Cognition 152 (2016) 141–149

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The sound of distance

Cristina D. Rabaglia [a], Sam J. Maglio [b,*], Madelaine Krehm [c], Jin H. Seok [d], Yaacov Trope [c]

[a] *University of Toronto Mississauga, Mississauga, Ontario, Canada*
[b] *University of Toronto Scarborough, Toronto, Ontario, Canada*
[c] *New York University, New York, NY, United States*
[d] *University of Missouri Columbia, Columbia, MO, United States*

ARTICLE INFO

Article history:
Received 14 February 2015
Revised 29 March 2016
Accepted 1 April 2016

Keywords:
Language
Sound symbolism
Distance
Spatial cognition

ABSTRACT

Human languages may be more than completely arbitrary symbolic systems. A growing literature supports *sound symbolism*, or the existence of consistent, intuitive relationships between speech sounds and specific concepts. Prior work establishes that these sound-to-meaning mappings can shape language-related judgments and decisions, but do their effects generalize beyond merely the linguistic and truly color how we navigate our environment? We examine this possibility, relating a predominant sound symbolic distinction (vowel frontness) to a novel associate (spatial proximity) in five studies. We show that changing one vowel in a label can influence estimations of distance, impacting judgment, perception, and action. The results (1) provide the first experimental support for a relationship between vowels and spatial distance and (2) demonstrate that sound-to-meaning mappings have outcomes that extend beyond just language and can – through a single sound – influence how we perceive and behave toward objects in the world.

1. Introduction

Upon asking people to match two never-before-seen words (“mil” and “mal”) with two tables of different sizes, Sapir (1929) found a startling consistency in participants’ choices: Nearly 80% assigned “mal” to the larger table, and “mil” to the smaller table. Since the words were unfamiliar, this uniformity in pairings cannot be attributable to word-specific prior knowledge or shared convention; rather, participants must have drawn on some *other* shared knowledge or intuition. Since the two labels differed only in a single speech sound, this shared intuition somehow must be associated with a very specific, elementary component of language.

Such intuitive sound-to-meaning mappings, or *sound symbolism*, are surprising when seen from perspectives in linguistics and cognitive science that stress arbitrariness as a core feature of human language. Such accounts offer convention among a language’s speakers as the mechanism linking words with their referents, and thus reject any *intrinsic* linkages between particular sounds and meaning (e.g., Hockett, 1958; Saussure, Bally, Sechehaye, Riedlinger, & Baskin, 1966). Sound symbolism constitutes a break in this arbitrariness, suggesting instead the existence of psychological links between referents and particular word forms, and as such can be viewed as a particular class of effect within the broader concept of *iconicity* in human language (Perniss, Thompson, & Vigliocco, 2010). In opposition to complete arbitrariness, iconicity posits a resemblance between the form of a linguistic sign (e.g., a word) and its referent. In more overt examples of iconicity (e.g., ideophones), the form of a word is intended to mimic some aspect of the referent (such as the sonic experience of a *gun going off:* “bang” in English, or the rapid heartbeat associated with *emotional excitement*: “doki doki” in Japanese). Sound symbolism proposes more subtle, intuitive linkages between individual language sounds (e.g., vowels produced in the front or the back of the mouth) and abstract concepts (e.g., size or shape). These sound symbolic effects – our focus in the present investigation – may derive from one or multiple origins, including intuitive, perhaps even synesthetic links between specific concepts and the acoustic qualities of the sound (Ramachandran & Hubbard, 2001) or the embodied experience of producing (or observing the production of) the individual phonemes that serve as the building blocks for human language (Tanz, 1971).

Investigations into iconicity and sound symbolism have provided mounting evidence against a complete arbitrariness in human language, with support converging from both cross-linguistic and experimental psychological approaches. Cross-linguistic evidence – by documenting consistencies in mappings between particular phonemes and particular concepts in words occurring in diverse individual languages and language families

* Corresponding author at: University of Toronto Scarborough, Management Division, 1265 Military Trail, Toronto, ON M1C 1A4, Canada.
E-mail address: sam.maglio@utoronto.ca (S.J. Maglio).

http://dx.doi.org/10.1016/j.cognition.2016.04.001
0010-0277/

(e.g., Dingemanse, 2012; Perry, Perlman, & Lupyan, 2015; Tanz, 1971) – supports the existence of iconic, sound-symbolic linkages across extant human languages. On the other hand, experimental work (using novel word forms) within particular communities of speakers – by documenting the existence of intuitive associations between particular phonemes and particular concepts – reveals the underlying psychological potential for sound-symbolism based iconicity in human language.

This latter, experimental psychology-based approach has drawn largely from two experimental paradigms to establish how people derive intuitive meaning from individual phonemes. First, foreign language intuition studies demonstrate that individuals are above chance (in closed-response tasks) at inferring the meaning of words in an unfamiliar language (e.g., Brown, Black, & Horowitz, 1955; Koriat & Levy, 1977; Parault & Schwanenflugel, 2006). Native English speakers, for example, perform significantly above chance when asked to match a pair of words in an unfamiliar language with appropriate English antonym-synonym pairs (e.g., hard/soft; warm/cool) reflecting sensory properties (Brown et al., 1955). A second approach examines biases in matching novel (experimenter-created) words to referents. As with "mil" and "mal" above, this methodology typically uses some version of a forced-choice task – either asking participants to pick the appropriate word for a novel object or the suitable referent for a particular word – to ascertain the intuitive meanings of specific sounds. This and related research has supported the existence of linkages between specific speech sounds and a variety of concepts: size (Sapir, 1929), shape (Maurer, Pathman, & Mondloch, 2006), creaminess (Yorkston & Menon, 2004), and speed/lightness/sharpness (Lowrey & Shrum, 2007), to sample only a few (for a comprehensive review, see Shrum & Lowrey, 2007).

