

Some consonants sound curvy: Effects of sound symbolism on object recognition

Mark E. Aveyard

Published online: 26 September 2011
© Psychonomic Society, Inc. 2011

Abstract Two experiments explored the influence of consonant sound symbolism on object recognition. In [Experiment 1](#), participants heard a word ostensibly from a foreign language (in reality, a pseudoword) followed by two objects on screen: a rectilinear object and a curvilinear object. The task involved judging which of the two objects was properly described by the unknown pseudoword. The results showed that congruent sound-symbolic pseudoword–object pairs produced higher task accuracy over three rounds of testing than did incongruent pairs, despite the fact that “hard” pseudowords (with three plosives) and “soft” pseudowords (with three nonplosives) were paired equally with rectilinear and curvilinear objects. [Experiment 2](#) reduced awareness of the manipulation by including similar-shaped, target-related distractors. Sound symbolism effects still emerged, though the time course of these effects over three rounds differed from that in [Experiment 1](#).

Keywords Sound symbolism · Iconicity · Language comprehension · Object recognition · Associative learning

Recent work on sound symbolism has revived a nearly century-old debate regarding possible constraints on the arbitrariness of language (for an overview, see Nuckolls, 1999). To a large extent, the arbitrariness of language is undeniable: Translating words between languages reveals widespread variation in the phonemes used to encode identical or similar concepts (e.g., *dog* in English, *inu* in

Japanese, *nay* in Tamil). Historically, though, linguistics and cognitive science have regarded arbitrariness as an unbounded linguistic principle, applicable in all circumstances or with only rare exceptions (such as onomatopoeia). Early research on sound symbolism failed to undermine confidence in this idea (as noted by Taylor & Taylor, 1965), but a recent body of research has uncovered a more expansive and comprehensive role for this form of iconicity in language. Several studies have shown, for example, that sound symbolism can facilitate word learning in both adults and children (for words in a foreign language, see Nygaard, Cook, & Namy, 2009; for unfamiliar words in a native language, see Parault & Schwanenflugel, 2006).

Numerous studies in sound symbolism have used a method that requires participants to associate pseudowords with realistic or abstract objects. In an early experiment of this kind, Sapir (1929) asked English-speaking participants to match pseudowords with objects of different sizes: *mil* and *mal* were compared to large and small tables, with participants choosing *mil* as a more appropriate term for the small table, as compared to *mal*. Sapir attributed this choice to associations between magnitude and vowel categories among English speakers. Around the same time, Köhler (1929) developed a similar paradigm by contrasting angular and curvy line-drawings with the pseudowords *taketa* and *maluma*, the latter representing an appropriate fit for curvy figures due to the open articulation involved in its pronunciation.

Köhler’s (1929) use of abstract, unfamiliar figures set an influential precedent for future research. More rigorous versions of this paradigm have appeared since (Kovic, Plunkett, & Westermann, 2010), but with a predominant focus on the role of vowels (Lowry & Shrum, 2007; Maurer, Pathman, & Mondloch, 2006). Despite frequent references to Köhler’s original work, the literature on consonant sound symbolism remains relatively sparse. In

M. E. Aveyard (✉)
Department of International Studies,
American University of Sharjah,
Sharjah, United Arab Emirates
e-mail: maveyard@aus.edu

one novel approach, Westbury (2005) presented “hard” and “soft” consonant–vowel–consonant (CVC) pseudowords (such as *tabe* and *neem*) inside boxes surrounded by curved or straight lines while participants completed a lexical decision task. Accurate responses were facilitated by congruency between the pseudoword and the surrounding visual information (e.g., the hard pseudoword *kide* surrounded by edgy visual information enabled faster decisions). In another study on consonant sound symbolism, Imai, Kita, Nagumo, and Okada (2008) created such pseudowords as *chokatoka* and *hyaihyai* and paired these stimuli with videos depicting a character walking in a smooth or choppy manner. Two- and three-year-old Japanese children learned to generalize the meanings of the novel verbs only when the training phase involved a sound-symbolic match between the verb and the depicted action (when *chokatoka* was paired with a choppy manner of walking).

Problematically, in these and other studies, there has been no consensus method for determining the congruency mappings between consonants and visual stimuli. Imai et al. (2008) pretested stimuli with British participants (English-speaking adults) and Japanese participants (adults and children), thereby providing a basis for congruency mappings in two subsequent experiments with Japanese children. The most common approach, however, has involved the application of articulatory phonetics to the properties of visual stimuli, usually based on previous research and researchers’ own judgments of congruency. Westbury (2005) created a set of 60 pseudowords, divided evenly between three categories based on manner of articulation: plosives (*buke*), continuants (*nole*), and mixed pseudowords (*furt*).

The stimuli chosen by Kovic et al. (2010) presented an interesting contrast with those of Westbury (2005) and Köhler (1929). Their categorization task involved two pairs of pseudowords (*mot* and *riff* in one experiment, *dom* and *shick* in another) with a focus on the articulatory features of the vowels (the rounded *o* in *mot* and the unrounded *i* in *riff*), consistent with the mappings used by Maurer et al. (2006). In both studies, the rounded vowel pseudowords produced congruency effects with rounded visual stimuli. Interestingly, it is not clear how to classify their stimuli within Köhler’s original analysis, which focused on the sound symbolism of consonants rather than vowels. The consonants in *riff* correspond more to the nonstop consonants of *maluma* than to the stop-filled *taketa*, but Kovic et al. considered *riff* as a congruent match with angular objects, not with the curvy objects Köhler matched with *maluma*. Likewise, whereas *mot* contains a rounded vowel, it also contains a stop consonant and a bilabial nasal; Kovic et al. paired *mot* with a curvilinear object, despite consonant features that Köhler would probably classify as angular in sound-symbolic terms.

