PAPER

WILEY **Developmental Science**

Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition

Mathilde Fort[1,2,*] | Imme Lammertink[3,*] | Sharon Peperkamp[4,5] | Adriana Guevara-Rukoz[4] | Paula Fikkert[6] | Sho Tsuji[4,7]

[1]Department of Speech and Cognition, GIPSA- Lab, CNRS UMR 5216, Inria, Grenoble Institute of Engineering, University of Grenoble Alpes, Grenoble, France

[2]CNRS UMR 5105, LPNC, University of Grenoble Alpes, University Savoie Mont Blanc, Grenoble, France

[3]Amsterdam Center for Language and Communication, University of Amsterdam, The Netherlands

[4]LSCP, Département d'Etudes Cognitives, ENS, EHESS, CNRS, PSL University, Paris, France

[5]Faculté de Médecine Paris Descartes, Maternité Port-Royal, AP-HP, France

[6]Center for Language Studies, Radboud University, Nijmegen, The Netherlands

[7]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Sho Tsuji, University of Pennsylvania, Department of Psychology, 425 S University Ave., Philadelphia, PA 19104, USA.
Email: tsujish@gmail.com

Abstract

Adults and toddlers systematically associate pseudowords such as "bouba" and "kiki" with round and spiky shapes, respectively, a sound symbolic phenomenon known as the "bouba-kiki effect". To date, whether this sound symbolic effect is a property of the infant brain present at birth or is a learned aspect of language perception remains unknown. Yet, solving this question is fundamental for our understanding of early language acquisition. Indeed, an early sensitivity to such sound symbolic associations could provide a powerful mechanism for language learning, playing a bootstrapping role in the establishment of novel sound–meaning associations. The aim of the present meta-analysis (SymBouKi) is to provide a quantitative overview of the emergence of the bouba-kiki effect in infancy and early childhood. It allows a high-powered assessment of the true sound symbolic effect size by pooling over the entire set of 11 extant studies (six published, five unpublished), entailing data from 425 participants between 4 and 38 months of age. The quantitative data provide statistical support for a moderate, but significant, sound symbolic effect. Further analysis found a greater sensitivity to sound symbolism for bouba-type pseudowords (i.e., round sound-shape correspondences) than for kiki-type pseudowords (i.e., spiky sound-shape correspondences). For the kiki-type pseudowords, the effect emerged with age. Such discrepancy challenges the view that sensitivity to sound symbolism is an innate language mechanism rooted in an exuberant interconnected brain. We propose alternative hypotheses where both innate and learned mechanisms are at play in the emergence of sensitivity to sound symbolic relationships.

RESEARCH HIGHLIGHTS

- By means of a meta-analysis a moderate, but significant, sensitivity to sound-symbolism in early language learning is evidenced.
- A greater sensitivity to bouba-type ("bouba" effect) rather than to kiki-type pseudowords ("kiki" effect) is demonstrated. The kiki effect only emerges as age increases.
- The view that sensitivity to sound symbolism is robustly rooted in an exuberant interconnected brain present at birth is challenged.

- We propose alternative hypotheses where both innate and learned mechanisms are at play in the emergence of the sensitivity to sound symbolism.

1 | INTRODUCTION

The sound of a word does not generally provide information about its meaning. The assumption that the relationship between the sound and the meaning of a word is arbitrary has been held as the conventional view (de Saussure, 1959). However, most languages

