triggered by the sharpness of the angles. Moreover, using models of geometric patterns, studies of human body shapes and facial expressions suggest that simple sharp shapes (e.g., a V-shaped corner) convey threat, whereas simple round shapes (e.g., an O-shaped element) convey warmth and calmness (Aronoff, Woike, & Hyman, 1992). Lastly, in a widely used nonverbal technique for assessing arousal, the Self-Assessment Manikin (Bradley & Lang, 1994), people are asked to rate the level of arousal of different types of stimuli by selecting from among a set of manikins that best fits their perceived feeling. The two extreme manikins on the left and on the right of the rating scale intuitively make use of spiky shapes versus rounded shapes to pictorially represent high levels of arousal versus low levels of arousal, respectively (see Fig. 1, bottom).

In sum, we hypothesized that in *bouba-kiki* experiments, people rely on the arousal evoked by sounds and shapes to match auditory and visual stimuli to one another. To test this, we conducted three experiments. In Experiment 1, we gathered a comprehensive list of the shapes and words used in previous studies examining the *bouba-kiki* effect. We then collected ratings of arousal for these shapes and words and compared these rating values across rounded–spiky and *bouba-kiki* categories. In Experiment 2, we generated a new set of pseudowords representative of English phonology. We then developed a model-driven measure of arousal for the previously published words and compared the

predicted values across *bouba-kiki* categories. Finally, to test the generalizability of arousal as a mediating factor, we asked participants in Experiment 3 (preregistered) to match a novel set of pseudowords varying in their arousal with the rounded–spiky shapes from previous studies.

Experiment 1

We asked participants to rate the degree of arousal of shapes and words in two separate conditions (shape condition and word condition) to determine whether *bouba*-like and *kiki*-like words and the corresponding shapes would yield different levels of arousal.

Method

Participants. Two groups of native-English-speaking Cornell University undergraduates participated for course credit. A total of 24 participants (14 women; age: $M = 19.5$ years, range = 18–21) rated the words and 28 participants (17 women; age: $M = 19.7$ years, range = 18–22) rated the shapes for affective arousal. The sample size was based on those used in similar rating studies of the affective arousal of words, pseudowords, or pictorial stimuli (18–30 ratings per item; e.g., Aryani, Conrad, et al., 2018; Bradley & Lang, 1994).

Materials. We selected previous studies investigating the *bouba-kiki* phenomenon by searching for the keywords “bouba kiki” on Google Scholar (<https://scholar.google.com>). From these studies, we selected those that used different sets of words and shapes, yielding a total of eight studies that used 29 different pairs of shapes and 45 different pairs of words. A professional male actor who was a native-English speaker recorded the stimuli. Words were spoken in a listlike manner to prevent affective prosody and recorded in a professional sound-recording booth.

Procedure. We asked participants to rate how exciting or calming each presented image felt (shape condition) or each presented word sounded (word condition), following the instructions of previous studies (Aryani, Conrad, et al., 2018; Bradley & Lang, 1994). The order of presentation within the lists of words and shapes was randomized across participants. Importantly, we did not use the Self-Assessment Manikin (Bradley & Lang, 1994) because the shapes in this instrument could potentially cause bias (see the introduction and Fig. 1, bottom). Instead, a 5-point scale was shown on the screen, characterized by five bars of different heights from *very calming* (1) to *very exciting* (5). For the words, participants were instructed to give their ratings solely on the basis of the sound of the item

and not its potential similarity to real words (for detailed instructions, see Figs. S1 and S2 in the Supplemental Material available online). Note that we used bars of increasing heights to visualize the intensity of arousal from low to high. This choice also provided a neutral counterpart to the Self-Assessment Manikin typically used to measure arousal and thus provides some continuity with that literature.

Analysis. Differences in the ratings of spiky shapes versus rounded shapes and *kiki*-like words versus *bouba*-like words were analyzed with *t* tests using the statistical software JMP Pro 14 (SAS Institute, 2018). For studies with more than one item in a category, we first calculated an average rating for each category and conducted the *t* test on the basis of the average ratings of participants.