Kovic et al. (2010) explained the congruency effects of the *mot-riff* experiment in terms of vocalic sound symbolism. But there could be a clash between vowel and consonant iconicity in these stimuli, raising the question whether the faster response times for certain stimulus pairs could be attributed to vocalic congruency effects. Similar results obtained in their second experiment, using *dom* and *shick*, would seem to support the presumed congruency relationships from the first experiment, since both of these pseudowords contain stop consonants, possibly providing some counterbalance between the two stimuli. Even so, Kovic et al. intentionally used voiced consonants in *dom* and unvoiced consonants in *shick* to maximize sound-symbolic effects beyond the influence of vowels (though it’s not explained why a voiced consonant should be more congruent with curvy stimuli than an unvoiced consonant is).

In addition to possible problems with conflicting sound-symbolic influences, studies on consonant sound symbolism typically provide participants with a limited set of stimuli. Kovic et al. (2010) used only two pseudowords (a necessity, given the difficulty of the categorization task they employed). Similarly, Imai et al. (2008) used only six pseudowords with six pairs of matching–mismatch videos. The iconicity of these six pseudowords seemed to result from a combination of universal and language-specific sound-symbolic influences, since British adults confirmed the congruency relationships, but not nearly as strongly as the Japanese adults, who seemed sensitive to Japanese-specific congruencies (described by the authors for each pseudoword used in the study). It is not clear, then, what kinds of sound symbolism contributed to the observed congruency effects: the articulatory properties of a class of consonants (e.g., stops or plosives), the unique properties of individual phonemes, the particular collection of consonants within each pseudoword, the order in which these consonants appeared, or some combination of these factors.

Among recent studies, only Westbury (2005) seems to have used a wider selection of phonemes to produce congruency effects within two articulatory categories. That study involved five plosives and four continuants, all appearing in both CVC consonant positions. In that study, the visual context surrounding words influenced lexical decision times. Would similar effects be found in a Köhler-style paradigm involving object recognition? Would these effects emerge even in a difficult task in which congruency relationships provided no advantage for learning the associations between unfamiliar words and objects? The two experiments here explored how adults use consonant airflow as a mechanism for associating novel words with novel objects. These experiments also represent the first sound symbolism investigation to report data on multiple rounds of testing, allowing for an analysis of the time course of these effects.

Experiment 1

Participants tried to learn word–picture pairs composed of unfamiliar English pseudowords and unfamiliar 3-D objects. On each trial, they first heard a word presented auditorily. Then two shapes appeared on screen, one on the right side and one on the left. Participants were asked to click on one of the two shapes using the mouse (one shape was a target and the other a distractor). No criteria were provided beforehand to guide their decisions, but feedback was provided after each trial to help participants improve their accuracy on specific word–picture pairs across three rounds of testing. Each word contained either plosive phonemes, in which airflow is temporarily obstructed (*kuh-der-pai*), or nonplosive phonemes, in which airflow is not obstructed (*fuh-lih-sai*). Each word was paired with a rectilinear or curvilinear 3-D target object (see Fig. 1 for sample stimuli). Sound symbolism predicts that participants would be faster to learn word–picture pairs with matching phoneme categories and 3-D shape compositions (plosives with rectilinear objects and nonplosives with curvilinear objects). If the natural relationship between sound and meaning is entirely arbitrary, sound information should not enable faster learning of matching word–picture pairs. Importantly, each participant encountered congruent and incongruent pairs in equal numbers, rendering sound-symbolic relationships irrelevant to overall improvement. For example, 16 rectilinear targets were paired with 8 “hard” pseudowords and 8 “soft” pseudowords. Therefore, participants could not improve accuracy on these targets if they used phonological information to try to determine the associated shape.

Method

In Experiment 1, framed as a foreign language study, 44 English-fluent undergraduates at the American University of Sharjah in the United Arab Emirates (who received course credit for their participation) were introduced to 32 pseudowords (ostensibly from “a dialect of Vietnamese”) by listening to the words individually over headphones. Participants were asked not to pronounce the words “so that

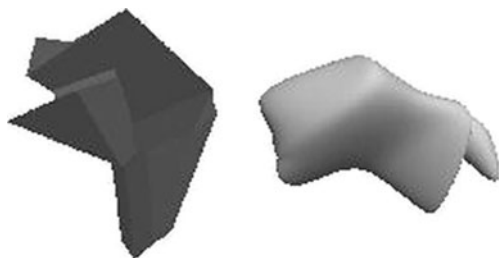


Fig. 1 Examples of rectilinear and curvilinear objects (left and right, respectively)

other participants would not be distracted.” Each audio file was generated by a voice simulation program (AT&T Labs Natural Voices Text-to-Speech Demo) to provide consistent audio quality. The experimenter then explained that the 3-D objects in the impending memory test had previously been viewed by students at a (fictitious) university in Vietnam, who provided 32 words from their language that described these pictures. Participants would listen to a word and then decide which of two 3-D objects correctly belonged to that word by clicking on that object with the mouse, with subsequent “Correct” and “Wrong” feedback as a guide for learning. The instructions stated that there would be “several rounds” of testing to measure improvement. The test was repeated three times, with 32 randomized trials in each round. In each round, targets were paired with different distractor objects, so that each trial in the experiment presented participants with a novel comparison between the two stimuli on screen.

The experiment was based on a within-subjects design, with 16 rectilinear and 16 curvilinear 3-D objects created in Silo Professional. Two categories of pseudowords were used: 16 “hard” pseudowords, such as *kuh-der-pai* /kʌ dɜr paɪ/ and *dee-gay-tau* /di geɪ təʊ/, and 16 “soft” pseudowords, such as *fuh-lih-sai* /fʌ li saɪ/ and *reh-sai-lau* /rɛ saɪ laʊ/. Hard pseudowords contained six plosives (/p/, /b/, /t/, /d/, /k/, and /g/); soft pseudowords contained six nonplosives (/l/, /w/, /r/, /s/, /f/, and /h/).