also contain a non-negligible proportion of *iconic* words in their lexicons (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016; Perry, Perlman. & Lupyan 2015; see Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015, for a larger review on different types of non-arbitrary relationships in the lexicon, including iconicity). In a study analyzing nearly two-thirds of the world's languages, Blasi et al. (2016) showed that a large portion of basic words carry associations between certain speech sounds and their meaning. For instance, they consistently found the vowel /i/, which is produced with a small mouth aperture, to be often associated with meanings related to "small"; they also evidenced that the phonemes /p/ and /b/, which are produced by constricting the mouth opening and building up pressure on the mouth's walls, are consistently associated with the concept of "full" (see Blasi et al., 2016, for additional examples of iconicity). Interestingly, more recent research has shown that iconicity is more frequent for words that are acquired early in life (Perry et al., 2015). Thus, rather than a marginal phenomenon in the structure of the (early) lexicon, the evidence for the presence of this sound symbolic pattern suggests that it may well play a role in language acquisition. For instance, sound symbolism could play a bootstrapping role in the establishment of novel sound–meaning associations in the very early stages of word learning, when children's vocabulary environment is relatively small and cannot take advantage of language cues learned through experience to deal with referential ambiguity (Imai et al., 2015; Imai & Kita, 2014; Monaghan, Mattock, & Walker, 2012; Monaghan, Shillcock, Christiansen, & Kirby, 2014; Perry et al., 2015).
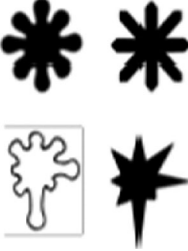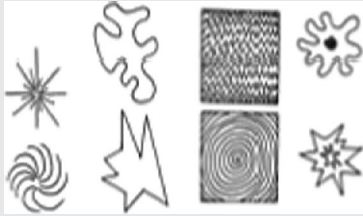
In line with the bootstrapping hypothesis, experimental evidence indicates that adults and toddlers are sensitive to sound symbolism when asked to form and memorize novel associations between phonemes and different properties of visual stimuli (see Lockwood & Dingemanse, 2015, for a review). For instance, they consistently associate certain pseudowords, such as "bouba" or "maluma", with round shapes, and others, such as "kiki" or "takete", with spiky ones, a phenomenon called the bouba-kiki (or, alternatively, the maluma-takete) effect (Köhler, 1947; Maurer, Pathman, & Mondloch, 2006; Ramachandran & Hubbard, 2001). Moreover, both adults and children memorize more easily novel sound–meaning mappings when the pseudowords are sound symbolic compared to when they are not, suggesting that sound symbolism facilitates word learning processes (Imai et al., 2015; Imai, Kita, Nagumo, & Okada, 2008; Kantarzis, Imai, & Kita, 2011; Monaghan et al., 2012). Similarly, more recent findings demonstrate that compared to adults, young children hear and tend to produce sound symbolic words more frequently than non-symbolic words in their spontaneous speech (Perry, Perlman, Winter, Massaro, & Lupyan, 2017). However, whether the sensitivity to sound symbolic relationships in language is present in their earliest stages of life (Ozturk, Krehm, & Vouloumanos, 2013; Peña, Mehler, & Nespor, 2011; Spector & Maurer, 2009; Walker et al., 2010, 2014) or emerges as a consequence of exposure to cross-modal statistical regularities in the environment (Fernández-Prieto, Navarra, & Pons, 2015; Lewkowicz & Ghazanfar, 2009; Lewkowicz & Minar,

2014) remains a matter of debate. Similarly, to what extent sensitivity to sound symbolic relationships is generalizable across different types of sound–shape correspondences is an open question. With regard to the bouba-kiki effect, it has been suggested that older participants are less sensitive to kiki-type than bouba-type associations (Lammertink, Tsuji, & Fikkert, 2015, Jones, Vinson, Clostre, Lau, Santiago, & Vigliocco, 2014, Ozturk et al., 2013). While prior research on infants has rarely reported results on round and spiky correspondences separately, a general preference for round shapes has been reported in infants (Jadva, Hines, & Golombok, 2010; Quinn, Brown & Streppa, 1997). Thus, the literature provides mixed results with regard to both age and type of sound–shape associations. Furthermore, the differences in methods, stimuli (shapes and pseudowords), and relevant participant characteristics (e.g., age and native language) render it difficult to find a unifying answer. The aim of the present paper is to quantitatively summarize the extant literature on the bouba-kiki effect in children before age 3, by means of a meta-analysis.

Several studies report that young children are sensitive to sound symbolic pairings between visual properties of objects and speech sounds (Asano et al., 2015; Imai et al., 2015; Imai et al., 2008; Kantartzis et al., 2011; Maurer et al., 2006; Spector & Maurer, 2013). For instance, using a two-alternative forced choice task, Maurer et al. (2006) showed that 3-year-old English-learning children show a bouba-kiki effect, associating pseudowords such as "kaykee" to spiky objects and "boobaa" to round objects. More recent evidence using event-related potentials (ERPs) demonstrates that even at 11 months of age, infants are sensitive to similar sound–shape associations. In each trial, infants were presented with a visual stimulus (e.g., a round shape) followed by a novel spoken pseudoword that either sound symbolically matched (e.g., "moma") or mismatched (e.g., "kipi") the shape (Asano et al., 2015). The results showed an increased N400 response in the mismatch as compared to the match condition, suggesting that sound–shape associations were more difficult to establish in the sound symbolically mismatching than in the sound symbolically matching condition. In addition, other evidence suggests that similarly to adults (Kovic, Plunkett, & Westermann, 2010; Monaghan et al., 2012; Nygaard, Cook, & Namy, 2009), children as young as 14 months of age benefit from sound symbolism when asked to associate a novel spoken label with a new meaning (Imai et al., 2008, 2015; Kantartzis et al., 2011).