A common feature in these approaches, however, is the overt involvement of language itself. Traditional sound symbolic work directly probes intuitions *about* language, asking people to guess the referent of a foreign-language word, to pick the label for a novel object, or to choose the preferred object for a label. Even the handful of studies outside these two dominant approaches still assess language-related outcomes, investigating the effects of sound symbolism on speed of lexical access (Westbury, 2005) or word learning (Imai, Kita, Nagumo, & Okada, 2008; Nygaard, Cook, & Namy, 2009; Yoshida, 2012) and typically using within-subjects designs that bring the linguistic nature of the task to the forefront. As a result, it remains unclear whether these sound-to-meaning linkages are mainly meta-linguistic in nature. That is, are sound symbolic effects (found within speakers of a given language) dependent on overtly linguistic tasks and outcomes, or do they reflect deeper associations that can impact perceptions of objects, events, and concepts in the world?

To probe these questions, the current studies approach the investigation of sound symbolism in a different way. One potential – but experimentally untested – sound symbolic relationship lends itself particularly well to such consideration: the link between the front/back vowel distinction and spatial distance. Front vowels (e.g., in General American English, /i/, the "ee" sound in "feet") are produced with the highest point of the tongue relatively forward in the mouth, while back vowels (e.g., /u/, the "oo" sound in "food") are produced with the highest point of the tongue relatively far back in the mouth. Converging cross-linguistic evidence suggests that front vowels may be associated with spatial proximity while back vowels may be associated with distance (Tanz, 1971; Ultan, 1978). For example, Tanz (1971) reported that in six separate language families, in which words meaning "here" (proximal) and "there" (distant) differed by only one vowel, the word for "here" always contained a front vowel and the word for "there" a back vowel. Cross-linguistic work thus suggests a potential association between vowel location and spatial distance, but this association as yet lacks direct experimental testing within speakers of a language.

Thus, the goals of the current studies were twofold. First, we sought to test the association between vowel backness and spatial distance. Despite limitations inherent to asking participants directly about language-related outcomes, our first experiment (Study 1) employs this type of design in order to offer continuity between prior investigations and ours. Second, to explore whether this sound symbolic effect persists even in tasks with non-linguistic outcomes, we bring manipulations of front/back vowels (at the independent variable level) to bear on dependent variables taken from outside the psycholinguistic literature (Studies 2–4) and ask if sound symbolic effects can impact even basic perceptual processes (Studies 3–4) and thus truly color our world.

2. Study 1: foreign language intuition

2.1. Stimuli

To establish the relationship between vowel backness and distance using more traditional sound symbolic methods, Study 1 adopted a foreign language intuition paradigm. We generated 200 pseudowords using the ARC psycholinguistic database (Rastle, Harrington, & Coltheart, 2002). All items only contained a single vowel and were orthographically and phonotactically valid in English. To ensure uniformity of target vowel pronunciation within the individual items, we chose items with few "body enemies" (orthographically similar items that would be pronounced differently). Half included front vowels, and the other half included back vowels. Within each vowel class, we used two different vowels, one each from relatively high and low vowel heights: /i/ (as in "beat", high front), /ɛ/ (as in "bet", lower front), /u/ (as in "boot", high back), and /ɒ/ (as in "bought", lower back). See Table 1 for a sub-sample of linguistic stimuli used in this study.

To test whether the construct of vowel frontness/backness is associated with spatial distance, we conceptualized this vowel dimension as a continuum, along which individual vowel exemplars can fall. In keeping with dominant traditions in phonetics, we operationalized individual vowel frontness/backness as a function of the ratio between two vowel formant frequencies: F1 (correlated with the area of the back oropharyngeal cavity and the degree of opening of the lips, indicating the height of the vowel) and F2 (correlated with the length of the front oral cavity,

Table 1
Vowels used in each study with experimental items.

Study	Front Vowel Item	Vowel	Back Vowel Item	Vowel
1	dweek	/i/	blune	/u/
	keam	/i/	koob	/u/
	hept	/ɛ/	thraw	/ɒ/
	zelm	/ɛ/	zaunt	/ɒ/
2	sneal	/i/	fluce	/u/
	fleeve	/i/	pruke	/u/
	belve	/ɛ/	blope	/o/
	sleph	/ɛ/	swoan	/o/
3	fleen	/i/	floon	/u/
4a	[dib]	/i/	[dob]	/o/
	[kiv]	/i/	[kuv]	/u/
	[nɛp]	/ɛ/	[nup]	/u/
4b	[dib]	/i/	[dob]	/o/

Note. When stimuli were presented as printed words, the "Item" columns present individual stimulus items as participants saw them. When stimuli were presented aloud, the "Item" columns present individual stimulus items transcribed in IPA as participants heard them. For Experiments 1 and 2, items shown are a randomly-selected subset from each vowel from the 200 and 76 total linguistic items, respectively, used in those experiments.

indicating its frontness/backness). Following typical phonetic tradition (Ladefoged, 2006), we express this in terms of the F1/F2 ratio, using average F1 and F2 formant frequency values for each vowel (Catford, 1988).

We decided ahead of time to create 50 different nonwords for each of the four specific vowel sounds to provide evidence for a robust effect across different frames (i.e., different consonants paired with the vowel to create the word). Accordingly, we recruited 200 participants so that each could evaluate one of our nonword targets and collected data until that number had been reached. This constituted the largest sample size in the present investigation, in keeping with the typical lower power of binary choice measures relative to point estimates (used in the subsequent studies).

2.2. Procedure

Two hundred American volunteers were recruited from the Amazon Mechanical Turk platform to complete a survey on language. After enrolling in the experiment on the Mechanical Turk platform, participants were directed to a separate website (hosted by Qualtrics, a survey design website) through which they completed the experiment individually. The instructions to participants read as follows: "You will see a word from a language spoken somewhere in the world and your task is to guess, using your intuition, what the meaning of the word is. First, you will see the word, and we would like you to read the word to yourself. Then, you will be given two possible meanings for that word and asked to try to guess which of these is the correct meaning." Each participant was then presented with one of the nonword stimuli we had created and asked "Does this word refer to something that is close to you in space, or something that is remote?"; participants chose via mouse-click, and the order of presentation for the two answers ("close"/"remote") was counterbalanced. We specifically chose these particular distance terms (i.e., "close" and "remote") as our dependent measure items because they are nearly matched in terms of their individual vowel content (i.e., both contain a back vowel; neither contain a front vowel). This helped ensure that any potential differences identified were not driven by an experimentally-created association between vowel and location (e.g., if we had instead used "near", containing a front vowel, and "far", containing a vowel produced farther back). We predicted that participants would prove increasingly likely to choose "remote" over "close" for target words containing vowels produced farther in the back of the mouth.