Results

A comparison of the arousal ratings in the shape condition revealed significantly higher arousal for spiky shapes than rounded shapes (spiky: $M = 3.45$, $SE = 0.039$; rounded: $M = 2.43$, $SE = 0.037$; $p < .0001$, $d = 0.92$). We then compared the ratings separately for the materials used in each study. The results showed significant differences in arousal ratings between stimulus types in each study, with spiky shapes rated as having significantly higher arousal than rounded shapes (see Fig. 2a and Table S1 in the Supplemental Material).

In the word condition, we compared arousal ratings for *bouba*-like and *kiki*-like words; *kiki*-like words were rated significantly higher in arousal than *bouba*-like words (*kiki*: $M = 3.16$, $SE = 0.033$; *bouba*: $M = 2.53$, $SE = 0.033$; $p < .0001$, $d = 0.57$). Separate analyses for each of the prior studies showed that *kiki*-like words always elicited higher mean arousal ratings than *bouba*-like words (see Fig. 2b and Table S1), similar to the shape condition.

In Experiment 1, the ratings for both shapes and words showed a consistent effect of category on arousal, with spiky shapes and *kiki*-like words rated higher in arousal than rounded shapes and *bouba*-like words across all of the studies. We took this as initial support for our hypothesis, showing that the typical shapes and words used in previous studies possess similar levels of arousal over different sensory modalities (i.e., visual vs. auditory).

Experiment 2

In Experiment 2, we asked whether any wordlike stimulus can convey emotional information on the basis of the basic perceptual features of its sounds. Specifically, we developed a model-driven measure of the level of

