To call a cloud 'cirrus': Sound symbolism in names for categories or items

Vanja Kovic [Corresp., 1] , Jelena Sucevic [2] , Suzy J Styles [3]

[1] Department of Psychology, University of Belgrade, Belgrade, Serbia
[2] Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom
[3] Division of Psychology, Nanyang Technological University, Singapore, Singapore

Corresponding Author: Vanja Kovic
Email address: vanja.kovic@f.bg.ac.rs

The aim of the present paper is to experimentally test whether sound symbolism has selective effects on labels with different ranges-of-reference within a simple noun-hierarchy. In two experiments, adult participants learned the make up of two categories of unfamiliar objects ('alien life forms'), and were passively exposed to either category-labels or item-labels, in a learning-by-guessing categorization task. Following category training, participants were tested on their visual discrimination of object pairs. For different groups of participants, the labels were either congruent or incongruent with the objects. In Experiment 1, when trained on items with individual labels, participants were worse (made more errors) at detecting visual object mismatches when trained labels were incongruent. In Experiment 2, when participants were trained on items in labelled categories, participants were faster at detecting a match if the trained labels were congruent, and faster at detecting a mismatch if the trained labels were incongruent. This pattern of results suggests that sound symbolism in category labels facilitates later similarity judgments when congruent, and discrimination when incongruent, whereas for item labels incongruence generates error in judgements of visual object differences. These findings reveal that sound symbolic congruence has a different outcome at different levels of labelling within a noun hierarchy. These effects emerged in the absence of the label itself, indicating subtle but pervasive effects on visual object processing.

1   To call a cloud 'cirrus': Sound symbolism in names for categories or items.

2   Vanja Ković*[1], Jelena Sučević[2] and Suzy J Styles[3]

4   [1] Department of Psychology, University of Belgrade, Belgrade, Serbia

5   [2] Department of Experimental Psychology, University of Oxford, Oxford, UK

6   [3] Division of Psychology, Nanyang Technological University, Singapore

8   Abstract:

9   The aim of the present paper is to experimentally test whether sound symbolism has selective
10   effects on labels with different ranges-of-reference within a simple noun-hierarchy. In two
11   experiments, adult participants learned the make up of two categories of unfamiliar objects
12   ('alien life forms'), and were passively exposed to either category-labels or item-labels, in a
13   learning-by-guessing categorization task. Following category training, participants were tested
14   on their visual discrimination of object pairs. For different groups of participants, the labels were
15   either congruent or incongruent with the objects. In Experiment 1, when trained on items with
16   individual labels, participants were worse (made more errors) at detecting visual object
17   mismatches when trained labels were incongruent. In Experiment 2, when participants were
18   trained on items in labelled categories, participants were faster at detecting a match if the trained
19   labels were congruent, and faster at detecting a mismatch if the trained labels were incongruent.
20   This pattern of results suggests that sound symbolism in category labels facilitates later similarity
21   judgments when congruent, and discrimination when incongruent, whereas for item labels
22   incongruence generates error in judgements of visual object differences. These findings reveal
23   that sound symbolic congruence has a different outcome at different levels of labelling within a
24   noun hierarchy. These effects emerged in the absence of the label itself, indicating subtle but
25   pervasive effects on visual object processing.

26

27

28   *Corresponding author. Vanja Ković, Čika Ljubina 18-20, 11000 Belgrade, Serbia,

29   vanja.kovic@f.bg.ac.rs

Introduction:

When cloud watchers talk about a 'cirrus' or a 'cumulus', they are naming cloud types which can be distinguished by their visual form. The cirrus (with its voiceless sibilants, and high front vowel), refers to a high-altitude ice cloud, with a typically thin, wispy form. The cumulus, (with its sonorant consonants, and rounded vowels), refers to a puffy, low-altitude cloud, prone to rising muffin-tops. Do the sounds encoded in the names of the clouds enhance visual differences between the clouds' shapes? And if so, does an alliance of sounds with visual forms differ when labels are used at different levels of description (e.g., 'cloud' versus 'cirrus')? This nebulous example[1] allows us to ask an important question about the relationship between labelling itself, and a label's 'range of reference,' for items that can be named at different levels within a noun hierarchy.

Since the first systematic studies into sound symbolism in linguistic labelling, research has tended to focus on the mapping of labels onto a small number of individual items (usually a pair), which can be discriminated on some visual dimension. For example, Sapir (1929) asked people which of the word-forms 'mil' or 'mal' would be a better label for pairs of items like a small or large table, and Köhler (1929:1947) asked whether a curved or an angular line-drawing would be a better match for word forms like 'baluma' and 'takete'. The majority of the literature on sound symbolism (for recent reviews see Imai & Kita, 2014; Lockwood & Dingemanse, 2015) has followed this trend for contrastive label-stimulus mappings, typically for a small number of items, usually two-dimensional line drawings, usually presented in pairs (prominent examples include Maurer, Pathman, & Mondloch, 2006; Ramachandran & Hubbard, 2001). Some alternative approaches include matching single words with pairs of pictures randomly generated to match certain visual characteristics of known congruence types (Drijvers, Zaadnoordijk, & Dingemanse, 2015; Nielsen & Rendall, 2011, 2013), or matching word-forms to a variety of natural and artificial objects, with varying agency (Flumini, Ranzini, & Borghi, 2014). However, as the majority of methods rely on the one-to-one match between a single label

---

[1] Although this lexical pair is not a canonical example of linguistic sound symbolism, when presented with a pair of pictures showing canonically wispy cirrus (cirrus uncinus) and canonically puffy cumulus (cumulus congestus) clouds, 83% of 135 Serbian speakers unfamiliar with the formal scientific cloud names guessed which cloud is called which. For further details see the Open Science Framework repository for this article (https://osf.io/2wvug/).

56 and single visual stimulus, it is difficult to interpret what range-of-reference these novel labels

57 might play in naturalistic noun-labelling hierarchy – do they refer to abstract superordinate

58 categories (e.g., animal), basic-level categories (e.g., dog), subordinate-nouns (e.g., Labrador), or

59 even proper nouns for individual exemplars (e.g., Fido).