2.3. Results and discussion

As choice rates did not vary as a function of the order in which the two options ("close"/"remote") were presented, $p = 0.29$, we collapsed across this factor. We regressed nonword meaning choice (dummy coded as "close" = 0 and "remote" = 1) on nonword vowel frontness/backness in a binary logistic regression. The formant ratios for each of the four vowel types, increasingly to the back of the mouth, were as follows: /i/ = 0.100; /ɛ/ = 0.321; /u/ = 0.420; /ɒ/ = 0.921 (Catford, 1988). As predicted, participants considering nonwords containing vowels produced increasingly to the rear of the mouth (i.e., scoring higher on the ratio of F1/F2) were more likely to guess that the word meant "remote", $\chi 2\ (1) = 7.863$, $b = 1.367$, odds ratio = 3.922, $p = 0.006$.

We note that the results fall along a linear trend (cf. Thompson & Estes, 2011), depicted in Fig. 1, rather than, for example, a step function in which the rates of choice for "remote" are identically low for front vowels and identically high for back vowels. This offers evidence in support of our contention that frontness/backness (conceptualized along a continuum) accounts for our effect rather than some other linguistic property that may co-occur with production location (e.g., rounding, which covaries with backness in English). To further support this claim, we conducted two subsequent binary logistic regressions, swapping out the format ratio equation for either just the F1 frequency for the vowel sounds (a proxy for height) or just the F2 frequency (a proxy for location). Whereas F1 did not prove a reliable predictor of choice, $p = 0.55$, F2 alone did, $p = 0.012$. Cues to vowel frontness/backness, rather than vowel height or rounding, thus predict intuitions about word meaning.

Absent any information other than the linguistic form of two words, participants proved more likely to intuit that a front-vowel containing word would be used to indicate physical proximity and a back-vowel containing word indicated greater physical distance. This provides initial evidence of an association between vowels and spatial distance in a paradigm typical of prior sound symbolism research. To extend beyond judgments specifically *about* word meanings – and to provide converging evidence for the vowel-distance association with more tightly controlled linguistic stimuli – we employ a new paradigm in Study 2.

3. Study 2: automaticity between vowel and distance

3.1. Stimuli

Prior sound symbolism research on shape (Westbury, 2005) has demonstrated that congruency in sound-to-meaning mappings can influence reaction times in a modified Stroop-like lexical decision task (for related work, see Kovic, Plunkett, & Westermann, 2010), suggesting that the process whereby people infer meaning from sound can occur automatically (i.e., effortlessly and uncontrollably, Yorkston & Menon, 2004). In Study 2, we explore the automaticity of the association between vowels and spatial distance in a task not overtly about word meanings (i.e., a modified Stroop task, Bar-Anan, Liberman, Trope, & Algom, 2007). Prior research has attested to the validity of this task in identifying an automatic relationship between spatial distance and seemingly unrelated constructs; we apply a similar logic in connecting another seemingly unrelated stimulus (vowels).

We generated a set of 32 experimental items by selecting monosyllabic pseudowords from the ARC nonword database, in the same way described above, in Study 1. These items comprised 16 nonwords (Table 1) containing front vowels (8 each of the vowels /i/ and /ɛ/) and 16 nonwords containing back vowels (8 each of the vowels /u/ and /ɒ/). An additional 40 nonwords containing central vowels served as filler items. We decided ahead of time on a sample size of 30 participants and collected data until that number had been reached. We arrived at this sample size based upon consideration of a similar investigation using a Stroop-like paradigm in order to investigate spatial distance (Bar-Anan et al., 2007).

3.2. Procedure

Thirty undergraduate students from New York University participated in the experiment in exchange for course credit. They participated in individual testing sessions in which they arrived at the research lab and met with a research assistant, who seated participants at a computer terminal and told them that they would perform a task related to analyzing pictures. The instructions indicated that they would be presented with a series of photographs of landscape scenes on the computer monitor, one at a time, which all included a road stretching into the distance and included an arrow overlaid in either a spatially near or far location with one of our nonword stimuli contained within the arrow (written in 24-point Helvetica font). The participants' task for each photograph was to

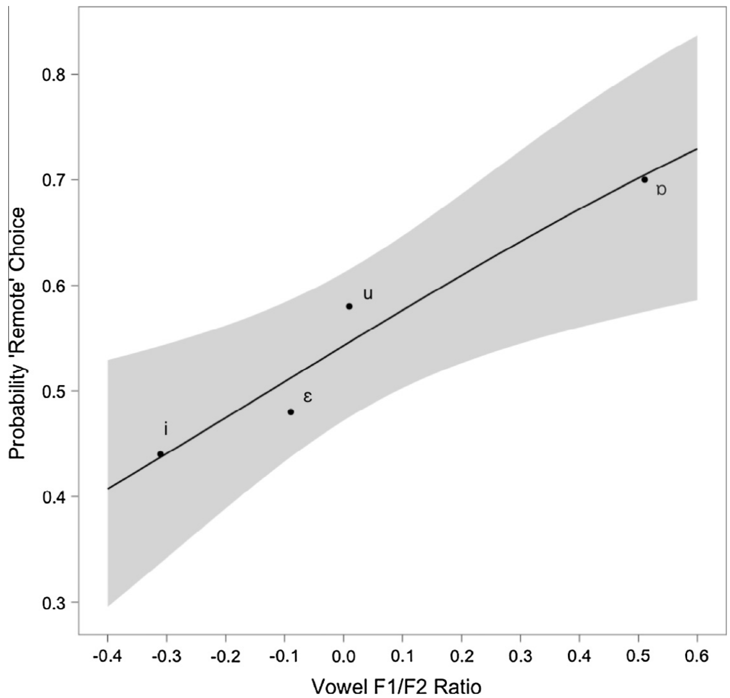

Fig. 1. Predicted probability of choosing "remote" as the meaning of a nonword. *Note*. Predicted probability (black fit line) of choosing "remote" is plotted as a function of centered vowel F1/F2 ratio, along with 95% CI of the fitted model (gray ribbon) and actual proportion of "remote" choices made for the four vowels of collected data (plotted points). Vowel ratios are thus plotted with reference to a theoretical center representing the midpoint in F1/F2 ratio between the two extreme vowels /i/ and /ɒ/.