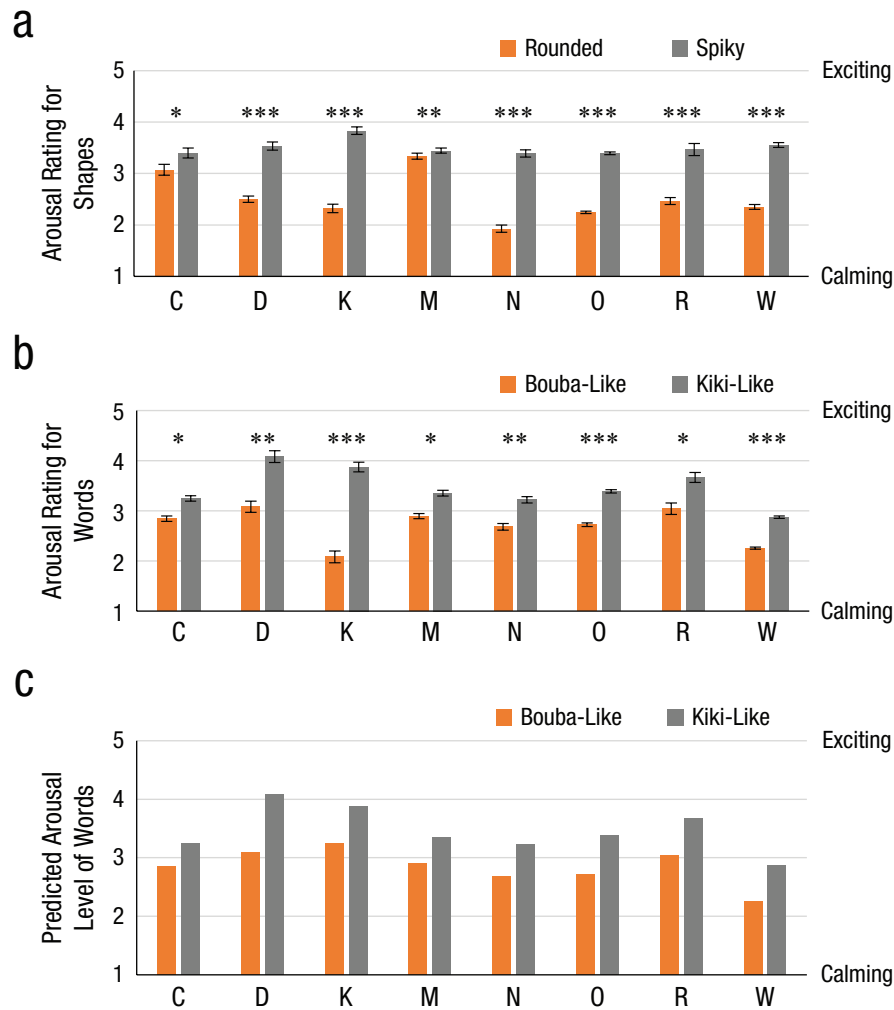


Fig. 2. Results from Experiments 1 and 2. For Experiment 1, arousal ratings are shown for each set of stimuli in the (a) shape condition and (b) word condition, separately for each of the eight previous studies. For Experiment 2 (c), model-driven predicted levels of arousal are shown for each word used in the previous studies. Asterisks indicate significant differences between stimuli ($*p < .05$, $**p < .01$, $***p < .001$). Error bars in (a) and (b) represent standard errors of the mean. The studies are referred to by the first author's initial: C = Cuskley, Simner, and Kirby (2017); D = Davis (1961); K = Köhler (1929/1947); M = Maurer, Pathman, and Mondloch (2006); N = Nielsen and Rendall (2011); O = Occelli, Esposito, Venuti, Arduino, and Zampini (2013); R = Ramachandran and Hubbard (2001); W = Westbury (2005).

arousal in the sound of *bouba-kiki* words that can overcome some of the typical issues related to direct subjective judgments in Experiment 1 (e.g., familiarity of items).

Method

A new group of participants rated a large set of novel auditorily presented pseudowords for their level of arousal. We then extracted the acoustic features of these pseudowords to develop acoustic models capable of predicting the variation in the ratings. These models

were applied to the word stimuli from Experiment 1 to predict their level of arousal solely on the basis of their acoustic features.

Participants. A group of native-English-speaking Cornell University undergraduates participated for course credit. This sample of participants was divided into two groups to develop two independent acoustic models to test the reliability of the results. Thus, we randomly assigned each of the 70 participants (53 women; age: $M = 18.8$ years, range = 17–22) to one of two groups of equal sizes ($n = 35$ each). The sample size was calculated

on the basis of the expected effect size from a similar previous study (Aryani, Conrad, et al., 2018; minimum $R^2 = 52.6\%$, $d = 0.72$) and assuming a power of 80%, resulting in a sample size of 32 participants per condition.

Materials. To generate novel words representative of the phonotactics of English, we used the Wuggy algorithm (Keuleers & Brysbaert, 2010). Wuggy generates pseudowords on the basis of the syllabic structure (i.e., onset, nucleus, and coda) of a given real word, thereby producing combinations of phonemes that are permissible in a given language. We chose the first 1,500 most frequent one- and two-syllable nouns from the English CELEX database (Baayen, Piepenbrock, & Gulikers, 1995) and adapted the program to generate five alternatives for each word. Candidate pseudowords that differed by fewer than two letters (added, deleted, or substituted) from the nearest real word were excluded (Coltheart distance = 1). For words with more than one remaining alternative, the one with the highest Levenshtein distance (i.e., the minimum number of insertions, deletions, or substitutions of a letter required to change the pseudoword into a real word) was selected. The list of words was checked for pseudohomophones, rare phonetic combinations, and similarity to real words by a native-English speaker. This resulted in 940 words, which were recorded as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1. We asked participants to evaluate 940 randomly presented pseudowords for how arousing they sounded (1 = *very calming*, 3 = *neutral*, 5 = *very exciting*). Ten practice items were presented at the beginning of the experimental session to familiarize participants with the task. The rating instructions were the same as those used in the word condition of Experiment 1 to elicit arousal ratings for previously published pseudowords. During a second inspection round (after collecting ratings), 11 pseudowords were identified as trisyllabic and excluded from subsequent analyses, resulting in 929 pseudowords (321 monosyllabic and 608 bisyllabic) in total.