Each word contained three consonants, with each phoneme appearing in different $C_1V_1-C_2V_2-C_3V_3$ positions across the 16 pseudowords within its category. For example, /d/ appeared as C_1 three times, C_2 three times, and C_3 two times; /k/ appeared as C_1 two times, C_2 three times, and C_3 four times. A wide array of vowels (front, back, high, and low) were used to minimize interference from vocalic sound symbolism: /i/, /ɪ/, /u/, /æ/, /ɛ/, /oʊ/, /ʌ/, /eɪ/, /ɔɪ/, /ɜr/, /aɪ/, and /aʊ/. The vowels also appeared in different positions across pseudowords: For example, /i/ appeared as V_1 one time, V_2 two times, and V_3 two times within both the plosive and nonplosive stimulus groups, while /eɪ/ appeared as V_1 two times, V_2 one time, and V_3 two times in both stimulus groups. Only once did a vowel appear more than two times in one position. Some vowels were used more than others: /eɪ/ appeared seven times in both groups, while /ɛ/ appeared only twice. More importantly, vowels appeared in nearly equal numbers between stimulus groups and in nearly equal positions. Among 48 vowel positions in the nonplosive pseudoword group, 44 occurrences had the same vowel in the same position as in the plosive pseudoword group. In the remaining 4 occurrences, comprehension problems in pretesting required changing the vowels between groups (/ɪ/, for example, appeared once in the plosive stimulus set but twice in the nonplosive group, whereas /aʊ/ appeared twice in the plosive group).

Two counterbalanced lists were generated such that a word associated with a rectilinear object in one list was paired with a curvilinear object in the other list. The experiment was executed using E-Prime by Psychology Software Tools.

Results

Trials with response times greater than two standard deviations from the within-cell means were excluded (4% of all responses). The low overall accuracy in this task ($M = .54$) underscores the difficulty of learning word–picture pairs with unfamiliar, similar-looking shapes and unfamiliar, similar-sounding words in only three rounds of testing. Even so, the congruent word–picture pairs ($M = .57$, $SD = .10$) provided a significant advantage in object recognition as compared to incongruent pairs ($M = .50$, $SD = .08$): $MD = .07$, $MD\ SD = .11$; $t(43) = 4.04$, $p < .001$, 95% CI = [.03, .10], $g = .77^1$ (Fig. 2). The experiment included three rounds to test whether participants would learn to ignore sound-symbolic information in order to improve their accuracy (or to use such information only on congruent trials). The congruency effect faded from Round 1 [$MD = .08$, $MD\ SD = .19$; $t(43) = 2.82$, $p = .007$, $g = .62$] to Round 3 [$MD = .03$, $MD\ SD = .20$; $t(43) = 1.05$, $p = .30$, $g = .20$] (Fig. 2). Overall accuracy improved from Round 1 ($M = .50$) to Round 3 ($M = .57$), with 68% of participants showing improvement. This improvement emerged more strongly on incongruent trials [$MD = .09$, $MD\ SD = .21$; $t(43) = 2.69$, $p = .01$, $g = .62$] than on congruent trials [$MD = .04$, $MD\ SD = .17$; $t(43) = 1.38$, $p = .17$, $g = .25$].

The design of Experiment 1 did not prevent participants from using non-sound-symbolic information to learn the word–picture pair associations. Indeed, the improvement shown on incongruent trials from Round 1 to Round 3 indicated that participants did use or develop non-sound-symbolic strategies. Responses to postexperiment questions also supported this interpretation. Four increasingly specific postexperiment questions onscreen asked participants about the strategies that they used to memorize the word–picture pairs, beginning with an open-ended question about the purpose of the experiment. Although most of the responses

were brief and 2 of 44 participants provided no responses, the data showed that conscious sound-symbolic strategies were used by some participants. Collapsing responses across questions, the strategies reported by participants included the following: using consonant sound symbolism ($N = 17$; e.g., “words that sounded slightly soft mostly had curved shapes”), using nonspecified sound symbolism ($N = 9$; e.g., “the pronunciation of the words reminds me of certain shapes”), using vowel sound symbolism ($N = 8$), relating the pseudoword to a real word ($N = 7$), relating the shape to a known object ($N = 5$), focusing on a specific part of an object ($N = 3$), focusing on one particular vowel or consonant ($N = 3$), rote memorization ($N = 3$), and random guessing ($N = 3$), with some participants reporting more than one strategy.

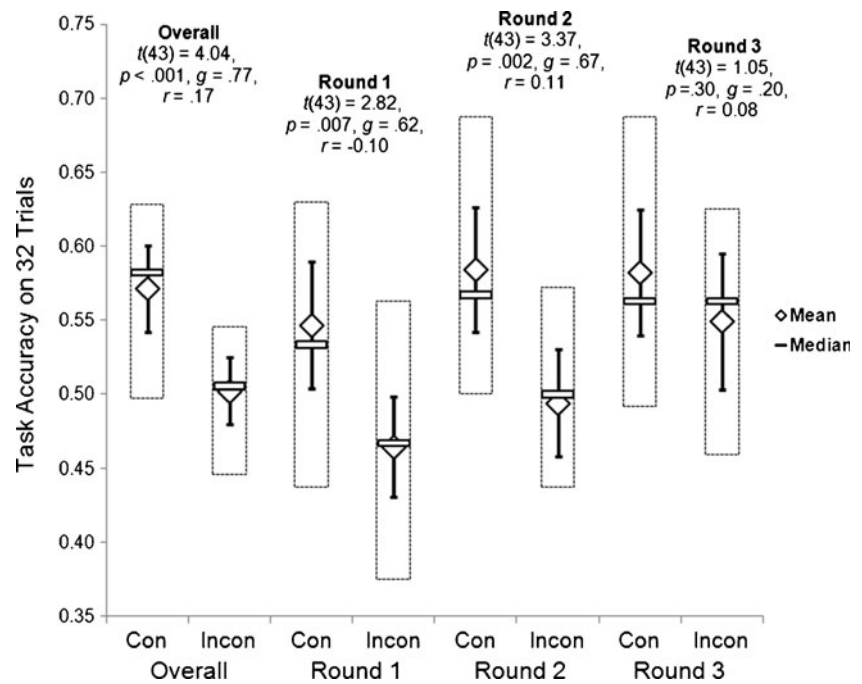
Did the use of a sound-symbolic strategy affect performance? After excluding participants who reported using a nonspecific sound-symbolic strategy ($N = 9$), the remaining participants were divided according to their use or nonuse of consonant-focused sound symbolism, producing two groups of nearly equal size: consonant-focused participants ($N = 17$) and non-consonant-focused participants ($N = 18$). There was a reliable congruency advantage for consonant-focused participants [$MD = .11$, $SD = .11$; $t(16) = 4.17$, $p < .001$, $g = 1.36$], but not for non-consonant-focused participants [$MD = .03$, $SD = .12$, $t(17) < 1$, $p = .35$, $g = 0.31$].