read the nonword aloud and then to indicate – via pressing a button on the keyboard – whether the arrow was located spatially near or far in the landscape. Participants were instructed to respond as quickly as possible without sacrificing accuracy. They completed four practice trials to familiarize themselves with the task before proceeding to the main block of trials. All experimental and practice trials were presented using the E-Prime experimental software, which also recorded accuracy and reaction times for each trial. Since reaction time was recorded as the amount of time that elapsed between the image onset and their button-press response, measured reaction times in this experiment included the time it took them to read the word aloud.

The experimental stimuli included a total of 72 nonwords: 16 including front vowels, 16 including back vowels, and 40 filler stimuli containing central vowels. Because each participant evaluated every nonword once, the main block comprised 72 trials for every participant. Separately, we collected six photographs that conveyed geographical distance, with a road spanning from a close to a far physical location. Within participants, the 72 experimental nonword items were presented in a random sequential order and were also randomly assigned to be presented in conjunction with one of six possible road photographs an equal number of times (i.e., using each photograph a total of 12 times) and with the included arrow presented in either a near or far location. This random assignment was constrained such that, within a given participant, items from each category of vowel stimulus (i.e., front, back, and central) appeared in a near arrow half of the time and a far arrow half of the time and each road photograph and arrow location was presented an even number of times across the experiment (i.e., fully counterbalanced for vowel stimulus category, arrow location, and specific photograph). Half of the participants were instructed to press the "j" key to indicate that the arrow appeared in a near location and the "f" key to indicate that the arrow appeared in a far location; the other half received the opposite mappings. The image (together with the arrow and the word within it) remained on the screen until the participant responded via keyboard press. We predicted that, if vowel frontness/backness is associated with distance, then congruency in sound-to-location pairings (i.e., front vowel words with near targets, back vowel words with far targets) should facilitate processing in a Stroop-like task.

3.3. Results and discussion

Prior to analysis, we excluded data from four participants with overall arrow location response accuracies of less than 90%. We also discarded reaction times equal to or greater than 3 standard deviations from each individual participant's mean (1.4% of the reaction time data).

As reaction times did not differ as a function of the key-press counterbalancing manipulation (nor did this factor interact with either arrow location or vowel location), we collapsed across this factor. Consistent with our hypothesis, a 2 (arrow location, near/-far) × 2 (vowel location, front/back) repeated-measures ANOVA revealed no main effects of arrow location ($p = 0.20$) or vowel ($p = 0.33$), indicating that participants were not faster to indicate the location of a near or far arrow in general, nor arrows containing front or back vowels in general. However, there was a significant two-way interaction between vowel location and arrow location, $F(1,25) = 10.18$, $p < 0.01$, $\eta_p^2 = 0.29$, indicating – as predicted – that the effect of vowel on reaction times differed as a function of arrow location. Reaction times showed evidence of a significant Stroop effect: Participants classified an arrow's location faster when presented with a vowel matching the hypothesized sound symbolic meaning (i.e., front vowels with near arrows and back vowels with far arrows, $M = 1076$ ms, $SD = 269$) than when presented with a mismatched vowel (i.e., front vowels with far arrows and back vowels with near arrows, $M = 1144$ ms, $SD = 317$), $p < 0.01$.

Thus, Study 2 provides converging evidence for a sound symbolic relationship between vowel location and spatial distance, extending beyond the traditional approach (i.e., Study 1) and demonstrating that this association is basic or primary enough to influence automatic processing. In order to ensure that participants attended not only to the arrow location but also to the word that it contained, we asked them to read the word aloud, which very likely contributed to our slower observed reaction times (i.e., in excess of one second) than are typically found in reaction time experiments (e.g., Bar-Anan et al., 2007). Of note, the absence of any main effect for vowel suggests that participants did not respond especially fast or slow to either front or back vowels (despite certain differences in vowel duration; Liu, Jin, & Chen, 2014) but, instead, showed evidence of a matching or fit facilitation effect, responding especially quickly when front vowels were paired with near arrows and back vowels were paired with far arrows.

The stated task was to indicate the location of the arrow, signifying that this sound symbolic effect held even when the outcome was not explicitly *about* word meanings. However, we note that reading a succession of unfamiliar linguistic stimuli in the context of a fairly simple task may have prompted consideration of language by our participants. This leaves open the question as to whether such effects exist in tasks completely lacking any overtly linguistic component and which our next study attempts to resolve.

4. Study 3: from vowel to distance

4.1. Stimuli

The design of Study 3 removed all explicit linguistic components from the experimental setting. Participants performed an "intuitive geography" task, estimating their distance from a purported city (cf. Alter & Oppenheimer, 2008).

We sought to avoid potential familiarity effects (Vitevitch & Luce, 1999) whereby a city might feel closer if its name sounded more familiar. Therefore, we generated two fictitious city names that were matched on all internal biphone probabilities (the likelihood that the first two sounds and the last two sounds of the word would co-occur in that position in an English word). Thus, we arrived at our experimental city names of Fleen (/flin/, containing a front vowel) and Floon (/flun/, containing a back vowel). Pilot testing confirmed that Fleen ($n = 30$, $M = 3.83$, $SD = 1.72$) and Floon ($n = 30$, $M = 3.30$, $SD = 1.82$) were equally plausible ($p > 0.2$) as city names.