Next, we extracted the acoustic features of these words and developed acoustic models that used these features to predict the variation in arousal ratings (cf. Aryani, Conrad, et al., 2018). We used the speech-analysis software Praat (Version 6.0.25; Boersma & Weenik, 2017) for the acoustic modeling. For each word, we extracted 14 acoustic features known to modulate emotional vocalization (Aryani, Conrad, et al., 2018; Juslin & Laukka, 2003): the mean of fundamental frequency f_0 (time step = .01, minimum = 75 Hz, maximum = 300 Hz); the mean, the minimum, the maximum, the median, and the standard deviation of intensity (time step = .01); and the mean and the bandwidth of the first three

formants, F1, F2, and F3 (time step = .01), from the spectral representation of the sound. Finally, the spectral centroid (spectral center of gravity) and the standard deviation of the spectrum were computed on the basis of fast Fourier transformations (time step = .01, minimum pitch = 75 Hz, maximum pitch = 300 Hz).

Analysis. For the regression models, standard least squares was chosen as the fitting method using the statistical software JMP Pro 14 (SAS Institute, 2018). We then applied these acoustic models to our list of 45 pairs of *bouba*-like and *kiki*-like words from the previous studies in Experiment 1 and predicted their affective arousal solely on the basis of their acoustic features.

Results

Across two groups of participants (Group 1 and Group 2), average arousal ratings for the entire set of words were highly correlated, $r = .73$, $p < .0001$. The results of linear regression models using the acoustic features to predict the arousal ratings showed that the acoustic features could account for 24.0% (for Group 1) and 22.9% (for Group 2) of the variance in arousal ratings (both R^2 adjusted, $ps < .0001$; for detailed model information, see Fig. S3 in the Supplemental Material). In both models, sound intensity and its related properties (e.g., minimum, maximum) were highly significant, replicating previous results on the dominant role of intensity in the arousal of different types of sounds (Aryani, Conrad, et al., 2018; Juslin & Laukka, 2003). We next took the acoustic models (i.e., the linear equations) and applied them to the extracted acoustic features of the *bouba*-like and *kiki*-like words to predict their arousal. We will refer to these two predicted values for arousal of *bouba*-like and *kiki*-like words as “predicted arousal based on G1 and G2” (PAG1 and PAG2, respectively).

A comparison of the predicted values of arousal of the set of *bouba*-like and *kiki*-like words from previous studies revealed significantly higher arousal for *kiki*-like words than *bouba*-like words, for both PAG1 (*kiki*: $M = 3.07$, $SE = 0.023$; *bouba*: $M = 2.90$, $SE = 0.020$), $t(88) = 5.91$, $p < .0001$, $d = 1.24$, and PAG2 (*kiki*: $M = 3.03$, $SE = 0.021$; *bouba*: $M = 2.85$, $SE = 0.019$), $t(88) = 6.17$, $p < .0001$, $d = 1.36$. Results of single comparisons across each of the eight selected studies showed that for both PAG1 and PAG2, the predicted values for the arousal of *kiki*-like words were consistently higher than those of *bouba*-like words in all of the studies (for Group 1 and Group 2 average ratings, see Fig. 2c). Because of the small number of *bouba*-like and *kiki*-like words in most of the studies, a statistical test of significance comparing the mean values at the item level within each study was not always possible (note that

this was not an issue in Experiment 1, in which the analyses were at the participant level). Instead, we tested a null hypothesis that the directions of these single comparisons (i.e., *kiki* > *bouba* vs. *bouba* > *kiki*) were determined by chance. Thus, we considered the success or failure of the results being in the direction expected by our hypothesis as several Bernoulli trials, with the words in each study being independent events with a probability of success (i.e., *kiki* > *bouba*) equal to chance ($p = .50$). Comparisons across all of the studies (see Table S2 in the Supplemental Material) showed that the predicted value of arousal for *kiki*-like words was higher than that for *bouba*-like words in all eight cases. Corresponding binomial tests thus rejected the null hypothesis: $X \sim B(8, .5)$, $p(X \geq 8) = .003$ for PAG1, and $X \sim B(8, .5)$, $p(X \geq 8) = .003$ for PAG2.