Only one study (Ozturk et al., 2013) suggested that younger, preverbal infants are sensitive to the bouba-kiki effect. Using a preferential listening procedure, Ozturk and colleagues showed that 4-month-old English-learning infants looked longer at mismatching (round shape + "kiki" or spiky shape + "bubu") than at matching/congruent (round shape + "bubu" or spiky shape + "kiki") sound–shape pairings, suggesting that infants considered as "novel" or "surprising" the incongruent/mismatching association (see Methods section for more discussion about the direction of this effect). Note that this study used only two stimulus pairs, that the tested sample was small (N = 12) and that the effect was limited in scope: Contrary to adult

TABLE 1 Overview of studies included in meta-analysis

No.	Author	Year	Age (days)	Language	Auditory stimuli	Visual stimuli	Method	Dependent variable	ES type	Number of ES
1.	Asano, Imai, Kita, Kitajo, Okada, & Thierry	2015	370	Japanese	**Bouba-type:** /moma/ **Kiki-type:** /kipi/		EEG	Peak Amplitude	RS	2
2.	Fort, Weiss, Martin, & Peperkamp	2013	Exp 1: 164 Exp 2: 206	French	**Bouba-type:** /buba/, /voga/, /ʒuvo/, /gudu/, /bodu/, /daʒo/, **Kiki-type:** /pɛ̃fi/, /tesi/, /sitɛ̃/, /kike/, /fekɛ̃/, /tepe/		LWL	Looking Ratio	RS	8
3.	Fort, Guevara-Rukoz, & Peperkamp	2015	130	French	**Bouba-type:** /bu/, /lu/, /mu/ **Kiki-type:** /pi/, /ti/, /ki/		LWL	Looking Ratio	RS	2
4.	Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada, & Kita	2015	442	Japanese	**Bouba-type:** /moma/ **Kiki-type:** /kipi/		CF + LWL	Looking Time	C	1
5.	Lammertink, Tsuji, & Fikkert	2015	Exp. 1: 274 Exp. 2: 761 Exp. 3: 1156	Dutch	**Bouba-type:** /buba/, /modo/ **Kiki-type:** /kiki/, /tekka/		LWL	Looking Ratio	RS	6
6.	Maurer, Pathman, & Mondloch	2006	973	English	**Bouba-type:** "Baa-moo"/"Boo-baa"/"Go-Gaa"/"Maa-boo-maa" **Kiki-type:** "Kuy-tay"/"Kay-kee"/"Tee-tay"/"Tuh-kee-tee"		FC	Target selection	C	1
7.	Ozturk, Krehm, & Vouloumanos	2013	Exp. 1: 126 Exp. 2: 130 Exp. 3: 131	English or Turkish?	**Bouba-type:** /bubu/, /kuku/, **Kiki-type:** /kiki/, /kuku/		CF	Looking Time	RS	6
8.	Pejovic, & Molnar	2017	Exp. 1: 126 Exp. 2: 365	Basque-Spanish	**Bouba-type:** /buba/ **Kiki-type:** /kike/		CF	Looking Time	RS	4

(Continues)

TABLE 1 (Continued)

No.	Author	Year	Age (days)	Language	Auditory stimuli	Visual stimuli	Method	Dependent variable	ES type	Number of ES
9.	Spector	2009	986	English	**Bouba-type:** /roro/, /lolo/, /wowo/, /yoyo/ **Kiki-type:** /riri/, /lili/, wiwi/, yiyi/		FC	Target Selection	RS	2
10.	Spector & Maurer	2013	1034	English	**Bouba-type:** Gofo/Bobo/Koko/Dodo **Kiki-type:** Gigi/Bibi/Kiki/Didi		FC	Target selection	RS	2
11.	Starr & Brennon	2012	Exp. 1: 361 Exp. 2: 366 Exp. 3: 365 Exp. 4: 189 Exp. 5: 184	English	**Bouba-type:** /buba/, /bamu/ **Kiki-type:** /kiki/, /kuhtay/, /taketa/		LWL	Looking ratio	RS	10