60   Research using words from unfamiliar languages provides a ready-made range-of-

61 reference, since participants will likely bootstrap the novel word to its translation equivalent. For

62 example, in a study by Berlin (1995), English speaking participants were above chance at

63 guessing which label in a pair was a type of fish, and which a type of bird, for stimuli from an

64 unfamiliar language. Studies of this kind demonstrate that sound symbolism can reside in labels

65 for categories of natural objects, not just as one-off names for arbitrary experimental tokens:

66 sound symbolism can enhance word learning for antonym pairs in unfamiliar languages (e.g.,

67 'fast'/'slow') (Lockwood, Dingemanse, & Hagoort, 2016; Nygaard, Cook, & Namy, 2009), and

68 facilitate word learning in tasks where adults and infants learn labels for novel action events, and

69 extend them to new exemplars of the action (Imai, Kita, Nagumo, & Okada, 2008; Kantartzis,

70 Imai, & Kita, 2011). Sound symbolism is thus not limited to arbitrary lab-based labels for

71 abstract visual items, but is evident in natural-category labels, descriptive phrases, and action

72 descriptions. However, among this body of research, it has remained unclear whether sound

73 symbolism plays different roles at different levels of labelling within a single hierarchy.

74   There is a growing body of research suggesting that words have a special status in

75 activating conceptual representations (Lupyan & Thompson-Schill, 2012). For example, labels

76 activate representations that are more abstract (Edmiston & Lupyan, 2015), and facilitate

77 acquisition of novel categories even when the label is redundant to the category (Lupyan,

78 Rakinson, & McClelland, 2007), whereas this is not the case for other correlated non-verbal

79 cues, such as environmental sounds. The speech-specific nature of labelling on category

80 formation has been shown in infants as young as 10-months-of-age (Althaus & Westermann,

81 2016), where labels have been shown to induce visual attention to parts of objects containing

82 featural similarities (Althaus & Mareschal, 2014). That is to say, even for babies, hearing objects

83 named-alike induces attention to shared object properties, and these attentional biases are

84 thought to underly the process of categorization. Bringing categorization and sound symbolism

85 together, a recent word learning study by Monaghan and colleagues (Monaghan, Mattock, &

86   Walker, 2012) showed that sound-symbolic labels facilitated learning the mapping between the

87   label and the *category* of items, but not between the label and the *individual* objects.

88   One series of studies (Kovic, Plunkett, & Westermann, 2010) has investigated sound

89   symbolism in labels which explicitly refer to whole categories of novel objects where multiple

90   members of the category are available to the participant. In Kovic et al. (2010), participants

91   learned to sort two classes of novel, schematic creatures, by allocating them to named categories.

92   Participants began by guessing which creature belonged to which category, and with feedback,

93   gradually refined their answers as they learned to categorize items. The creatures differed

94   according to a number of schematic dimensions (e.g., number of legs, shape of tail), but no

95   single feature was solely responsible for which item belonged to which category. Two groups of

96   participants learned to form categories with labels which were sound-symbolically congruent or

97   incongruent with the typical shape of the creatures' head (e.g., 'dom' or 'shick' for mostly-round

98   or mostly-pointy heads). In two separate behavioural tasks, the pairings of labels to categories

99   did not influence training speed or recall accuracy. However, during later test phases (in which

100   people heard a label, then decided whether a subsequently presented picture matched) the

101   congruence of the shape of the creatures' heads with the newly learned labels influenced the

102   *speed of decisions*. This finding suggested that although the efficiency and success of word

103   learning did not differ between conditions, links between the auditory label and the category

104   members were more efficiently accessed for participants who had learned sound symbolically

105   congruent labels.

106   The same study also reported electrophysiological evidence of differences between

107   responses of the two groups of participants – during the test phase, ERPs showed characteristics

108   of enhanced visual object processing for people who had learned the congruent category labels.

109   Taken together, these findings suggest that when sound symbolism is present for a recently

110   acquired category label, it enhances our ability to access information relevant to the mapping

111   between form and meaning. In terms of mechanisms, this finding could suggest that a congruent

112   label invokes (primes) salient features of its member items prior to their visual appearance, thus

113   enhancing object recognition. Alternatively, encoding with symbolism may cause the individual

114   items to be represented with richer feature representations, meaning that their individual shapes

115   were recognized more efficiently. This could be achieved by more efficient access to the stored

116 knowledge about each item during the test phase, following sound-symbolic encoding during
117 training.

118   However, notable questions still remain about the limits of sound symbolism in
119 naturalistic language environments. As argued in a recent review (Dingemanse, Blasi, Lupyan,
120 Christiansen, & Monaghan, 2015), if all concepts with similar meanings had similar-sounding
121 names, the phonological space for semantically related items would become extremely crowded,
122 leading to a kind of 'systematicity' in which multiple similar form-meaning relationships
123 generate high confusability. This crowding may necessitate a tug-of-war between systematicity
124 and arbitrariness in natural language systems, with systematicity arising from similar meanings
125 being encoded by similar sounds, and arbitrariness facilitating discriminability between items
126 with similar meanings. It has therefore been suggested that systematicity plays a role in grouping
127 together categories, rather than distinguishing between exemplars. This claim has been tested on
128 meaning relations arising from different grammatical categories (Monaghan & Christiansen,
129 2006; Monaghan, Christiansen, & Fitneva, 2011). However, it remains to be tested whether
130 similar effects are observed for labels at different levels of noun taxonomy.

131   The aim of the present paper is to experimentally test whether sound symbolism has a
132 different effect in labels with a different range-of-reference within a simple noun-hierarchy. We
133 conducted a study in which people learned the make up of two categories of unfamiliar objects
134 ('alien life forms'), and were passively exposed to either labels for the categories, or labels for
135 each item, while they learned. Within this framework, for different groups of participants, the
136 labels were either congruent or incongruent with the objects. Since previous research (Kovic, et
137 al., 2010) has demonstrated that sound symbolism does not alter the speed or the accuracy of
138 category formation, we did not predict the influence of sound symbolism to be present during
139 training, but to emerge in tests conducted after category learning. We therefore asked whether
140 congruence between the label and the visual form during training might cause people to attend to
141 the objects differently, and influence visual object processing, even in the absence of the auditory
142 training stimulus (the label). The test phase was implemented as a silent, visual same/different
143 judgement, to tease apart whether effects are driven by sound symbolism altering the strength of
144 the mapping between the word-form and the visual features of the object, or by sound symbolism
145 altering the representation of the visual objects themselves. To the best of our knowledge, no

146   other studies have attempted to look at differences in visual object processing caused by sound
147   symbolism, in the absence of the sound symbolic stimulus.