By showing that the predicted values for arousal significantly differ across *kiki*-like and *bouba*-like words, Experiment 2 provided further support for the hypothesis that arousal may mediate the *bouba-kiki* effect. Crucially, the present findings go beyond the limited selection of *bouba-kiki* words used in previous studies: Any wordlike stimulus can potentially convey emotional information solely on the basis of its perceptual acoustic characteristics (cf. Aryani, Conrad, et al., 2018; Aryani, Hsu, & Jacobs, 2018; Aryani & Jacobs, 2018), making it possible to match it with emotionally similar concepts (e.g., shapes).

Experiment 3

To further establish the generalizability of the emotional-mediation hypothesis, we sought to replicate the classic *bouba-kiki* effect with a new set of words in a preregistered study (<https://aspredicted.org/mq97g.pdf>). We hypothesized that any word (or pseudoword) would be matched with spiky shapes versus rounded shapes depending on the level of arousal in its sound.

Method

Participants. A group of 64 native-English-speaking Cornell University undergraduates participated for course credit (49 women; age: $M = 19.2$ years, range = 18–22). Assuming an effect size (d) of 0.5 (a medium effect), this sample size provides adequate power of 80%.

Materials. On the basis of the ratings collected in Experiment 2, we selected 168 words and assigned them to three distinct groups of high arousal (> 3.12), medium arousal (2.87–3.01), and low arousal (< 2.75; see Table S3 in the Supplemental Material). The number of selected words was motivated by the number of available spiky and rounded shape pairs from previous studies ($N = 28$)

and was chosen so that each pair would be presented an equal number of times across the experiment. Because of an existing confound between arousal-rating value and both the number of phonemes and the number of syllables in the words ($r = .54$, $p < .0001$, and $r = .53$, $p < .0001$, respectively), we controlled the number of phonemes across the three arousal categories (see Table S3) and selected the same number of monosyllabic words ($n = 16$) and bisyllabic words ($n = 40$) for each of these categories.

Procedure. The order of presentation of the 168 words was randomized for each participant. During each presentation, a pair of spiky or rounded shapes from previously published studies was displayed. The position of each shape was counterbalanced; each shape in a pair appeared thrice throughout the course of the experiment on either side of the screen. Participants were given a two-alternative forced-choice task in which they used the keyboard to indicate which of the shapes best fitted a word they had heard.

Analysis. As was preregistered, responses that took longer than 5 s were excluded from analysis (230 of 10,752), resulting in 10,522 responses that were analyzed. We fitted a logistic generalized linear mixed model to our data set using the *glmmTMB* package (Brooks et al., 2017) in the R software environment (Version 3.6.1; R Core Team, 2019). We modeled response to shapes as a binary outcome (1 = spiky, 0 = rounded), word category (high, medium, low) as the predictor, and subject and word as random effects, and we used the binomial link function (the entire R code is available at <https://osf.io/a8sd4>).

Results

In line with our hypothesis, results showed that the average probability of selecting a spiky shape over a rounded shape (i.e., reporting a response of 1) was significantly higher for the words in the high-arousal group than for those in the medium-arousal group, which was, in turn, higher than those in the low-arousal group. A summary of the actual results is shown in Figure 3 (for a summary of the results of the model output and the results of pairwise comparisons between each pair of word groups, see Tables S4 and S5 in the Supplemental Material).