An alternative explanation for the main effect could emphasize the accuracy feedback received on each trial as the basis for accuracy improvement across rounds, without making any reference to the use of sound symbolism in learning the word–picture pairs. This alternative interpretation, though, cannot account for the significant accuracy differences observed between congruent and incongruent trials in Rounds 1 and 2. If participants only used the onscreen feedback for developing a response strategy, without incorporating sound symbolism information, this would not explain why congruent trials would show an advantage in Round 1, nor would it explain why they would continue to show this advantage in Round 2 after receiving feedback on half of the trials in Round 1 that undermined a general sound-symbolic approach.

Even so, if participants used a sound-symbolic approach in Round 1, they must subsequently have adopted one of two strategies to improve their accuracy in Rounds 2 and 3: Use a sound-symbolic approach only for congruent word–picture pairs, while developing an alternative approach for incongruent pairs, or develop an alternative approach for both congruent and incongruent pairs. It could then be argued that participants began with a sound-symbolic strategy that produced the initial congruency advantage in Round 1, then switched to a general non-sound-symbolic strategy that would increase overall accuracy by Round 3.

¹ MD here refers to the difference between the means; $MD\ SD$ refers to the standard deviation of the difference between the means. The accuracies on congruent and incongruent trials were not significantly correlated, $r(44) = .17$, $p = .28$. Normality tests indicated normal distributions for both congruent trials ($K^2 = 2.36$, $p = .31$) and incongruent trials ($K^2 = 1.12$, $p = .57$). The outlier removal procedure had almost no effect on the main comparison; an analysis without removal showed similar results, $t(43) = 3.99$, $p < .001$, $g = .77$. Additionally, an analysis by items reflected the results by participants: congruent pairs ($M = .57$, $SD = .10$) versus incongruent pairs ($M = .50$, $SD = .08$), $t(31) = 3.37$, $p < .01$, $g = .77$.

Fig. 2 Experiment 1 overall accuracy across all rounds, along with accuracy by round. Vertical bars represent 95% confidence intervals, and boxes represent the 1st and 3rd quartiles. Con, congruent trials; Incon, incongruent trials



Possibly this could result from non-consonant-focused participants developing a successful general strategy that overcame the limitations of a sound-symbolic approach, with consonant-focused participants failing to modify their initial sound-symbolic strategy. But in fact, consonant-focused participants showed a large improvement from Round 1 ($M = .50$) to Round 3 ($M = .59$), while non-consonant-focused participants showed a smaller improvement ($M = .48$ in Round 1 to $M = .52$ in Round 3).

Importantly, consonant-focused participants increased their performance not only on congruent trials ($M = .56$ in Round 1 to $M = .63$ in Round 3), but also on incongruent trials ($M = .45$ in Round 1 to $M = .54$ in Round 3), indicating that they adopted a strategy that enabled them to improve their accuracy for incongruent word–picture pairs while also increasing their accuracy for congruent pairs. Possibly they could have learned to ignore sound symbolism altogether by adopting a general strategy that enabled improvement in both conditions. However, if that were the case, non-consonant-focused participants should have had an advantage over consonant-focused participants from the start, in Round 1, by not using sound symbolism at all (given that it could not be effective as a general strategy). Yet these participants were not significantly more accurate in Round 1 ($M = .48$) than were consonant-focused participants ($M = .50$). At the very least, non-consonant-focused participants should have been able to adopt a general strategy that generated improvement in both conditions across all three rounds. But this did not happen: Non-consonant-focused participants showed improvement on incongruent trials (from $M = .46$ in Round 1 to $M = .53$ in

Round 3) but made virtually no improvement on congruent trials ($M = .50$ in Round 1 to $M = .51$ in Round 3).

To support the above interpretation, there should be some evidence that at least some participants used more than one strategy. Although the postexperiment questions probably do not show the full range of strategies employed for each participant, they do reveal that some participants used multiple strategies during the task, with 12 participants reporting using more than one strategy in the experiment. For example, 1 participant wrote, “At first I was trying to match the sound of the word to what the shape would look like, then I tried memorizing the sounds and the images.” Another participant reported, “I tried linking different words to existing words of my language. Also I linked each word to whether it is sharp or smooth edged.” To summarize, there is evidence indicating that participants gradually approached congruent and incongruent trials with different learning strategies, with sound symbolism providing an initial accuracy advantage for those who reported using it during the experiment.