148   We predicted that if sound symbolism is effective in highlighting featural similarities
149   among members of a category, then congruent labels would enhance performance (faster RT and
150   fewer errors) for categories learned with category-labels, but would not influence performance
151   for categories learned with item labels. On the other hand, if sound symbolism enhances object
152   processing on a trial-by-trial basis, via a generalized form of crossmodal congruence between
153   individual sounds and individual object features, we predicted that training with congruent labels
154   would enhance performance (faster RT, fewer errors) for categories learned with all levels of
155   labelling.

156   Method

157   Participants

158   Eighty-two participants from a local Science Center and second-year psychology
159   undergraduate students at the first author's institution (13 male, 17-19 years old), took part in the
160   present experiment and received course credit for their participation. All participants reported
161   normal or corrected-to-normal vision and were unaware of the purpose of the present study. This
162   research was approved by one branch of the Serbian Psychological Association, the Ethics
163   Committee from the Department of Psychology, University of Niš (document number 3/16).

164   Stimuli

165   Visual stimuli were abstract pencil drawings of novel complex shapes designed for use in
166   a number of different studies. The forms were inspired by the shape of the historic Vinča figures,
167   products of a Neolithic culture from the Balkan region. These abstract drawings were created to
168   form two categories, differing in their visual properties: one category consisted of curved shapes,
169   in rounded forms, and the other category, angular shapes, in a vertical orientation (see Figure
170   1B). Each category contained six members. Given their abstract forms (reminiscent of Picasso),
171   we term these stimuli Vinčasso illustrations.

172   Given the novelty of stimuli, a rating task was run in order to establish that the two sets
173   of Vinčassos did indeed represent a contrast between soft/round and sharp/angular visual forms,

174   with between category difference, and within-category homogeneity. 135 psychology and

175   education second year undergraduate students (age: 17.5-19.5 years) were asked to judge each

176   item's shape on a 7-point scale ranging from 'round' (at the leftmost side of the scale) to angular

177   (at the right). The online rating task was implemented using the Qualtrics survey platform, where

178   each item was presented in a random order. The results showed that the pencil drawn Vinčassos

179   did indeed fall into two categories (Round: $M = 1.58$, $SD = .10$; Sharp: $M = 5.68$, $SD = 0.43$;

180    $t(5.51) = 22.98$, $p < .001$), with high between category difference, and high within-category

181   homogeneity. The same participants also rated the complexity of the pictures from 'simple' to

182   'complex'. The results showed no overall differences between the two Vinčasso categories

183   (Round: $M = 4.3$, $SD = .79$; Sharp: $M = 3.94$, $SD = .45$; $t(1, 11) = .95$, $p = .36$), which showed

184   substantial within-category and within-subject variability. The original Vinčasso stimuli at

185   screen-resolution can be found in the Open Science Framework repository for this article

186   (https://osf.io/2wvug/), and further details of the rating study can be found in the Supplementary

187   Materials (S1).

188   Twelve auditory pseudowords were used as labels for the pictures. The pseduowords,

189   have a C-V-C-V-C structure, and were read aloud by the first author, a female native speaker of

190   Serbian, using a neutral intonation. In the current study, we elected to use natural speech in

191   preference to artificially generated speech, as we are interested in the naturalistic processing of

192   realistic whole language, without the potential confounds introduced by participants' awareness

193   of the unnatural characteristics of synthesized or flat-prosody speech. Stimuli were recorded in a

194   single session and filtered to remove hiss and hum using GoldWave. Stimuli were approximately

195   equal in duration (M = 790ms, SD = 82.3ms). Half of the stimuli contained soft phonological

196   structure (e.g., 'volab') and the other half, sharp phonological structure (e.g., 'šičak', See Figure

197   1 for further details). The 'sharpness' and 'softness' of the phonemes were based on findings of

198   several previous studies of Serbian, the language of test, and the twelve test items were

199   previously validated for their 'sharp' and 'soft' associations (Sučević, Savić, Popović, Styles, &

200   Ković, 2015). The audio files used for presentation in the experiment can be found in the Open

201   Science Framework repository for this article (https://osf.io/2wvug/). As it has been suggested

202   both phonetic and prosodic characteristics combine to generate sound symbolic effects

203   (Dingemanse, Schuerman, Reinisch, Tufvesson, & Mitterer, 2016; Nygaard, et al., 2009), further

204   details about the acoustic envelope of the stimuli can be seen in the Supplementary Materials

205 (S2), where the stimuli for the different shape conditions show substantial overlap in their
206 prosodic contours.

207

208 Experimental design and procedure

209   Participants were asked to imagine they have just landed on an unexplored planet and
210 they need to learn to differentiate between two species that live on the planet in order to survive.
211 Participants were seated in front of a computer and were given headphones. Instructions were
212 presented on screen.

213   The experiment consisted of a training and a test phase. In the training phase, participants
214 were familiarized with the categorical structure of the presented objects. In the first half of the
215 training phase, members of the two categories were presented one-by-one, with an auditory
216 label: Each trial began with a fixation cross presented for 500 ms (±100ms), followed by the
217 auditory label, then a single Vinčasso, approx 12cm in height. While the creature remained
218 onscreen, participants were asked to guess which of the two categories it belonged to, by pushing
219 one of two buttons ('C' and 'N' on a standard keyboard), with key allocation counterbalanced
220 across participants using ABAB sequence. When participants made their selection, feedback was
221 displayed on the screen for 600ms ('Correct!' or 'Incorrect!') (Figure 1). All twelve stimuli were
222 presented once in a random order.

223   In the second stage of training, labels were omitted, so that the creature was presented
224 immediately after the fixation cross (which remained on the screen for 500ms ±100ms). All
225 twelve stimuli were presented once in a random order.

226   After the training phase, participants were instructed to identify whether the two creatures
227 presented side-by-side were the same ('identical') or different. In the Test block, a fixation cross
228 was presented for 500ms (±100ms), followed by the presentation of two Vinčassos side-by-side.
229 The Creatures remained on the screen until participants responded. The test pairs depicted the
230 same item (Identical Match), or items drawn from different categories (Category Mismatch). For
231 mismatching pairs, each image was randomly paired with a picture from the other category, and
232 pairs were yoked across participants. 12 trials in each condition were presented in a random

233   order, making a total of 24 trials of these types[2].

234

235   Training Conditions

236   *Label Type* was manipulated between experiments: In Experiment 1, participants heard each
237   Vinčasso named with a different label (Item Labels), meaning that the auditory labels were not
238   meaningfully correlated with the to-be-learned category structure. In Experiment 2, participants
239   heard only two labels, each of which labeled one of the categories (Category Labels), meaning
240   that the auditory labels were correlated with the to-be-learned categories, but were redundant, as
241   the task could be solved without them.