These results suggest that the extent to which a word in a *bouba-kiki* experiment is matched with a rounded shape or spiky shape depends on the level of arousal elicited by its sound. Importantly, the results generalize the mediating role of arousal to a new set of pseudowords and show that varying one factor in the characteristics of

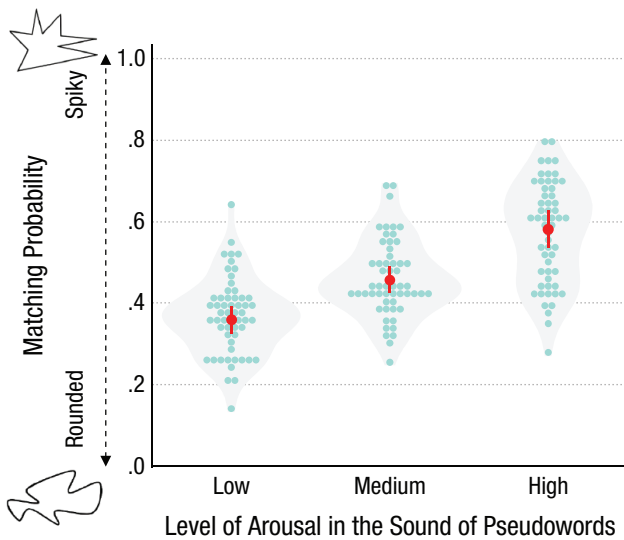


Fig. 3. Results from the two-alternative forced-choice task in Experiment 3: average probability of matching a pseudoword with a spiky shape or a rounded shape as a function of the level of arousal in its sound. Blue dots represent individual data points, and red dots represent group means. The width of the gray plots indicates the density of the data. The vertical red lines indicate standard errors of the mean.

stimuli (i.e., arousal) in one modality (i.e., auditory words) can significantly impact the outcome of matching preferences in the other modality (i.e., visual images).

General Discussion

The present results substantiate the emotional-mediation hypothesis (e.g., Marks, 1975; Walker et al., 2012) in the *bouba-kiki* paradigm, suggesting that information from different sensory domains can be linked through shared emotional associations. Our findings show that the tendency to match pseudowords with spiky shapes versus rounded shapes is associated with the level of arousal in their sound.

Although our data do not permit a direct conclusion about the causal role of arousal in driving the *bouba-kiki* effect, many of the effect's characteristics, such as its prevalence across different age groups (e.g., Ozturk et al., 2013) and cultures (e.g., Bremner et al., 2013), can be explained by the emotional-mediation hypothesis. Particularly, recent findings on the occurrence of the effect even in the absence of visual awareness (Heyman et al., 2019; Hung et al., 2017) suggest the involvement of automatic processes that bypass higher order cognitive analysis. This type of stimulus–response processing has been frequently reported for highly emotional stimuli; for instance, individuals with cortical blindness show spontaneous facial and pupillary

reactions to nonconsciously perceived expressions of fear and happiness (Tamietto et al., 2009). Although further studies are needed to determine the causal role of emotion in cross-modal mappings, our data suggest that affective arousal serves as a potent mediator between these domains.

Our results may also provide an answer to the problem of grounding abstract concepts during development. Despite the valuable insights of previous investigations, it is still a matter of debate as to how abstract sound categories (i.e., phonemes) can be linked to abstract shape forms (e.g., spikiness) merely on the basis of the physical characteristics of the stimuli. In line with theories emphasizing the role of affect in the representation of abstract words (Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Vigliocco, Meteyard, Andrews, & Kousta, 2009), our results suggest that the lack of direct sensorimotor information in an abstract concept may in part be compensated for by its affective relevance, which may help build links between specific abstract concepts and specific affective states. Accordingly, an abstract concept such as spikiness may be grounded in the affective system by the embodied experience of harmfulness and, hence, high arousal. Thus, learning to build abstract categories on the basis of their emotional relevance may be a crucial stepping stone in the development of abstract semantic representations during early word learning.