We decided ahead of time on a minimum sample size of 25 participants for each of the two conditions and collected data until that number had been reached. We arrived at this sample size due to the statistical noise inherent in estimating an ambiguous magnitude of distance and in keeping with similar investigations (e.g., Liberman & Förster, 2009; Maglio & Trope, 2011). Because this study was conducted in the field, we collected data until the end of the day and accordingly stopped above the target.

4.2. Procedure

Fifty-five volunteers were recruited from the New York University student center to complete a survey ostensibly about geography (using a paradigm adapted from Alter & Oppenheimer, 2008). After participants had provided consent to participate, the research assistant handed to the participant – in individual sessions – an iPad tablet onto which the experimental materials had been pre-loaded (via a wireless internet connection to the same Qualtrics website as in Study 1). The electronically-presented instructions to participants read as follows: "We will present you with the name of a city in New York state and ask you to estimate – using your intuition – how far away it is from you here in New York City (in miles)." Next, to reduce variability in distance estimates, the electronic instructions read as follows: "As a point of reference, the greatest distance between New York City and any other city in New York state is about 400 miles." Participants were randomly assigned to estimate their distance in miles from either Fleen ($n = 28$) or Floon ($n = 27$). We predicted that, if vowels have an association with distance that functions independently of priming meta-linguistic knowledge, a city name containing a back vowel should be estimated as further away than a city with a front vowel containing name.

4.3. Results and discussion

As predicted, participants estimated Floon, NY as significantly further ($M = 201.93$ miles, $SD = 101.61$) than Fleen, NY ($M = 152.64$ miles, $SD = 63.52$), $t(43.36) = 2.15$, $p = 0.04$, $d = 0.58$ (Fig. 2). Thus, in a task lacking any overtly linguistic component, participants exhibited an intuitive association between vowels and spatial distance, estimating a greater distance between themselves and a city with a name containing a back versus a front vowel.

5. Study 4a and 4b: vowel and visual estimation of distance

Next, Study 4 tests the downstream consequences of sound-to-meaning associations. Because top-down expectations can bias even the perception of distance (Balcetis & Dunning, 2007, 2010), we considered whether a visible target would be perceived differently simply based on the vowel contained in its name. We manipulated the name given to an object and predicted that people would judge it to be spatially further when the name included a back (versus front) vowel.

5.1. Study 4a: verbal estimation

5.1.1. Stimuli

Here, we assessed the impact of object names – including a front or a back vowel – on how people visually perceive the distance separating themselves from that object. In this case, to control for the potential confound of familiarity, we manufactured an entirely novel object (i.e., a dog toy spray painted in a multitude of colors) as the target of estimation. Participants made verbal estimates of the physical distance between themselves and that object after they were told its name (i.e., one of several nonwords containing either a front or back vowel).

We generated a set of 6 experimental items for Study 4a. These items comprised 3 nonwords containing front vowels ("deeb" [dib], "nep" [nεp], and "keev" [kiv]) and 3 nonwords containing back vowels ("doab" [dob], "noop" [nup], and "koov" [kuv]). As in Study 3, we balanced the biphone probabilities across each front/back pair to avoid the potential confound of familiarity. We decided ahead of time on a minimum sample size of 20 participants for each of the six conditions in consideration of (1) the average sample size and effect size for verbal distance estimation studies and (2) in order to have sufficient power to detect any potential differences between the specific items in each of the object label conditions. Before this sample size was reached, a maintenance worker using a lawnmower in Washington Square Park destroyed the dog toy inadvertently and despite the best efforts of the research assistants to save it. Because other similar distance perception investigations have relied upon samples slightly under 20 participants per cell (e.g., Cole, Balcetis, & Dunning, 2013), we decided to perform our statistical analyses on the sample as it stood.

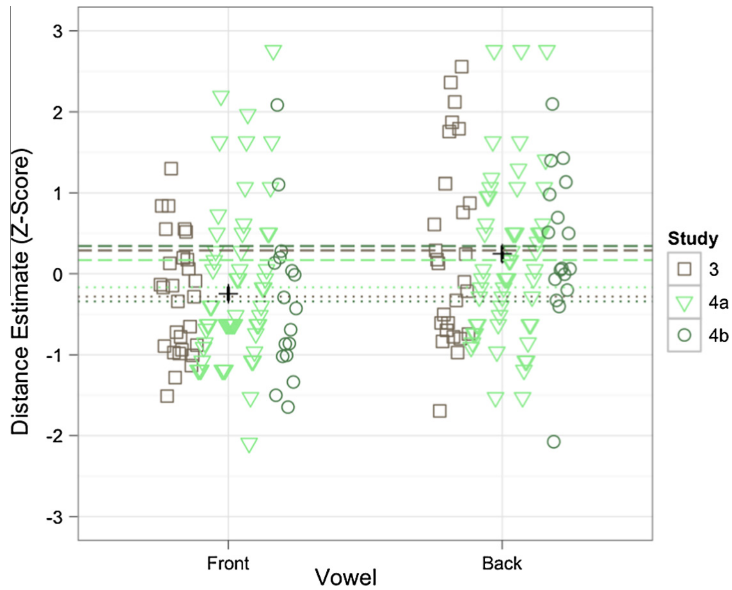

Fig. 2. Estimated distance as a function of target names containing either front or back vowels in Studies 3, 4a and 4b. *Note*. Distance estimates for individual participants in the three studies are plotted as z-scored values relative to other distance estimates within each study (values are jittered around each category for clarity). Means (horizontal black lines) and standard errors (±1, vertical black lines) for these z-scored distance estimates in each vowel type, collapsing across study, are overlaid, along with means for front (dotted lines) and back (dashed lines) vowels within each individual study.

5.1.2. Procedure

One hundred fourteen volunteers were recruited from Washington Square Park near New York University to complete a study ostensibly investigating sports psychology. After participants had provided consent to participate, the research assistant asked participants – in individual sessions – to stand at a particular location (marked with a piece of wood) facing a novel object (the painted dog toy) that the research assistant had positioned in a location 30 ft away from the marked location. The research assistant then verbally read the following instructions to participants: "Do you see the object over there? It is called a [object name]. Your task is to estimate how far away the [object name] is. Please make your estimate in feet. For your reference, this ruler is one foot long." At this point, the research assistant held a standard ruler in front of the participant, then said "When you're ready, let me know how far away the [object name] is," and recorded each individual participant's response on a piece of paper.