242   *Label Congruence* was systematically manipulated between participants: Half of the participants
243   heard labels that were congruent with the visual properties of the Vinčassos (soft phonology –
244   rounded shapes; sharp phonology – angular shapes), and half heard incongruent labels.
245   Participants were randomly assigned to a Congruence condition.

246   Since the categories were defined by their visual properties in both experiments (rounded,
247   angular), the labels were an irrelevant feature of the training environment. In other words,
248   participants could have successfully solved the category structure in this task without paying any
249   attention to the labels. Here, we ask the question whether labels presented during a first
250   encounter with a visual object alter subsequent visual processing of the object.

251

---

2 We originally included a third condition in the Test Phase, in which participants saw different
pictures from the same category. However, around two thirds of participants misunderstood the
task instructions, and answered on the basis of whether the picture was from the 'same category',
instead of whether the pictures were 'identical'. The large scale of the dropout rendered this
condition un-analysable here (Further details can be found in the Supplementary materials S3
and the Open Science Framework repository for this project: https://osf.io/2wvug/)

Figure 1. Experimental Design. A. The three blocks of the study, showing the structure of trials, and event timing. B. Stimuli and experimental conditions. Structure of the visual categories, and the twelve auditory labels used in Item Label training condition (names are given in standard Serbian Latin orthography, and glossed in IPA), and the 2 labels used in the Category Label training condition, along with examples of the test trial types, and the training conditions.

Data handling and analysis

Mean error rates were computed for individuals in each of the four training conditions, and each of the three types of test trial. Reaction times were transformed into inverse RT (iRT=1/RT), to approximate normal distribution for each participant. Based on visual inspection

264   of the distribution, extreme outliers (exceeding 100-4000ms) were excluded from further

265   analysis. For correct trials each individual's mean inverse RT was computed. Following analysis,

266   reaction times were back-transformed into milliseconds (RT=1/iRT) for graphing, and for

267   presentation of means.

268   For analysis of the training phase, mixed ANOVAs investigated the influence of Training

269   Condition (Congruent, Incongruent) on Error Rate and on Reaction Time across the two training

270   Blocks (First, Second). For analysis of the test trials, mixed ANOVAs investigated the influence

271   of Training Condition (Congruent, Incongruent) on Error Rate and Reaction Time across the two

272   match types (Identity Match, Mismatch). Power analyses were conducted using G*Power 3.1

273   (Faul, Erdfelder, Buchner, & Lang, 2009).

274

275   Experiment 1: Item Labels

276   Each Vinčasso was presented with a different label during the first training block. Half of the

277   participants heard congruent, and half, incongruent item labels. Labels were yoked to pictures for

278   all participants in a given condition. The categorization task involved two categories of six

279   visually coherent items (See Figure 1).

280   Results and Discussion

281   **Training.** Figure 2 shows the results from the training and the test phase of Experiment 1, with

282   the means for individual participants shown separately. Means are given in Table 1. Participants

283   solved the categorization task, with a reduction in errors from the first to the second block (Error:

284   $F (1, 40) = 31.55, p < .001, {}_p\eta^2 = .44$, Observed Power = 1.00). No further effects were observed

285   for Error Rate or RT. As seen elsewhere (Kovic, et al., 2010) in this training-by-guessing

286   paradigm, sound symbolism is not observed to influence performance during training[3].

---

3   One participant with an unusually high error rate at the end of the second block can be observed
in the Incongruent training condition. This participant also has unusually slow RT during training.
Removal of this participant from this, and all subsequent analyses, has no impact on the pattern or
direction of the effects, with the exception of reduced power in most tests – here, the difference is not

287 **Test.** Error Rates differed across the different Match conditions (Match Type: F (1, 40) = 8.08,

288 p = .007, $_p\eta^2$ = .17, Observed Power = .79), with overall lower error rates for the Mismatch

289 condition than for the Match condition. This effect was larger for participants who were trained

290 with item labels that were congruent with the shape of the individual items, and was absent for

291 participants in the incongruent condition (Match Type x Congruence: F (1, 40) = 6.38, p = .016,

292 $_p\eta^2$ = .14, Observed Power = .69). This finding suggests that the correct detection of large

293 differences between paired images in the mismatch condition was disrupted by having earlier

294 encountered incongruent labels, suggesting sub-optimal processing of the visual properties of the

295 objects. Rejecting a mismatch would typically be easier in a task of this kind, since low-level

296 visual features (e.g., vertical edges, curves) can provide definitive information about a mismatch

297 drawn from different visual categories. Interestingly, for participants who heard incongruent

298 labels during their training, there was no general difference between match and mismatch trials.

299 Effectively, they were worse than would normally be expected at correctly rejecting mismatches.

300 However, as this effect was not predicted, and was somewhat under-powered, it should be

301 treated with caution, and a larger sample size would be required for future investigations seeking

302 to replicate this effect.

303 Reaction times did not differ significantly for participants in the different training conditions[4].

---

notable: F (1, 39) = 39.83, p < .001, $p\eta^2$ = .51, Observed Power =1.00.

4   In Figure 2B, this can be observed as the outcome of approximately two thirds of participants showing slightly faster responses for ID match than for mismatch (Congruent: 14, Incongruent, 13), overlapping with the remainder of participants, who showed faster responses for mismatch than for ID match (Congruent: 7, Incongruent: 8), but exhibited a larger reaction time difference between trial types. The degree of individual sensitivity to sound symbolism has previously been shown to correlate with extent of sound symbolic congruence in learning (Lockwood & Dingemanse, 2015). Here, since the direction of this split is the same for both training conditions, it does not appear to arise from sound symbolism, but may index a difference in visual object processing.