Experiment 1 also showed that different kinds of phonological and perceptual stimuli produce sound-symbolic effects that differ in strength and development over time. The overall congruency advantage was stronger for pictures associated with nonplosive pseudowords [$MD = .09$; $t(43) = 2.14$, $p = .04$, $g = .54$] than for pictures associated with plosive words [$MD = .05$; $t(43) = 1.63$, $p = .11$, $g = .38$]. Overall, congruency advantages emerged for both rectilinear objects [$MD = .07$; $t(43) = 2.77$, $p = .01$, $g = .45$] and curvilinear objects [$MD = .07$; $t(43) = 3.46$, $p < .01$, $g = .50$]. Any conclusions drawn from round-by-round

data were limited by the lower observations per cell (7.6 on average per condition per round) and greater variability ($SD = .21$ on average per condition, as compared to $SD = .15$ per condition for the overall analysis). But it is interesting to note that plosive pseudowords produced a congruency advantage only in Round 1 ($MD = .11$ in Round 1 vs. $MD = -.01$ in Round 3), whereas nonplosive pseudowords produced congruency advantages across all three rounds ($MD = .05$ in Round 1, $MD = .15$ in Round 2, and $MD = .07$ in Round 3). Also, while rectilinear objects produced consistent congruency advantages across all three rounds ($MD = .06$ in Round 1, $MD = .08$ in Round 2, and $MD = .05$ in Round 3), the congruency advantage for curvilinear objects disappeared in the last round ($MD = .10$ in Round 1, $MD = .11$ in Round 2, and $MD = .01$ in Round 3).

Experiment 2

The results from Experiment 1 provided further evidence that consonants play an important role in sound symbolism with respect to object recognition in a word–picture association task. Nonetheless, it could be argued that Experiment 1 highlighted the difference in shape information (rectilinear versus curvilinear) by providing two response options that differed only along this dimension. A similar criticism of Sapir's (1929) original work was offered by Bentley and Varon (1933), who argued that Sapir's sound symbolism effects emerged partly from the use of a forced-choice task with only two objects contrasting on a single dimension. Thus, in modifying the design of Experiment 1 above, the previously observed congruency advantages could possibly be eliminated by reducing awareness of the shape attributes that relate to sound symbolism.

Experiment 2 tested the persistence of the pattern observed in Experiment 1 by providing four response options: a target (congruent or incongruent), a target-related distractor (in the same shape category), and two target-unrelated distractors (in the other shape category). For example, the pseudoword *kuh-der-pai* would be followed by a display with four objects: a congruent target (rectilinear object), a target-related distractor (a different rectilinear object), and two unrelated distractors (curvilinear objects), counterbalanced across participants. This design reduced the likelihood that participants would use only rectilinear–curvilinear contrasts in their decision-making in Round 1. Using that strategy would eliminate two response options, but not the other two. Thus, the decision between the two remaining options must be based on some criteria other than the linearity of the shapes. Otherwise, the method for this experiment was identical to that for Experiment 1, with 58 new participants.

Results

As in Experiment 1, trials with response times greater than two standard deviations from the within-cell means were excluded (4% of all responses). Even with other information available for selecting targets (such as the shading of targets and target-related distractors, or the orientation of the same stimuli), a sound symbolism advantage emerged again for congruent trials ($M = .29$, $SD = .07$) over incongruent trials ($M = .24$, $SD = .07$): $MD = .04$, $MD\ SD = .09$; $t(57) = 3.56$, $p = .001$, 95% CI = [.02, .07], $g = .67^2$ (Fig. 3). Pictures associated with nonplosive pseudowords again provided the main force behind the congruency effects [$MD = .09$; $t(57) = 3.72$, $p < .001$, $g = .77$], as compared to plosive pseudowords [$MD = .01$; $t(57) < 1$]. As in Experiment 1, congruency effects emerged for both rectilinear targets [$MD = .05$; $t(57) = 2.95$, $p = .004$, $g = .50$] and curvilinear targets [$MD = .04$; $t(57) = 2.55$, $p = .01$, $g = .35$].

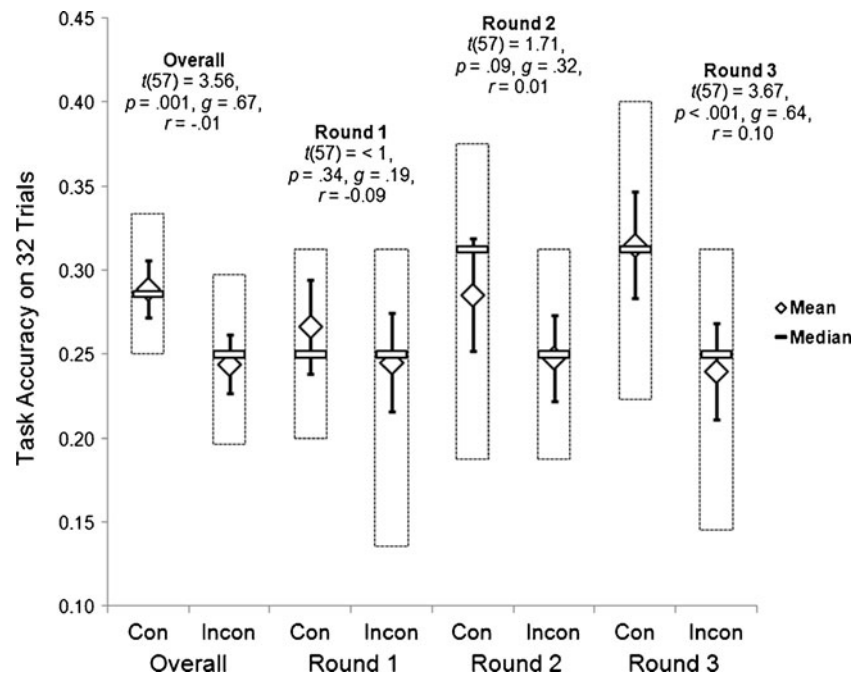
While there was only a small improvement in accuracy from Round 1 ($M = .26$) to Round 3 ($M = .28$), accuracy on congruent trials increased significantly [$MD = .05$, $MD\ SD = .16$; $t(57) = 2.28$, $p = .03$, $g = .42$], as compared to incongruent trials [$MD = .00$, $MD\ SD = .14$; $t(57) < 1$, $p = .77$, $g = .05$]. In other words, unlike participants in Experiment 1, participants in Experiment 2 failed to develop a strategy that would be effective for improving accuracy on incongruent trials. The improvement on congruent trials, however, indicated that at least some participants were able to use sound symbolism selectively on those trials. As a result, the congruency advantage increased across rounds, from Round 1 [$MD = .02$, $MD\ SD = .16$; $t(57) < 1$, $p = .34$, $g = .19$] to Round 3 [$MD = .08$, $MD\ SD = .16$; $t(57) = 3.67$, $p < .001$, $g = .63$], essentially inverting the pattern observed in Experiment 1, where congruency effects dropped off by Round 3 (Fig. 3). Since Round 1 involved trials without prior feedback, target and target-related selection rates could be combined to represent the total sound-symbolic selection rate relative to target-unrelated trials. This analysis showed virtually no difference between these conditions: $MD = -.03$; $t(57) < 1$, $p = .42$, $g = .22$.³