Notes: Explanation of abbreviations. Method: EEG = Event-Related Potential; LWL = Looking While Listening; CF = Central Fixation; FC = Forced Choice. ES Type: C = one effect size for round and spiky shapes combined was calculated; RS= A separate effect size was calculated for round and spiky shapes.

control subjects, the infants failed to show a preference in two additional experiments (*N* = 12 in each) in which either the consonants ("kiki" vs. "kuku") or the vowels ("bubu" vs. "kuku") were held constant. Taken together, these results suggest that this type of sound symbolism is weak in the earliest stages of life, challenging the view that this effect is, at birth, robustly rooted in an interconnected brain (Spector & Maurer, 2009).

Finally, several other studies (published and unpublished) failed to replicate the finding that infants are sensitive to the bouba-kiki effect early on in development (Fort, Weiss, Martin, & Peperkamp, 2013; Fort, Guevara-Rukoz, & Peperkamp, 2015; Lammertink et al., 2015; Pejovic & Molnar, 2017; Starr & Brannon, 2012). In order to shed light on this discrepancy, we aimed to provide more general insights into the emergence of the bouba-kiki effect in early language development. To this end, we conducted a meta-analysis of both published and unpublished studies that investigated the emergence of the bouba-kiki effect in young children. We reviewed 11 studies (a total of 22 experiments, 425 participants) on this specific correspondence between pseudowords and shapes, including results from behavioral and electrophysiological methods applied to infants and young children between 4 and 38 months of age. We used 38 months as a cut-off to avoid including children that have undergone orthographical training, since that could be an additional factor influencing sound symbolic sensitivity (Cuskley, Simner, & Kirby, 2015). In addition to estimating the overall evidence for the sensitivity to sound symbolism in these studies, we specifically assessed the effect of age and type of sound–shape correspondence on this sensitivity.

2 | METHODS

2.1 | Systematic literature review

We followed the PRISMA statement (Moher, Liberati, Tetzlaff, Altman, & the PRISMA Group, 2009) for selecting and reporting the studies to be included in our meta-analysis. We decided to include articles if they were assessing the online processing of sound symbolically matching or mismatching sound–shape correspondences related to the bouba-kiki effect (thus, testing both "round" (bouba-type) and "spiky" (kiki-type) correspondences) in children up to and inclusive of 38 months of age. "Matching" responses refer to children's responses to congruent sound–shape associations (bouba-type pseudoword + round object; kiki-type pseudoword + spiky object) and "mismatching" responses refer to children's responses to incongruent sound–shape associations (bouba-type pseudoword + spiky shape; kiki-type pseudoword + round shape), respectively (see Table 1 for more details about the bouba-type and kiki-type auditory and visual stimuli used in each study). Since we were already aware of five published journal articles, as well as five conference presentations or conference proceedings papers (one of which was published during our meta-analysis) that fit our inclusion criteria, and since we considered our strict inclusion criteria to lead to a rather small selection of articles, we continued our literature

search with a seed strategy rather than a broad literature search. We began by assembling five key articles that fit the inclusion criteria (Asano et al., 2015; Imai et al., 2015; Maurer et al., 2006; Ozturk et al., 2013; Spector & Maurer, 2013), as well as two recent review papers on sound symbolism including infant studies (Imai & Kita, 2014; Lockwood & Dingemanse, 2015). For each of these articles, we screened all potentially relevant references cited in it, as well as all articles citing it on scholar.google.com. This search led to one additional eligible study, a dissertation.

We contacted the first or last authors of the eight original articles and conference presentations that were not co-authored by the two first authors of the present article with a request for additional data that was either critical for the computation of effect sizes, or would improve the computation that was already possible based on the available information. Authors of seven of the articles replied, of which five were able and willing to provide us with additional data. We also asked these authors whether they were aware of any additional literature fitting our search criteria, which did not lead to any further hits, leaving us with data from six published articles, four conference presentations, and one dissertation (henceforth collectively referred to as "articles") with sufficient information to include in our meta-analysis (see Table 1).