The research assistant inserted as the object name one of the six experimental stimuli. Participants were randomly assigned to one of two object label conditions (vowel location, front/back), and, within each of these, to one of three specific label items. In the front vowel condition, the object was labeled a "deeb", ($n = 18$), "nep", ($n = 19$), or "keev", ($n = 20$), while in the back vowel condition, the object was labeled a "doab" ($n = 18$), "noop" ($n = 19$), or "koov" ($n = 20$). We predicted that distance estimates would be greater for back vowel labels than front vowel labels.

5.1.3. Results and discussion

Prior to analysis, we excluded data from one participant with a distance estimate 3 standard deviations from the mean of the overall vowel location condition (cf. Balcetis & Dunning, 2010). The following pattern of results remains consistent if this outlier is included in the analyses.

We examined the effect of vowels on estimates of spatial distance using a mixed-effects ANOVA with vowel condition (front/back) as a fixed effect and individual item names nested in vowel condition as random factors. As predicted, vowels significantly impacted estimates of distance, $F(1,4) = 17.44$, $p = 0.01$. We observed no effect of names nested within condition, $F < 1$, suggesting that the results were not driven by any particular item(s). Accordingly, we collapsed across the individual names in our subsequent analyses. In line with our hypothesis, participants perceiving an object named with a back vowel saw it as physically farther ($M = 27.04$ ft, $SD = 8.79$) than participants for whom the object was named with a front vowel ($M = 23.59$ ft, $SD = 8.10$), $t(111) = 2.17$, $p = 0.03$, $d = 0.41$ (Fig. 2).

5.2. Study 4b: action-based estimation

5.2.1. Stimuli

Next, we sought to replicate Study 4a with a different distance estimation measure. Participants were instructed to throw a beanbag as close as possible to an object (i.e., a different dog toy from the one used in Study 4a) whose name included either a front or a back vowel. Having established no significant effect of individual item names, we chose one specific pair of names from Study 4a ("deeb" [dib], including a front vowel, and "doab" [dob], including a back vowel) for Study 4b.

We adopted this paradigm from Balcetis and Dunning (2010; Study 3a), who observed a notably larger effect size for this action-based measure of distance perception compared to verbal distance estimates. Accordingly, we decided ahead of time on a minimum sample size of 15 participants for each of the two conditions and collected data until that number had been reached. Because this study used an undergraduate participant pool, we collected data until the end of the day and accordingly stopped above the target.

5.2.2. Procedure

Thirty-four undergraduate students from New York University participated in the experiment in exchange for course credit. They participated in individual testing sessions in which they arrived at the research lab and were escorted by a research assistant to an empty hallway in the New York University psychology building for an ostensibly physical task. After participants provided consent to participate, the research assistant asked participants to stand on a particular location (marked with a piece of masking tape on a carpeted hallway) facing a novel object (the dog toy) that the research assistant had positioned 13 ft away from the marked location. The research assistant then verbally read the following instructions to participants: "Do you see the object over there? It is called a [object name]. Your task is to toss this beanbag so that it lands as close as possible to the [object name]." At this point, the research assistant handed a standard-sized beanbag to the participant, then said "When you're ready, go ahead and toss the beanbag toward the [object name]."

Participants were randomly assigned to condition ($ns = 17$) in which they were told that the object was named either "deeb" or "doab". After the participant threw the beanbag, the researcher used a measuring tape to measure the distance from the landing position of the center of the beanbag to the target in inches. Responses were recorded in negative terms when the beanbag fell short of the object and in positive terms when the beanbag fell beyond the object. We predicted that participants would overthrow the beanbag (i.e., obtain generally higher scores on our distance measure) when aiming at a target named with a back vowel relative to a target named with a front vowel.

5.2.3. Results and discussion

The distance from the landing position of the beanbag to the toy constituted each participant's distance estimate: Negative scores indicate that the beanbag landed between the toy and the participant and positive scores indicate that the beanbag landed beyond the toy. As expected, participants perceiving the object named "doab" saw it as significantly further ($M = 4.64$ in., $SD = 15.48$) than participants for whom the object was named "deeb" ($M = -6.61$ in., $SD = 15.74$), $t(32) = 2.10$, $p = 0.04$, $d = 0.72$ (Fig. 2). This action-based measure suggests that vowels, via their impact on the perception of spatial distance, can influence real behavior toward a target. Taken together, Study 4 suggests that the relationship between vowels and distance can affect even basic perceptual processes. This obtained independent of the modality of assessment (self-report or action-based measures) or any specific items used to manipulate vowel.

6. General discussion

The current investigation provides experimentally-based support for the sound symbolic association between vowels and spatial distance. Through converging evidence from 5 studies, we show how this association influences participants' intuitions about language meaning (Study 1), information processing and reaction times (Study 2), and estimates of distance to both imagined (Study 3) and real targets (Study 4). These effects held across a variety of controlled linguistic stimuli, suggesting that these results are not due merely to differences in a feeling of familiarity of the stimuli and are not dependent on specific linguistic items. Rather, similar results from a range of approaches support the existence of a robust, non-arbitrary, intuitive association between vowel backness and spatial distance.

More generally, these results make an important contribution in defining the scope of sound symbolic effects. Prior research has provided evidence for sound-to-meaning mappings in explicitly metalinguistic settings and/or on linguistic outcomes (e.g., intuitions about word meanings, word learning). As a result, it has remained largely unclear what effects these associations may have on non-linguistic outcomes. Here, we have provided evidence for sound-to-meaning mappings as made manifest on dependent variables that asked not about language per se (save for Study 1), but rather about outcomes resulting therefrom, suggesting that sound symbolic effects may operate independently of meta-linguistic knowledge. These results both reinforce the robustness of sound symbolic effects and provide a key extension of their scope, demonstrating that intuitive linkages between specific phonemes and meaning can have measurable effects on the representation of and behavior toward real (nonlinguistic) objects in the world. These studies suggest that intuitive mappings between meaning and even relatively subtle characteristics of language can exert substantial psychological effects.