304
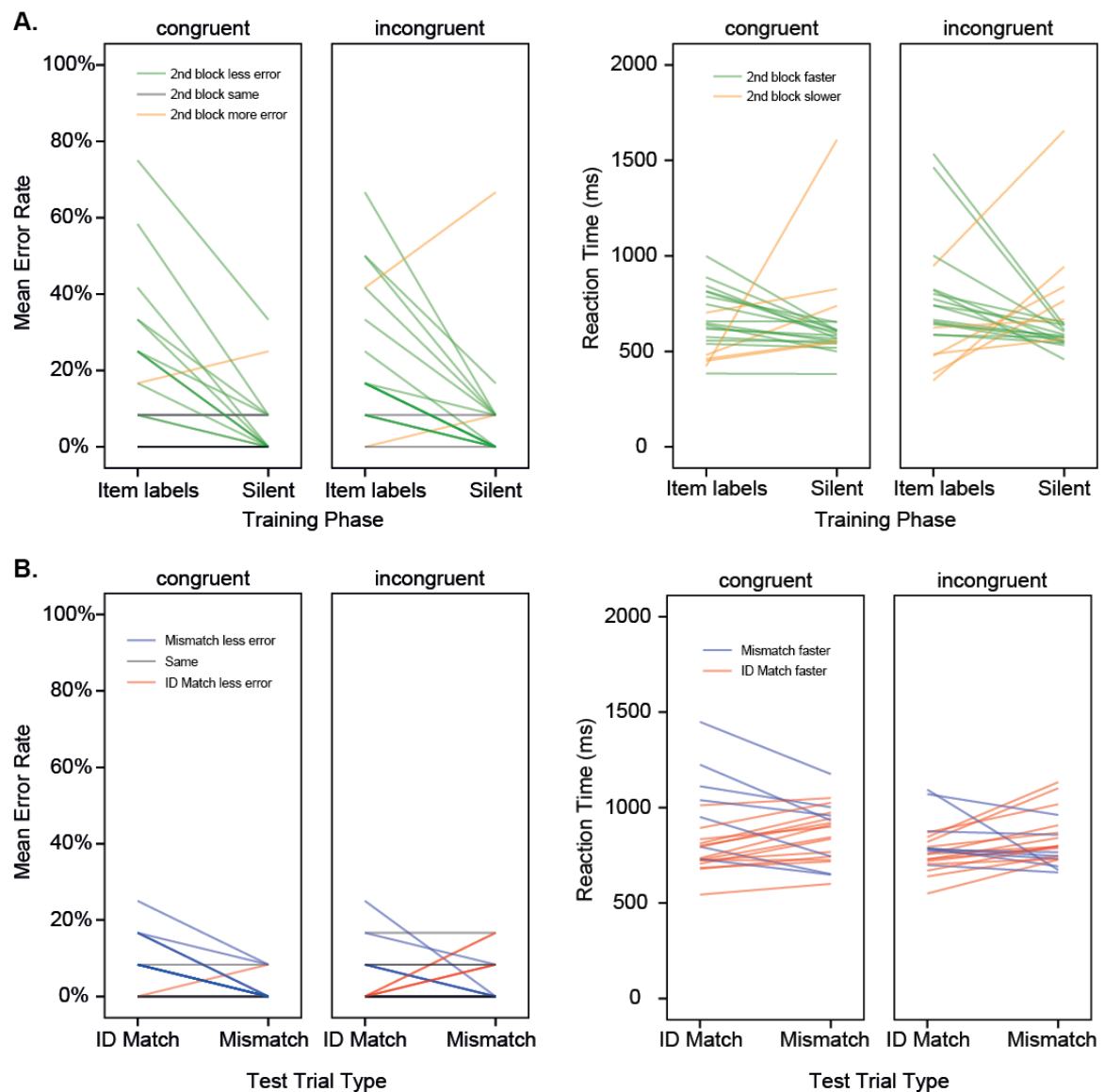
Figure 2. Experiment 1. Training with item labels. A. Results for the two training phases shown for individual participants in different training conditions. Mean error rate, and mean reaction time (back-transformed from iRT) are shown, where GREEN indicates better performance in the second block (more accurate, faster). C&D. Results for the two test trial types shown for individual participants in different training conditions. B. Mean error rate, and mean reaction time (back-transformed from iRT) are shown, where BLUE indicates better performance for the Mismatch condition (more accurate, faster).

313

314

315

316

317   Table 1. Mean Error rates and Reaction Times (back-transformed from iRT) for Experiment 1.

		Training				Test			
		Phase 1		Phase 2		ID Match		Mismatch	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Error Rate	Congruent	.19	.20	.05	.09	.08	.07	.02	.03
	Incongruent	.22	.18	.07	.15	.05	.07	.05	.06
RT (ms)	Congruent	650	167	637	239	856	212	861	151
	Incongruent	750	299	671	251	784	125	825	132

318

Experiment 2: Category Labels

320   Each Vinčasso was presented with one of two auditory labels, each of which labeled a category
321   of six visually coherent Vinčasso creatures during the first training block. Half of the participants
322   heard congruent, and half, incongruent item labels. Labels were yoked to categories for all
323   participants in a given condition.

324   **Training.** Figure 3 shows the results from the training and the test phase of Experiment 2, with
325   the means for individual participants shown separately. Means are given in Table 2. As in
326   Experiment 1, participants solved the categorization task, with a reduction in errors from the first
327   to the second block (Training Phase: $F(1, 38) = 9.49$, $p = .004$, $_p\eta^2 = .20$, Observed Power =
328   .85). However, unlike in Experiment 1, participants exposed to labels correlated with the
329   category structure were, on the whole, *slower* to make their category judgments in the second
330   training block, where category labels were not presented ($F(1, 38) = 7.70$, $p = .009$, $_p\eta^2 = .17$,
331   Observed Power = .72). This effect was not predicted, but suggests that in the first block,
332   participants may have been relying on the diagnostic properties of the audio label to make their
333   category guesses, and subsequently found judgments harder in the silent second block when this
334   label was absent. As elsewhere, no influence of sound symbolism was observed on error rates or
335   reaction times during training. As this effect is somewhat under-powered, they should be treated
336   with caution, and a larger sample size would be required for future investigations seeking to
337   replicate this effect.

338   **Test.** As in Experiment 1, participants made more errors accepting an identical pair than they

339   did rejecting a mismatch (Match Type: F (1, 38) = 14.00, p = .001, $_p\eta^2$ = .27, Observed

340   Power = .95). Unlike in Experiment 1, Reaction Time was influenced by a combination of Match

341   Type and Congruence (Match x Congruence: F (1, 38) = 21.89, p < .001, $_p\eta^2$ = .37, Observed

342   Power = .99), such that the majority (17 of the 21) of participants trained with congruent labels

343   were faster to accept matches than to reject mismatches, while the majority (17 of the 19) of

344   participants trained with incongruent labels were faster to reject mismatches than to accept

345   matches. This finding suggests that passive exposure to labels during training generates

346   differences in object processing, which are in line with the idea that congruent labels facilitate

347   recognition of salient feature similarities, while incongruent labels highlight featural differences.