Why did this shift occur toward sound symbolism from Round 1 to Round 3? In contrast to Experiment 1,

² Normality tests indicated normal distributions for both congruent trials ($K^2 = 0.18$, $p = .91$) and incongruent trials ($K^2 = 0.82$, $p = .66$). The outlier removal procedure had almost no effect on the main comparison; an analysis without removal showed similar results, $t(57) = 3.33$, $p = .002$, $g = .64$.

³ A similar analysis for Round 3 on congruent trials showed a shift to a sound-symbolic approach [$MD = .09$; $t(57) = 2.43$, $p = .02$, $g = .64$]. Problematically, though, targets and target-related distractors could not have been treated identically in Round 3, because participant responses should have begun to distinguish these two categories based on prior feedback.

Fig. 3 Experiment 2 overall accuracy across all rounds, along with accuracy by round. Vertical bars represent 95% confidence intervals, and boxes represent the 1st and 3rd quartiles. Con, congruent trials; Incon, incongruent trials



Experiment 2 was designed to make the congruent and incongruent word–picture relationships less obvious, thereby encouraging participants to develop non-sound-symbolic strategies for decision-making. This could explain why few participants used a sound-symbolic strategy from the start in Round 1. But what explains the growing congruency effect by Round 3? As in Experiment 1, a multiple-strategy approach seems to account for the data best. If participants switched to a purely sound-symbolic approach by Round 3, accuracy on incongruent trials would have declined significantly, but this did not occur. Instead, accuracy on congruent trials increased while accuracy on incongruent trials remained the same. It is important to recognize here that participants in Experiment 2 were getting negative feedback on their performance on 73% of trials on average, across all three rounds, as compared to 46% of trials for participants in Experiment 1. Faced with such strong negative feedback, participants most likely began searching for alternative strategies during and after Round 1, including sound symbolism.

Support for this interpretation could again be found in the postexperiment questions (identical to those used in Exp. 1). As before, a variety of strategies were described: relating the pseudoword to a real word ($N = 10$), using nonspecified sound symbolism ($N = 9$), using consonant sound symbolism ($N = 8$), using vowel sound symbolism ($N = 7$), rote memorization ($N = 6$), relating the shape to a known object ($N = 6$), random guessing ($N = 4$), and using word length as a cue ($N = 3$). Relative to Experiment 1, where 39% of the participants mentioned using a consonant-focused strategy (and 70% reported using some kind of sound-symbolic strategy), in Experiment 2 only

14% reported using a consonant-focused approach (with 40% reporting any kind of sound-symbolic strategy). Generally, participants were less able to articulate the strategies they used, with 15 participants providing an unclear explanation of their approach. With only 8 participants reporting a consonant-focused approach, a post hoc analysis contrasting these participants with others could not reveal much. It is worth noting, though, that although these participants showed a strong congruency advantage overall for congruent pairs ($M = .30$) versus incongruent pairs ($M = .22$), this advantage emerged not in Round 1, where there was no difference between the two conditions ($M = .27$ for both congruent and incongruent pairs), but in Round 3, where there was a larger difference ($M = .36$ on congruent trials vs. $M = .23$ on incongruent trials). Interestingly, though, these consonant-focused participants alone were not sufficient to produce the overall congruency advantage across all 58 participants. Indeed, excluding participants who reported a consonant-focused strategy or a nonspecified sound-symbolic approach, an overall congruency tendency remained [$MD = .03$; $t(41) = 1.84$, $p = .07$, $g = .41$] that grew stronger from Round 1 [$MD = .01$; $t(41) < 1$, $p = .73$, $g = .09$] to Round 3 [$MD = .05$; $t(41) = 2.09$, $p = .04$, $g = .42$]. In other words, although consonant-focused participants showed the strongest effects (as in Exp. 1), even participants who clearly reported no such strategy showed a tendency to incorporate a sound-symbolic approach to some extent by Round 3. This tendency might not have been strong enough, however, to elicit a postexperiment report.