2.2 | Sample description

The 11 articles were published or presented between 2006 and 2017 by nine different first authors and included data from children between 4 and 38 months of age (range: 126–1156 days). Children were acquiring English (four articles), French (two articles), Japanese (two articles), Dutch (one article), or were from a mixed sample of Basque or Spanish monolinguals or bilinguals (one article). For one additional article, it was not obvious whether children were acquiring English or Turkish.

The dependent variable varied slightly across the 11 articles. The results in seven of the articles were based on looking times, in three articles children were asked to pick a shape (behavioral choice), and in one article the results were based on event-related potentials (ERPs). The way the shapes were presented during the test phase also differed across studies. In eight articles, children were simultaneously presented with a round and a spiky object side-by-side (either live or on a screen) while they were hearing either a bouba-type or a kiki-type pseudoword (looking-while-listening). In four articles, children were presented with either a round or a spiky object while hearing either a bouba-type or a kiki-type pseudoword (central fixation; with one article using both of the former presentation methods in different experiments). In one article, the test phase was preceded by a habituation phase.

The 11 articles included 22 unique experiments. We were able to obtain results of different granularities for different experiments (while aiming for obtaining information as fine-grained as possible for the computation of effect sizes). We obtained results based on a "combined" score (an overall score reflecting the amount of matching

compared to mismatching responses) for two experiments, and separate results for instances in which the match (or the mismatch) was for bouba-type or kiki-type pseudowords, for the rest of the experiments. Of these, we obtained even finer-grained results for one experiment, separated by whether the vowels and consonants in the sound stimuli matched in their sound symbolic shape (e.g., "tiki") or whether they did not (e.g., "tuku"). Overall, we thus computed 44 effect sizes (two experiments with one effect size each, 19 experiments with two effect sizes each, and one experiment with four effect sizes). Of these, 16 effect sizes were derived from published studies, and 28 effect sizes from unpublished studies.

2.3 | Effect size computation

For all the results we calculated Hedges' g effect size, a variant of the standard Cohen's d effect size that corrects for small sample sizes (Hedges, 1981). For results based on two between-participant measurements (e.g., looking times of participants tested in either the match condition or mismatch condition), we used the means and pooled standard deviation of these two measurement points. Similarly, for results based on two within-participant measurements (e.g., individual looking times to match and mismatch conditions, respectively), we also used the means and pooled standard deviation of these two measurements. One complication in this case is that, in order to compute the standard error of the effect size, it is necessary to know the correlation between the two within-participant measurements, which is rarely reported in articles. We were able to obtain these correlations from the authors for 10 out of 16 cases, and we imputed values randomly drawn from the 10 extant values for the six missing cases. In order to test the stability and generalizability of these imputations, we resampled them 100 times and conducted all analyses over these resampled values. These resampling analyses demonstrated stability of our results across different instantiations of imputations, and are reported only when they were informative in addition to the results of the original analysis. Finally, for effect sizes based on a single score (e.g., percentage of correct matches), we calculated effect sizes by subtracting the appropriate chance level from the reported score before dividing by the standard deviation.

Before computing effect sizes, we needed to make a decision regarding the directionality of effect sizes. In the majority of studies, a stronger response for matching compared to mismatching sound–shape correspondences (longer looking times, more pointing, or higher peak ERP amplitudes) was interpreted as expressing sensitivity to sound symbolic matches. This directionality is straightforward for designs in which infants were presented two images simultaneously, such that a higher amount of looks or pointing to images matching the pseudoword would be interpreted as a preference for the match. However, designs that only presented one image while playing either a matching or a non-matching pseudoword were divided on their interpretation of directionality. Two of the four articles with this design (Asano et al., 2015; Fort et al., 2013) interpreted a preference for sound symbolically matching compared to