348   There was also a general difference between the RTs of people in the different training

349   conditions (F(1, 37) = 4.65, p = .035, $_p\eta^2$ = .11, Observed Power = .56), with people in the

350   incongruent condition slightly faster overall during testing. The low power of this group

351   difference suggests it may be artefactual, and a larger sample size would be required for

352   replication.

353
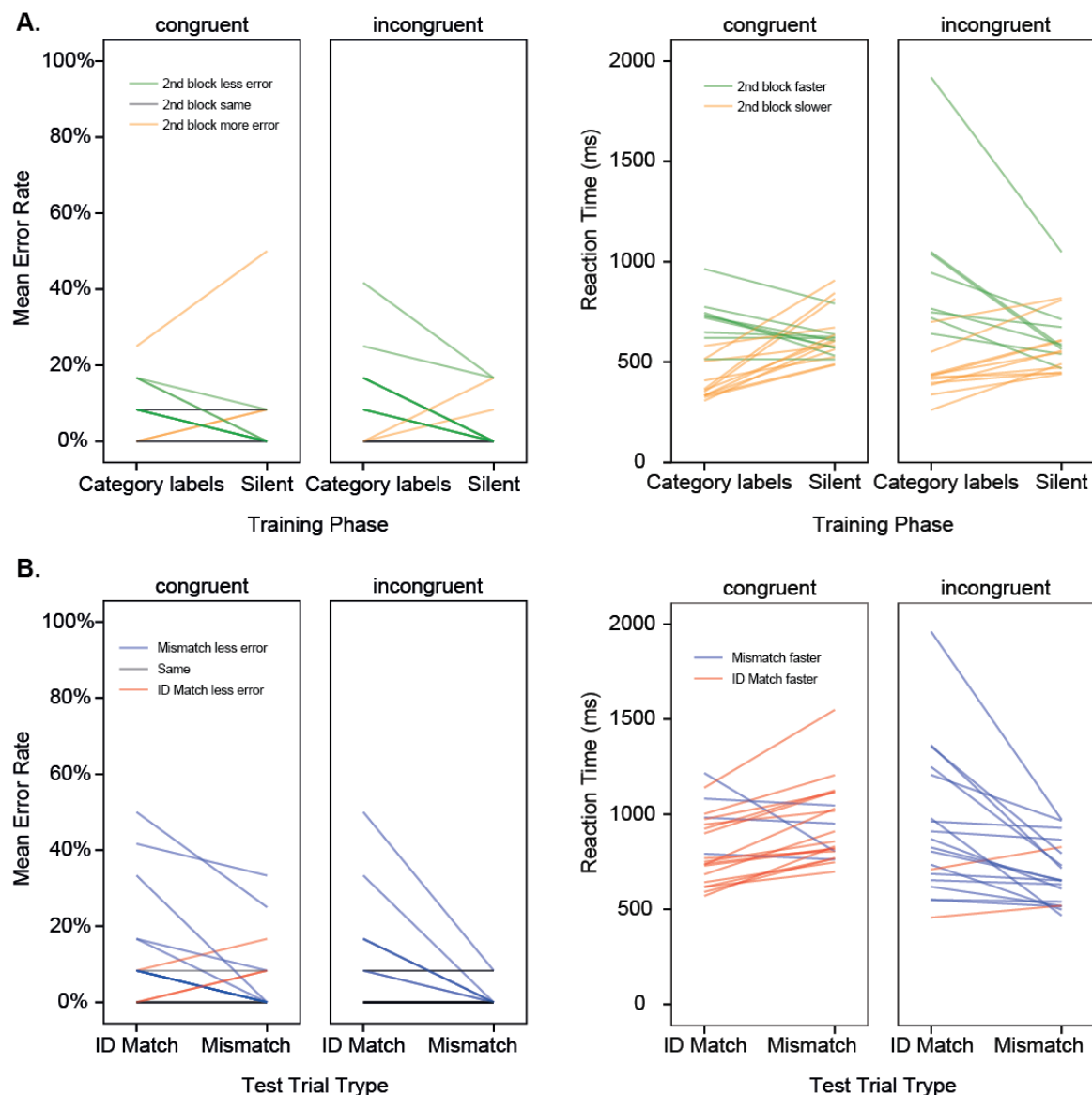
Figure 3. Experiment 2. Training with category labels. A. Results for the two training phases shown for individual participants in different training conditions. Mean error rate, and Mean reaction time (back-transformed from iRT) are shown, where GREEN indicates better performance in the second block (more accurate, faster). B. Results for the two test trial types shown for individual participants in different training conditions. Mean error rate, and Mean reaction time (back-transformed from iRT) are shown, where BLUE indicates better performance for the Mismatch condition (more accurate, faster).

361

362

363

364   Table 2. Mean Error rates and Reaction Times (back-transformed from iRT) for Experiment 2.

		Training				Test			
		Training 1		Training 2		ID Match		Mismatch	
		M	SD	M	SD	M	SD	M	SD
Error Rates	Congruent	.08	.07	.05	.11	.11	.14	.05	.09
	Incongruent	.10	.11	.03	.06	.09	.13	.02	.03
RT (ms)	Congruent	534	193	631	117	829	190	940	204
	Incongruent	663	385	601	156	918	371	687	165

365

Discussion

367   The key finding of the present research is that the sound symbolism in labels with

368   different ranges of reference (category-label vs item-label) has different outcomes on object

369   processing. Passive exposure to category or item labels during category learning influences later

370   abilities to detect similarities and differences between pairs of visual stimuli: as predicted,

371   category labels influenced the speed of recognizing identical pairs if trained with congruent

372   labels, and the speed of rejecting visually dissimilar pairs, when trained with incongruent labels.

373   In addition, labels disrupted the ability to correctly reject a gross visual difference between two

374   images if the items were trained with incongruent labels. These effects are best understood as

375   arising from differences in visual object encoding grounded in the sound symbolism encountered

376   during the initial training stage. Sound symbolism therefore modulates visual object attention

377   differently when labels align with task-relevant category structure.

378   Until now, the majority of previous studies into sound symbolism have focused on the

379   learnability, guessability or recall of the mapping between a label and its referent. For example,

380   Kovic and colleagues (2010) demonstrated sound symbolic effects in a test phase where

381   participants heard an auditory label, then saw a matching or mismatching picture; Monaghan and

382   colleagues' (2011) participants heard an auditory label and guessed from an array of 12 pictures;

383   and Imai and colleagues (2008) tested participants on extension of action descriptions to novel

384   agents in a two-alternative forced choice task. In all of these studies, the range of reference of the

385   label was unclear, as there was no experimental control of how labels aligned with other

386   information about category structure.