Rates of selecting target-related and -unrelated distractors could not be directly compared, because each trial presented one target-related distractor and two target-unrelated

distractors, with the latter images being conceptually indistinguishable and therefore inseparable in the analysis. To permit a direct comparison with target-related distractors, the selection rates for target-unrelated distractors were divided by two (Table 1 shows the modified and nonmodified selection rates). Overall, there were no differences between the selection rates for target-related and target-unrelated distractors for either congruent pairs ($MD = -.01$) or incongruent pairs ($MD = .01$). Contrary to the sound symbolism hypothesis, participants did not favor the nontarget objects that carried sound-symbolic information consistent with the preceding pseudoword. But a round-by-round analysis provides further support for a growing influence of sound symbolism from Rounds 1–3. If participants were using sound symbolism in Round 1, one would expect target-related distractors to be selected more often than target-unrelated distractors. In Round 1, though, participants showed a tendency to select one of the unrelated distractors rather than the target-related distractor on congruent trials [$MD = -.04$; $t(57) = -1.83$, $p = .07$, $g = .42$]. By Round 3, this difference had disappeared [$MD = 0.01$; $t(57) < 1$, $p = .75$, $g = .07$], indicating that target-related distractors became relatively more attractive than unrelated distractors as participants began incorporating sound symbolism. Comparatively, incongruent trials showed little change from Round 1 ($MD = -.01$) to Round 3 ($MD = .01$) between target-related and -unrelated distractors, supporting the idea that

participants managed to apply sound-symbolic information selectively between conditions.⁴

Within pseudoword categories, some interesting differences appeared. On incongruent trials with a curvilinear target preceded by a plosive pseudoword, participants chose the target-related (curvilinear) distractor significantly more often than either of the unrelated (rectilinear) distractors [$MD = .05$; $t(57) = 2.69$, $p = .01$, $g = .59$]. In other words, when participants failed to choose the target-related curvilinear shape after hearing a plosive pseudoword, they chose the only available curvilinear distractor somewhat more often than the other, unrelated targets. When a nonplosive pseudoword preceded a rectilinear target, there was a tendency for incorrect responses to gravitate toward the target-unrelated curvilinear distractors [$MD = -.03$; $t(57) = -1.63$, $p = .11$, $g = .35$]. In this case, when participants chose the wrong picture after hearing a nonplosive pseudoword, they chose one of the available curvilinear distractors rather than the target-related rectilinear distractor. In short, there was a moderate bias for selecting curvilinear distractors under conditions of uncertainty. This bias cannot explain the congruency effects, however, because rectilinear objects also showed congruency effects in both experiments.

General discussion

In two experiments, sound symbolism among congruent word–picture pairs produced higher response accuracies in an object recognition task, relative to incongruent word–picture pairs. For congruent trials, pseudowords were paired with shapes based on a natural correspondence between their articulatory features and the linearity of the shapes. For example, *kuh-der-pai*, with three plosives, was paired with a rectilinear object on congruent trials but with a curvilinear object on incongruent trials. A congruency advantage emerged in both experiments, despite the fact that a sound-symbolic response strategy to word–picture pairs could not provide participants with an advantage for increasing their overall accuracy. Using a sound-symbolic strategy on each trial would produce overall accuracy rates no better than chance across all three rounds: .50 (Exp. 1, with two choices on screen) or .25 (Exp. 2, with four choices).

Table 1 Selection rates for targets (T), target-related distractors (TR), and target-unrelated distractors (TU) in Experiment 2

Trial Type	Congruent Trials			Incongruent Trials		
	T	TR	TU	T	TR	TU
All Trials						
Overall	.29	.23	.24 (.48)	.24	.26	.25 (.50)
Round 1	.27	.22	.26 (.51)	.24	.24	.26 (.52)
Round 2	.29	.23	.24 (.48)	.25	.28	.24 (.47)
Round 3	.31	.23	.23 (.45)	.24	.26	.25 (.50)
Rectilinear						
Overall	.27	.24	.25 (.49)	.22	.24	.27 (.53)
Round 1	.26	.22	.26 (.53)	.23	.22	.28 (.56)
Round 2	.27	.24	.24 (.49)	.24	.24	.26 (.52)
Round 3	.29	.25	.23 (.46)	.20	.26	.27 (.53)
Curvilinear						
Overall	.31	.22	.24 (.47)	.26	.28	.23 (.46)
Round 1	.28	.22	.25 (.50)	.26	.26	.24 (.48)
Round 2	.30	.23	.24 (.48)	.25	.32	.21 (.43)
Round 3	.34	.22	.22 (.44)	.28	.25	.24 (.48)

The raw, nonmodified TU selection rates are presented in parentheses, after the modified rates used for comparisons with TR selection rates

⁴ As noted previously, a similar analysis on congruent trials, which compared sound-symbolic consistent choices (target and target-related objects) with inconsistent choices (target-unrelated objects), showed the same pattern. For incongruent trials, there were no significant differences between target-unrelated selection rates and the combined target and target-related selection rates in Round 1 ($MD = -.04$) and in Round 3 ($MD = .00$).

Previous research demonstrating advantages for sound symbolism in language tasks has often used small sets of stimuli (Imai et al. 2008; Köhler, 1929). The experiments reported here produced sound-symbolic effects with 32 pseudowords divided equally into “hard” and “soft” articulatory categories, by using a variety of plosive and nonplosive consonants. Together with Westbury (2005), the results here confirm that accuracy or speed in language-related tasks can be affected by the influence of sound symbolism in broad categories of phonemes. This tendency emerges even when participants never pronounce novel words out loud, indicating that phonological representations contain information about airflow apart from the actual experience of pronunciation.

Responses to curvilinear and rectilinear objects were equally affected by sound symbolism, but nonplosive pseudowords in both experiments produced stronger sound symbolism effects than did plosive pseudowords. Westbury (2005) also reported that continuant pseudowords produced somewhat stronger congruency effects in a lexical decision task, similar to the results of [Experiment 1](#) here, but unlike the results of [Experiment 2](#), in which only nonplosives produced an overall congruency effect. Whether this represents an aberration or a genuine pattern is difficult to determine from only two studies. But, as future research begins to explore sound symbolism in more detail, the contrast between plosives and nonplosives could help illuminate that landscape of sound-symbolic effects.