The present findings may moreover shed new light on the early stages of language evolution and the emergence of linguistic signs by highlighting the role of arousal. The *expression* and *perception* of affective states are fundamental aspects of human communication and, since Darwin (1871), have been viewed as a key impetus for language evolution. At the level of vocal expressions, arousal can be readily reflected in human vocalizations and thus extends to acoustic features of the speech signal. At the perception level, humans can reliably identify levels of arousal in the vocalizations of other humans (Juslin & Laukka, 2003) as well as in diverse vertebrate species (Filippi et al., 2017). Thus, an initial form of vocalization may have been a motor reflection of arousal in the vocal tract. Such vocalizations might gradually have been used to refer to external objects that were associated with similar affective experience (e.g., sharp rocks becoming associated with high-arousal sounds). Grounding linguistic signs in the affective system may thereby have provided human verbal communication with an essential building block that is easily learnable and, thus, likely to be retained over the course of the cultural evolution of language (Christiansen & Chater, 2008).

Transparency

Action Editor: Rebecca Treiman

Editor: D. Stephen Lindsay

Author Contributions

A. Aryani and M. H. Christiansen designed the research, A. Aryani and E. S. Isbilen conducted the research, and A. Aryani analyzed the data. All the authors wrote the manuscript and approved the final version for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research was funded by grants to A. Aryani from the Direct Exchange Program of Freie Universität Berlin and to E. S. Isbilen from the National Science Foundation Graduate Research Fellowships Program (No. DGE-1650441).


Open Practices

All data and stimuli have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/a8sd4>. The design and analysis plans were preregistered at <https://aspredicted.org/mq97g.pdf>. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797620927967>. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Arash Aryani  <https://orcid.org/0000-0002-0246-9422>

Morten H. Christiansen  <https://orcid.org/0000-0002-3850-0655>

Acknowledgments

We are grateful to Felix Thoemmes for guidance on the statistical analyses and to Stewart McCauley for generating an earlier version of the pseudowords. We also thank Dante Dahabreh, Emma Goldenthal, Phoebe Ilevbare, Eleni Kohilakis, Jake Kolenda, Farrah Mawani, Jeanne Powell, and Olivia Wang for their help in preparing the stimuli and conducting the experiments. We especially thank Emma Goldenthal and Eleni Kohilakis for their help with the International Phonetic Alphabet transcriptions of the stimuli from Experiment 2.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797620927967>

Note

1. Note that we use the term *word* to refer to both words with lexical meaning and pseudowords without lexical meaning (e.g., *kiki*, *bouba*).

References

- Aronoff, J., Woike, B. A., & Hyman, L. M. (1992). Which are the stimuli in facial displays of anger and happiness? Configurational bases of emotion recognition. *Journal of Personality and Social Psychology*, *62*, 1050–1066.
- Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why ‘piss’ is ruder than ‘pee’? The role of sound in affective meaning making. *PLOS ONE*, *13*(6), Article e0198430. doi:10.1371/journal.pone.0198430
- Aryani, A., Hsu, C. T., & Jacobs, A. M. (2018). The sound of words evokes affective brain responses. *Brain Sciences*, *8*(6), Article 94. doi:10.3390/brainsci8060094
- Aryani, A., Hsu, C. T., & Jacobs, A. M. (2019). Affective iconic words benefit from additional sound-meaning integration in the left amygdala. *Human Brain Mapping*, *40*, 5289–5300. doi:10.1002/hbm.24772
- Aryani, A., & Jacobs, A. M. (2018). Affective congruence between sound and meaning of words facilitates semantic decision. *Behavioral Sciences*, *8*(6), Article 56. doi:10.3390/bs8060056
- Auracher, J. (2017). Sound iconicity of abstract concepts: Place of articulation is implicitly associated with abstract concepts of size and social dominance. *PLOS ONE*, *12*(11), Article e0187196. doi:10.1371/journal.pone.0187196
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, *17*, 645–648.
- Boersma, P., & Weenik, D. (2017). Praat: Doing phonetics by computer (Version 6.0.25) [Computer software]. Retrieved from <http://www.praat.org/>
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*, 49–59.
- Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). “Bouba” and “Kiki” in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, *126*, 165–172. doi:10.1016/j.cognition.2012.09.007
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*, 378–400.
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral & Brain Sciences*, *31*, 489–509. doi:10.1017/S0140525X08004998
- Cuskley, C., Simner, J., & Kirby, S. (2017). Phonological and orthographic influences in the bouba–kiki effect. *Psychological Research*, *81*, 119–130. doi:10.1007/s00426-015-0709-2
- Darwin, C. (1871). *The descent of man and selection in relation to sex* (Vol. 1). London, England: John Murray.
- Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology*, *52*, 259–268.