387   By contrast, in the current study, we wanted to know whether sound symbolic relations

388 encountered during training would change the way that visual objects were encoded, leading to
389 downstream differences in visual object processing, even in the absence of the label. We found
390 that sound symbolism between a label and a category of objects generated predicted differences
391 in visual processing, but that sound symbolism between individual items and their labels did not
392 generate the enhancements that might typically be predicted, in this categorization paradigm.
393 That is to say, sound symbolism affects the formation of category attributes, not item attributes,
394 and those attributes influence later visual object processing. To the best of our knowledge, this is
395 the first study to show that sound symbolism influences visual object processing outside of the
396 mapping between label and referent.

397   Participants in both experiments were clearly able to infer the category structure from the
398 trials presented in the first training block, as their accuracy improved in the second (silent)
399 training block, where errors fell to below 10% for participants in all training conditions. In
400 previous research (Kovic, et al., 2010), sound symbolism did not influence the trajectory of
401 category learning in a learning-by-guessing categorization task, with task irrelevant category
402 labels. In the present research, we replicated and extended this observation, with no difference in
403 category learning for congruent versus incongruent labels, when trained with *either* item or
404 category labels (Experiment 1 and 2, respectively).

405   Even though the labels presented during the first training block were irrelevant to the task
406 presented in the subsequent test blocks, sound symbolism had pervasive influences on
407 performance in the silent, visual discrimination task: Firstly when trained on categories where
408 each item had different label, participants were worse at rejecting differences than would be
409 typically expected in a task of this kind if they have been trained with congruent labels. This
410 suggests that the normal pattern of object processing may have been disrupted by exposure to
411 incongruent labels. This effect was not predicted, and was somewhat underpowered, so should be
412 treated with some degree of caution. However, if it replicates in future studies, it provides an
413 interesting counterpoint to the perspective that arbitrariness enhances object individuation – it
414 may be the case that *incongruent sound-symbolism* is qualitatively different to other kinds of
415 artbitrariness.

416   By contrast, when trained on items in labelled categories, participants were better at
417 detecting a match when the labels were congruent, and better at detecting a mismatch when the

418   labels were incongruent (as shown by their RTs). Hence, sound symbolism in category labels
419   enhances similarity judgments when congruent, and enhances object discrimination when
420   incongruent. These results are consistent with the idea that congruent sound symbolism may
421   enhance the processing of shared object features, in a way that enhances similarity judgments,
422   via efficient visual retrieval processes. We also saw incongruent category labels enhance the
423   discrimination of grossly mismatching picture pairs, which may have been due to enhanced
424   recognition of item differences. We therefore see discrete effects of congruent and incongruent
425   sound symbolism at different levels of labelling.

426   The preceding eighty-odd years of research into linguistic sound symbolism has
427   confirmed that certain sounds in language simply 'go better' with certain perceptual experiences,
428   either via general cross-modal correspondences (Spence, 2011), or communication oriented
429   mechanisms of iconicity (Perniss, Thompson, & Vigliocco, 2010). These effects are known to be
430   automatic (Parise & Spence, 2012), unconscious (Hung, Styles, & Hsieh, 2017), and to enhance
431   memory for individual word meanings under certain conditions (Imai, Haryu, & Okada, 2005;
432   Kantartzis, et al., 2011; Lockwood, et al., 2016). The finding of differential effects at different
433   level of labelling draws an interesting parallel with existing research into category learning in the
434   absence of sound symbolism.

435   Previous experimental and modelling work on massed word learning by Monaghan et al.
436   (2011), has shown that both human learners and neural networks will learn to map word-forms to
437   meaning classes faster if the mappings are systematically linked to the meanings, while learning
438   to individuate specific form-meaning pairs is more effective if the mappings are arbitrary. In
439   their studies, the systematicity arises from sound-alike mean-alike relationships analogous to
440   morphology in linguistic systems like grammatical gender (e.g., prefix X always occurs with
441   word type A). Hence, their kind of systematicity is effective at functionally grouping items
442   which share a similar form, while arbitrariness is better at distinguishing the meanings of
443   individual words. By contrast, linguistic sound symbolism (arising as does from iconic matches
444   between the senses), generally functions at the level of individual pairings between sounds and
445   meanings, rather than at the level of grammatical categories or word classes. Hence, where
446   symbolic congruence gives rise to sound-alike mean-alike relations (due to shared sound-
447   meaning matching processes), it may be responsible for a type of systematicity, as the

448 phonological space available for words with similar meanings becomes crowded by an
449 overabundance of similar-sounding words. Gasser (2004), demonstrated that it is possible to
450 efficiently encode a relatively small number of sound symbolic words. The modelling and
451 experimental work of Monaghan, Shillcock, Christinasen and Kirby (2014) has shown a higher
452 proportion of sound-alike mean-alike relations in the early lexicon than the adult lexicon,
453 suggesting that sound symbolism is most useful when the lexicon is still small. This modelling
454 perspective sits well with findings that sound symbolism can facilitate early word learning (Imai,
455 et al., 2008), and may act as a 'Bootstrap' for the acquisition of language (Imai & Kita, 2014).
456 However, as vocabulary increases, confusability between similar sounding items also increases,
457 as the phonological space becomes crowded, therefore necessitating the emergence of more-
458 arbitrary linguistic structures.

459   In the current study, our findings therefore support a dominant theoretical perspective in
460 the field of category labelling: that sound symbolism (like other forms of systematicity) does
461 indeed have its greatest impact at the level of Category labels, rather than Item labels (at least in
462 this task, where category structure is task-relevant), but we also show that the effects of sound
463 symbolism go beyond the 'priming' effects which may be generated by a label, or the strength of
464 the link between the label and referent, as they also influence downstream visual object
465 processing, in the absence of the label.

466   It should be noted that the two testing conditions were not of the same kind: in the Match
467 condition, the decision can be made purely on visual grounds, and the category of the object
468 cannot form part of the decision process. Whereas in the Mismatch condition, either the learned
469 category difference or the visible item differences could be cues to correctly identifying the
470 mismatch. This discrepancy arises because of an unexpected ambiguity in the wording of out test
471 question, which rendered a third experimental condition un-analysable (For further details, see
472 Supplementary Materials S3). This means at this stage we are not in a position to diagnose
473 whether the effects arise out of purely visual processing, or a combination of visual and category
474 cues. Further studies will be needed to clarify this point, along with the remaining question of
475 whether non-sound-symbolic but otherwise systematic labels generate similar effects.