While our results highlight the effects of sound-to-meaning mappings beyond linguistic outcomes, they remain restricted to judgments about referents with which participants had no prior experience. Thus, an open question for future research may ask whether (or to what extent, if at all) these associations hold for referents for which people hold preexisting knowledge. Our studies explored associations between specific phonemes (i.e., front versus back vowels) contained in *unfamiliar* nonwords and behavior toward *unfamiliar*, *novel* targets. Accordingly, these results have clear implications for applied contexts in which novel names are created for novel objects or in situations when people consider unfamiliar objects (Shrum & Lowrey, 2007; Yorkston & Menon, 2004). Future work might explore how these effects interact with prior semantic knowledge, such as when labels hold conventional meaning (i.e., real, familiar words) or when people have prior information associated with the referents (i.e., familiar objects).

Research on sound symbolism in general also awaits an answer to the underlying nature of its sound-to-meaning mappings. At a very basic level, it remains unclear what specific aspects of speech sounds themselves drive their associations with meaning, as language constitutes a multi-modal experience (i.e., words can be visually read, auditorily heard, and motorically spoken or subvocalized). Ours is the first experimental psychology-based investigation into the relationship between vowels and spatial distance within a community of speakers. Accordingly, the only prior speculation regarding the nature of this specific sound-symbolic relationship was offered by Tanz (1971). Tanz (1971) proposed something akin to an embodied account, whereby articulating the different types of vowels may resemble, at a motoric level, projection into space for back vowels or the keeping of something close for front vowels (see also Topolinski, Maschmann, Pecher, & Winkielman, 2014). However, even for the long-noted, robust sound-to-meaning mappings that predate our investigation (e.g., size, shape), there is little consensus on this issue, and the propositions have been various. These include aspects of sound quality, such as intrinsic pitch or frequency, sensory-motor mappings from articulatory motor gestures made during production, and aspects of the visual form of articulatory gestures (Ohala, 1984; Sapir, 1929). Any of these might be at play, and because the spectral qualities of language sounds (i.e., pitch, duration) are a direct consequence of articulatory actions within the resonator system created by the human vocal tract, the precise locus of these effects will likely be difficult, if not impossible and/or inappropriate, to isolate.

However, we believe the approach adopted in Study 1 – examining how the sound-symbolic effect may co-vary with specific aspects of vowel quality beyond a simple front/back distinction – may prove useful in future work to identify which specific linguistic features drive sound-symbolic effects. Specifically, we show in Study 1 that the apparent magnitude of the association between vowels and distance varies with the phonetic qualities that serve as cues to vowel frontness/backness. This approach thereby distinguishes the sound-symbolic effect from one that operates only via, for example, vowel rounding (a motoric experience of shaping the mouth into a more round shape during pronunciation, which tends to covary with vowel backness yet need not), showing instead a gating of the effect with F2 frequency. This result suggests that pitch – the auditory percept of sound frequency – may contribute to the relationship between vowel frontness/backness and distance (i.e., a higher F2 frequency, indicating a relatively front vowel, has a higher pitch). Indeed, sounds with higher pitch or frequency tend to have shorter wavelengths; shorter wavelengths are more directional and radiate less, whereas longer wavelengths radiate more and travel more diffuse routes. As a result, a listener hearing sounds produced at a greater physical distance will perceive primarily the longer wavelengths and lower frequencies of a sound because the high-frequency components of the sound dissipate sooner; if internalized, this co-variation could provide an origin for our effects. While our results highlight pitch as an account for the specific linkage between spatial distance and vowel backness, other sound symbolic associations may well originate from a variety of linguistic features, ranging from other aspects of sound quality (such as length or voicing) to embodied linkages between meanings and sensorimotor experiences (Perniss & Vigliocco, 2014). Examining how the magnitude of sound-symbolic associations covaries with specific linguistic features of speech sounds may thus be a valuable tool in uncovering the origins of sound-symbolic effects more generally.

Given the existence of cross-modal relationships, one cannot help but wonder why the human brain affords these *specific* cross-modal linkages. One intriguing proposal is that cross-activation of sensory and motor maps may lead to natural biases mapping certain sound profiles to certain object properties. From this view, sound symbolic effects are just one example of a broader category of the sensory-to-motor synesthesia pervasive throughout the normal population (Ramachandran & Hubbard, 2001). This would imply that such effects represent a direct mapping between sound and aspects of the world that is not dependent on linguistic experience or the particular idiosyncrasies of one's native language. Indeed, recent evidence demonstrating such synesthetic associations in pre-linguistic infants – for instance, between the pitch of a tone and physical thickness (Dolscheid, Hunnius, Casasanto, & Majid, 2014) and between classes of phonemes and shape (Ozturk, Krehm, & Vouloumanos, 2013) – supports the possibility that these sound-meaning correspondences are not dependent on linguistic experience (for a review, see Imai & Kita, 2014).

These findings support the contention that at least some sound symbolic associations are not learned over the course of one's lifetime through repeated pairings of (vowel-sound-containing) names referring to certain types of objects, but rather reflect an inherent and perhaps universal human linguistic phenomenon. Our results – to the extent that they identify sound symbolic effects as robust, intuitive associations exhibited outside of explicitly linguistic contexts – are consistent with this view. However, Baxter and Lowrey (2011) have identified a developmental difference whereby a previously-established sound symbolic association in adults (i.e., a preference to pair names including a front vowel with an "icy and sweet" ice cream and names including a back vowel with a "creamy and rich" ice cream; Lowrey & Shrum, 2007) is more manifest among older children (i.e., 10–12 years of age) than younger children (i.e., 6–7 years of age).