- De Saussure, F. (1916). *Course in general linguistics*. New York, NY: McGraw-Hill.
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences, 19*, 603–615. doi:10.1016/j.tics.2015.07.013
- Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., . . . Güntürkün, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences, 284*(1859), Article 20170990. doi:10.1098/rspb.2017.0990
- Heyman, T., Maerten, A.-S., Vankrunkelsven, H., Voorspoels, W., & Moors, P. (2019). Sound-symbolism effects in the absence of awareness: A replication study. *Psychological Science, 30*, 1638–1647. doi:10.1177/0956797619875482
- Hoshi, H., Kwon, N., Akita, K., & Auracher, J. (2019). Semantic associations dominate over perceptual associations in vowel–size iconicity. *i-Perception, 10*(4). doi:10.1177/2041669519861981
- Hung, S.-M., Styles, S. J., & Hsieh, P.-J. (2017). Can a word sound like a shape before you have seen it? Sound-shape mapping prior to conscious awareness. *Psychological Science, 28*, 263–275.
- Isbilen, E. S., & Krumhansl, C. L. (2016). The color of music: Emotion-mediated associations to Bach's *Well-Tempered Clavier*. *Psychomusicology: Music, Mind, & Brain, 26*, 149–161.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770–814. doi:10.1037/0033-2909.129.5.770
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*, 627–633. doi:10.3758/BRM.42.3.627
- Köhler, W. (1947). *Gestalt psychology*. New York, NY: Liveright. (Original work published 1929)
- Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140*, 14–34. doi:10.1037/a0021446
- Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin, 82*, 303–331.
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science, 9*, 316–322. doi:10.1111/j.1467-7687.2006.00495.x
- Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology, 65*, 115–124. doi:10.1037/a0022268
- Ocellli, V., Esposito, G., Venuti, P., Arduino, G. M., & Zampini, M. (2013). The Takete-Maluma phenomenon in autism spectrum disorders. *Perception, 42*, 233–241.
- Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin, 49*, 197–237.
- Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology, 114*, 173–186. doi:10.1016/j.jecp.2012.05.004
- Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology, 1*, Article 227. doi:10.3389/fpsyg.2010.00227
- Ramachandran, V. S., & Hubbard, E. H. (2001). Synaesthesia—A window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3–34.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- SAS Institute. (2018). JMP Pro 14 [Computer software]. Cary, NC: Author.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*, 971–995. doi:10.3758/s13414-010-0073-7
- Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., & de Gelder, B. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proceedings of the National Academy of Sciences, USA, 106*, 17661–17666.
- Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*, 219–247. doi:10.1515/LANGCOG.2009.011
- Walker, L., Walker, P., & Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception, 41*, 1186–1192. doi:10.1068/p7149
- Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language, 93*, 10–19. doi:10.1016/j.bandl.2004.07.006
- Westbury, C. (2018). Implicit sound symbolism effect in lexical access, revisited: A requiem for the interference task paradigm. *Journal of Articles in Support of the Null Hypothesis, 15*(1), 1–12.