[Experiment 1](#) provided participants with two clearly contrasting rectilinear and curvilinear shapes, thereby foregrounding the possible relevance of sound-symbolic relationships between these shapes and their preceding pseudowords. As such, some participants showed a tendency to use a consonant-focused strategy right from the start in Round 1. Interestingly, accuracy on incongruent trials also improved by Round 3 without a corresponding decline on congruent trials, indicating that participants were able to employ different strategies for congruent and incongruent pairs. When provided with two additional response options onscreen that minimized the relevance of rectilinear–curvilinear visual features in [Experiment 2](#), participants were slower to incorporate sound symbolism into their decision-making, with the strongest influence emerging in Round 3. In effect, the structure of [Experiment 2](#) masked the relevance of sound-symbolic relationships at the start. The usefulness of that dimension became apparent to some participants only after receiving overwhelmingly negative feedback in Round 1 for ineffective non-sound-symbolic strategies.

Why would sound symbolism affect the ability to associate words with objects? At least one study has

indicated that children benefit from sound-symbolic effects in learning the names of objects (Imai et al. 2008). As such, sound symbolism could provide language learners with a mechanism for binding words with objects in a predictable way at a time when they are engaged in multiple language acquisition tasks (learning how to pronounce words, how to use grammar, how to choose useful words, etc.). This mechanism might be useful throughout life, however, for resolving ambiguities in comprehension, for predicting the meanings of unfamiliar words, for learning new languages, or for inventing new words. Not all of these uses would necessarily be connected to the original role of sound symbolism in language development. Researchers have speculated, for example, that protolanguage began with sound symbolism and that human language gradually became less sound-symbolic over time (Allott, 2001; Foster, 1978). The postexperiment feedback from [Experiment 2](#) pointed to the possibility that sound-symbolic effects influence language processes regardless of the conscious strategies used in experimental tasks. If so, this tendency would lend further support to the idea that sound symbolism played a role in the evolution of language in humans, an idea that would be more credible if demonstrably subconscious sound symbolism effects could be discovered.

In any case, a growing body of research has demonstrated that adults can employ sound symbolism for a variety of language-related tasks in a way that probably transcends its older purpose. Thus, sound symbolism stands as a useful tool for the adult language comprehender as well as for the first-language learner. Future experiments could extend the results reported here by introducing additional rounds of testing, but the difficulty of this task makes it likely that testing fatigue would become an issue. A better approach might involve the use of verbal protocols to explore the various online strategies being employed. Such an approach could also help uncover the relative contributions of conscious and subconscious processes in sound-symbolic relationships. Another approach for exploring the time course of sound symbolism’s effect on language processes would be to create a more interesting and naturalistic task in which participants try to accomplish some goal using complex symbols with congruent and incongruent sound-symbolic relationships. In this way, observing a longer time course of sound-symbolic development might be possible without fatiguing participants with artificial and difficult experimental tasks. Additionally, the practical benefits of using sound symbolism in natural language could be more highly specified, yielding discoveries that could ultimately contribute techniques for improving communication.

Appendix

Table 2 List of pseudoword stimuli in IPA notation (with breaks between syllables)[COMP: Preserve phonetic characters and columns below.]

Plosive Pseudowords		
k ^ʌ d ^ɜ r p ^a ɪ	bu p ^ɔ ɪ d ^æ	b ^æ pi k ^æ
peɪ ki g ^ɔ ɪ	g ^æ tu beɪ	doʊ gl beɪ
dɪ b ^ʌ kaɪ	k ^ɔ ɪ t ^æ gi	gaɪ b ^ɛ ku
di geɪ toʊ	teɪ baɪ kaʊ	toʊ k ^æ pi
paʊ gaɪ t ^ʌ	baʊ du p ^ʌ	
t ^ɛ kaɪ doʊ	taɪ doʊ p ^ɔ ɪ	
Non-Plosive Pseudowords		
f ^ʌ ɪ saɪ	wu h ^ɔ ɪ r ^æ	h ^æ si r ^æ
seɪ wi h ^ɔ ɪ	s ^æ su reɪ	roʊ sɪ feɪ
ɪ f ^ʌ saʊ	f ^ɔ ɪ r ^æ si	faɪ r ^ɛ lu
hi leɪ soʊ	weɪ waɪ laʊ	loʊ h ^æ fi
raʊ faɪ w ^ʌ	saʊ hu w ^ʌ	
r ^ɛ saɪ loʊ	faɪ loʊ h ^ɔ ɪ	

References

- Allott, R. (2001). *The natural origin of language: The structural interrelation of language, visual perception, and action*. Knebworth, Hertfordshire: Able.
- Bentley, M., & Varon, E. J. (1933). An accessory study of “phonetic symbolism.” *The American Journal of Psychology*, *45*, 76–86
- Foster, M. L. (1978). The symbolic structure of primordial language. In S. L. Washburn & E. R. McCown (Eds.), *Human evolution: Biosocial perspectives* (pp. 77–122). Merlo Park, CA: Benjamin Cummings.
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*, 54–65. doi:10.1016/j.cognition.2008.07.015
- Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.
- Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*, 19–28. doi:10.1016/j.cognition.2009.08.016
- Lowry, T. M., & Shrum, L. J. (2007). Phonetic symbolism and brand name preference. *Journal of Consumer Research*, *34*, 406–414.
- 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
- Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, *28*, 225–252. doi:10.1146/annurev.anthro.28.1.225
- Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*, 181–186. doi:10.1016/j.cognition.2009.04.001
- Parault, S. J., & Schwanenflugel, P. J. (2006). Sound-symbolism: A piece in the puzzle of word learning. *Journal of Psycholinguistic Research*, *35*, 329–351. doi:10.1007/s10936-006-9018-7
- Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*, 225–239. doi:10.1037/h0070931
- Taylor, I. K., & Taylor, M. M. (1965). Another look at phonetic symbolism. *Psychological Bulletin*, *64*, 413–427. doi:10.1037/h0022737
- 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