These results, to the extent that they show a developmental trajectory in sound-symbolic effects within older children, may appear to support an experiential based account. However, considered alongside previous work showing sound-symbolic associations in infants, we note that they diverge both (1) in the different stages of development targeted by the separate investigations (i.e., infants versus children, coupled with a host of cognitive developmental differences) and (2) in the tasks required of the young participants (i.e., looking time versus explicit choice). The latter may simply suggest that certain types of sound-symbolic associations with more complex concepts (e.g., that ice cream is creamy) may emerge more gradually than basic perceptual concepts (e.g., shape). More broadly, that Baxter and Lowrey (2011) do nonetheless observe a pattern of results (though weaker) that is consistent with sound-symbolic inferences among children 6–7 years of age might very well support the notion that these effects are multiply determined and, while they may accentuate with age, cannot sustain a purely experiential account.

If such cross-modal linkages between sound and physical properties evolved prior to the development of language, an intriguing possibility proposed by Ramachandran and Hubbard (2001) is that this specific type of sensory-motor synesthesia may have played an important role in language evolution. The approach taken in the present investigation – which attempts to elucidate the nature of sound symbolic effects in tasks that are not specifically "about" language and on non-linguistic outcomes – could lend further insight to these hypotheses. Similarly, it has been proposed that sound symbolism, along with other examples of iconicity in language more broadly, may have played an important role in language evolution and may continue to play an important role in language development by assisting with the mapping of linguistic form to referents (Perniss et al., 2010; Perniss & Vigliocco, 2014). Further work attempting to outline the domains, scope, and underlying roots of sound-symbolic linkages may thus provide a window onto the very origins of our linguistic capacity itself.

Acknowledgements

This research was supported by a University of Toronto Connaught New Researcher Grant to the second author and a NSF Grant #BCS-1053128 to the fifth author. The authors thank Catherine Lee and Michael Feder for help with execution of the studies.

References

Alter, A. L., & Oppenheimer, D. M. (2008). Effects of fluency on psychological distance and mental construal (or why New York is a large city, but *New York* is a civilized jungle). *Psychological Science, 19*, 161–167.

Balcetis, E., & Dunning, D. (2007). Cognitive dissonance and the perception of natural environments. *Psychological Science, 18*, 917–921.

Balcetis, E., & Dunning, D. (2010). Wishful seeing: Desired objects are seen as closer. *Psychological Science, 21*, 147–152.

Bar-Anan, Y., Liberman, N., Trope, Y., & Algom, D. (2007). Automatic processing of psychological distance: Evidence from a Stroop task. *Journal of Experimental Psychology: General, 136*, 610–622.

Baxter, S., & Lowrey, T. M. (2011). Phonetic symbolism and children's brand name preferences. *Journal of Consumer Marketing, 28*, 516–523.

Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology, 50*, 388–393.

Catford, J. C. (1988). *A practical introduction to phonetics*.Oxford: Clarendon Press.

Cole, S., Balcetis, E., & Dunning, D. (2013). Affective signals of threat increase perceived proximity. *Psychological Science, 24*, 34–40.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass, 6*, 654–672.

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science, 25*, 1256–1261.

Hockett, C. F. (1958). A course in modern linguistics. *Language Learning, 8*, 73–75.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*. http://dx.doi.org/10.1098/rstb.2013.0298.

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

Koriat, A., & Levy, I. (1977). The symbolic implications of vowels and of their orthographic representations in two natural languages. *Journal of Psycholinguistic Research, 6*, 93–103.

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

Ladefoged, P. (2006). *A course in phonetics* (5th ed.)Boston, MA: Thomson Wadsworth.

Liberman, N., & Förster, J. (2009). Distancing from experienced self: How global-versus-local perception affects estimation of psychological distance. *Journal of Personality and Social Psychology, 97*, 203–216.

Liu, C., Jin, S.-H., & Chen, C.-T. (2014). Durations of American English vowels by native and non-native speakers: Acoustic analyses and perceptual effects. *Language and Speech, 57*, 238–253.

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

Maglio, S. J., & Trope, Y. (2011). Scale and construal: How larger measurement units shrink length estimates and expand mental horizons. *Psychonomic Bulletin & Review, 18*, 165–170.

Maurer, D., Pathman, T., & Mondloch, C. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science, 9*, 316–322.
Nygaard, L., Cook, A., & Namy, L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112*, 181–186.
Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F_0 of voice. *Phonetica, 41*, 1–16.
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.
Parault, S., & Schwanenflugel, P. (2006). Sound-symbolism: A piece in the puzzle of word learning. *Journal of Psycholinguistic Research, 33*, 329–351.
Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology, 1*, 1–15.
Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*. http://dx.doi.org/10.1098/rstb.2013.0300.
Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE, 10*. http://dx.doi.org/10.1371/journal.pone.0137147.
Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia – A window into perception, thought and language. *Journal of Consciousness Studies, 8*, 3–34.
Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 Nonwords: The arc nonword database. *The Quarterly Journal of Experimental Psychology: Section A, 55*, 1339–1362.
Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 225–239.
Saussure, F., Bally, C., Sechehaye, A., Riedlinger, A., & Baskin, W. (1966). *Course in general linguistics*.New York: McGraw-Hill.
Shrum, L. J., & Lowrey, T. M. (2007). Sounds convey meaning: The implications of phonetic symbolism for brand name construction. In T. M. Lowrey & L. J. Shrum (Eds.), *Psycholinguistic phenomena in marketing communications* (pp. 39–58). Mahwah, NJ: Erlbaum.
Tanz, C. (1971). Sound symbolism in words relating to proximity and distance. *Language and Speech, 14*, 266–276.
Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology, 64*, 2392–2404.
Topolinski, S., Maschmann, I. T., Pecher, D., & Winkielman, P. (2014). Oral approach–avoidance: Affective consequences of muscular articulation dynamics. *Journal of Personality and Social Psychology, 106*, 885–896.
Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.). *Universals of human language* (Vol. 2, pp. 525–568). Stanford, CA: Stanford University Press.
Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374–408.
Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language, 93*, 10–19.
Yorkston, E., & Menon, G. (2004). A sound idea: Phonetic effects of brand names on consumer judgments. *Journal of Consumer Research, 31*, 43–51.
Yoshida, H. (2012). A cross-linguistic study of sound symbolism in children's verb learning. *Journal of Cognition and Development, 13*, 232–265.