476

477   ———

478   Acknowledgements. Thanks go to Ivana Todorović for creating the original Vinčasso

479   drawings, according to the design parameters, and to Anđela Šoškić for assisting in the collection

480   of stimulus rating and cloud guessing data. The Cloud Appreciation Society deserves thanks for

481   the final author's sub-ordinate category knowledge of clouds. Preliminary results from some

482   conditions of this project were presented in the poster 'Ković, V., Sučević, J. & Styles, S.J.

483   "What's in a name?" Category label vs identity label in learning novel categories.' Dubrovnik

484   Conference on Cognitive Science: Language and Conceptual Development, Dubrovnik, Croatia,

485   May 2014.

486

487

488   References:

489   Althaus, N., & Mareschal, D. (2014). Labels direct infants' attention to commonalities during
490   novel category learning. *PloS one, 9.*(7), e99670.
491   Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-
492   month-old infants. *Journal of experimental child psychology, 151*, 5-17.
493   Berlin, B. (1995). Evidence for pervasive synesthetic sound symbolism in ethnozoological
494   nomenclature. In L. Hinton, J. Nichols & J. J. Ohala (Eds.), *Sound Symbolism* (pp. 76-
495   93): Cambridge University Press.
496   Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015).
497   Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences,*
498   *19*(10), 603-615.
499   Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What
500   sound sybolism can and cannot do: Testing the iconicity of ideophones from five
501   langauges. *Language, 92*(2), e117-e133.
502   Drijvers, L., Zaadnoordijk, L., & Dingemanse, M. (2015). *Sound-Symbolism is Disrupted in*
503   *Dyslexia: Implications for the Role of Cross-Modal Abstraction Processes.* Paper
504   presented at the 37th Annual Meeting of the Cognitive Science Society.
505   Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues.
506   *Cognition, 143*, 93-100.
507   Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using
508   G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods,*
509   *41*, 1149-1160.
510   Flumini, A., Ranzini, M., & Borghi, A. M. (2014). Nomina sunt consequentia rerum–Sound–
511   shape correspondences with every-day objects figures. *J Mem Lang, 76*, 47-60.
512   Gasser, M. (2004). *The origins of arbitrariness of language.* Paper presented at the 26th Annual
513   Conference of the Cognitive Science Society, .
514   Hung, S.-M., Styles, S. J., & Hsieh, P.-J. (2017). Can a word sound like an object looks before

515   you have seen it? Sound-shape mapping prior to conscious awareness. *Psychological*
516   *Science, 28*(3), 263-275.
517   Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action
518   events: Are verb meanings easier to learn than noun meanings for Japanese children?
519   *Child Development, 76*(2), 340-355.
520   Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language
521   acquisition and language evolution. *Philosophical Transactions of the Royal Society of*
522   *London. Series B, Biological Sciences, 369*(20130298), 1-14.
523   Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism between a word and an
524   action facilitates early verb learning. *Cognition, 109*, 54-65.
525   Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning
526   in English-speaking children. *Cognitive Science, 35*, 575-586.
527   Köhler, W. (1929:1947). *Gestalt psychology*. New York: Liveright Publishing.
528   Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition,*
529   *114*, 19-28.
530   Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral,
531   developmental, and neuroimaging research into sound-symbolism. *Frontiers in*
532   *Psychology, 6*(1246), 1-14.
533   Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-Symbolism Boosts Novel Word
534   Learning. *Journal of Experimental Psychology: Learning, Memory & Cognition,*
535   *Advance online*. doi: http://dx.doi.org/10.1037/xlm0000235
536   Lupyan, G., Rakinson, D. H., & McClelland, J. L. (2007). Language is not Just for Talking:
537   Redundant Labels Facilitate Learning of Novel Categories. *Psychological Science,*
538   *18*(12), 1077–1083. doi: doi:10.1111/j.1467-9280.2007.02028.x.
539   Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of
540   concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General,*
541   *141*(1), 170-186. doi: 10.1037/a0024904
542   Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound–shape
543   correspondences in toddlers and adults. *Developmental Science, 9*(3), 316-322.
544   Monaghan, P., & Christiansen, M. H. (2006). *Why form-meaning mappings are not entirely*
545   *arbitrary in language* Paper presented at the Proceedings of the 28th Annual Conference
546   of the Cognitive Science Society.
547   Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The Arbitrariness of the sign:
548   Learning advantages from the structure of the vocabulary. *Journal of Experimental*
549   *Psychology: General*(140(3)), 325-347.
550   Monaghan, P., Mattock, K., & Walker, P. (2012). The Role of Sound Symbolism in Language
551   Learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 38*(5),
552   152-1164.
553   Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is
554   language? *Philosophical Transactions of the Royal Society of London. Series B,*
555   *Biological Sciences, 369*(20130299).
556   Nielsen, A. K. S., & Rendall, D. (2011). The Sound of Round: Evaluating the Sound-Symbolic
557   Role of Consonants in the Classic Takete-Maluma Phenomenon. *Canadian Journal of*
558   *Experimental Psychology, 65*(2), 115-124.
559   Nielsen, A. K. S., & Rendall, D. (2013). Parsing the Role of Consonants Versus Vowels in the
560   Classic Takete-Maluma Phenomenon. *Canadian Journal of Experimental Psychology,*

561   *67*(2), 153-163.
562 Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences
563   facilitate word learning. *Cognition, 112*, 181-186.
564 Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound
565   symbolism: a study using the implicit association test. *Experimental Brain Research,*
566   *220.3-4*(3-4), 319-333.
567 Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of
568   language: evidence from spoken and signed languages. *Frontiers in Psychology, 1*(227).
569   doi: 10.3389/fpsyg.2010.00227
570 Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A Window Into Perception,
571   Thought and Language. *Journal of Consciousness Studies, 8*(123-34).
572 Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 225–
573   239.
574 Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception and*
575   *Psychophysics, 73*, 971-995.
576 Sučević, J., Savić, A. M., Popović, M. B., Styles, S. J., & Ković, V. (2015). Balloons and
577   bavoons vs spikes and shikes: ERPs reveal shared neural processes for shape-sound-
578   meaning congruence in words, and shape-sound congruence in pseudowords. *Brain &*
579   *Language, 145-146*, 11-22. doi: http://dx.doi.org/10.1016/j.bandl.2015.03.011
580
581