mismatching sound–shape correspondences (thus, longer looking times or higher peak ERP amplitudes) as an indicator of sensitivity to sound symbolism. However, two others (Ozturk et al., 2013; Pejovic & Molnar, 2017) interpreted the reverse, namely a preference for mismatching correspondences, as an indicator for sensitivity to sound symbolism. Indeed, the problem of not being able to predict which direction of preference infants will show is inherent to designs relying on infants' direction of preference (Bergmann & Cristia, 2016). Faced with this uncertainty, one option (that the authors of the four concerned articles likely followed) is to interpret any difference in response to matching and mismatching trials as a sign that these trial types were discriminated, disregarding the directionality of the effect. However, in a meta-analysis this would mean treating behaviors that go in opposite directions as equivalent without strong theoretical underpinnings. For reasons of consistency and interpretability, we therefore decided to code all studies with the same directionality, thus subtracting responses to sound symbolically mismatching correspondences from responses to matching correspondences. Therefore, positive effect sizes in our analysis indicate a stronger response to matching than to mismatching correspondences, while negative effect sizes indicate a stronger response to mismatching than to matching correspondences. We confirmed in an additional analysis (reported in footnotes) that reversing effect size estimates for the two articles in question did not lead to qualitatively different results.

2.4 | Coding of moderator variables

In addition to assessing the overall median effect size of expected sound symbolic matches, we included three moderator variables in the model: First, based on the mixed results and the question related to the ontogeny of sound symbolism in the literature, we included mean age in days per experiment as a continuous predictor variable. When age was reported in months, we multiplied this number with 30.42. Second, based on the reported difference in infants' processing of round and spiky shapes (Jadva et al., 2010), we included pseudoword type as a second, categorical moderator, as well as its interaction with age, allowing us to assess whether matches involving round and spiky correspondences yield different results and how this interacts with infant age. Third, in order to assess whether published articles lead to systematically different effect sizes than unpublished ones, we added publication status as a third moderator variable. Since we did not have predictions on this variable's interaction with the other moderators, no interaction was included.

2.5 | Data pre-processing and analysis

Prior to our analysis, we checked our dataset of effect sizes for outliers which deviated more than 3 SD from the mean overall effect sizes. No records had to be removed by this criterion. The continuous moderator variable age was log-scaled and centered in order to account for the uneven distribution of age in the sample (see Figure 2). The categorical

moderator variables pseudoword type and publication status were effect coded (−1,1).

First, we assessed funnel plot asymmetry (Egger et al., 1997) using the metafor package version 1.9-8 (Viechtbauer, 2010) in R version 3.3.0 (R Core Team, 2016). Next, the main analysis was performed in two steps, the first of which was an intercept-only model in order to estimate the overall mean effect size without taking into account moderator variables. In the second step, we added pseudoword type, age, and publication status as moderator variables. For both analyses, we used the robumeta package version 1.6 (Fisher & Tipton, 2015) to implement robust variance estimation (Hedges, Tipton, & Johnson, 2010) with small sample correction. In a nutshell, robust variance estimation allows for dealing with dependent effect sizes for which the covariance structure is unknown. In the present meta-analysis, the effect sizes for the bouba-type and kiki-type pseudoword conditions (in case they were reported separately) were always based on measurements on the same children, making them statistically dependent. In order to perform a meta-analysis, this case usually requires the knowledge of the covariance structure of the effect sizes, which was unknown to us. Therefore, we used robust variance estimation, which provides an estimator of these unknown covariances by taking into consideration the estimated residual vectors for each dependent set of effect sizes. Effect size weighting was performed based on inverse variance weighting of these estimated variances.

2.6 | Community access

A common problem with meta-analyses is that they are crystallized at the point of publication, at which point no new data points can be added. To avoid this problem, we adopted the approach of community-augmented meta-analyses (Tsuji, Bergmann, & Cristia, 2014) by making the entered data openly accessible on MetaLab (metalab.stanford.edu), allowing interested researchers to interact with the data for their own purposes, as well as to request to add new data points. In addition, we made our analysis scripts available on our Open Science Framework (OSF) project page at https://osf.io/wshdy/. There, the interested reader can also find some supplementary, exploratory analyses, for instance how testing method and native language relate to effect sizes.

3 | RESULTS

3.1 | Assessment of publication bias

Figure 1 shows the funnel plot of the dataset. Funnel plots are a potential indicator of publication bias such that an asymmetry can indicate selective publishing. In a funnel plot, effect sizes are plotted against their standard error as a measure of study size and study precision. Studies with higher precision are expected to be closer to the true effect size and thus cluster around the middle, while studies with lower precision are expected to spread to both sides. To test whether this expectation was met, we performed a regression test on funnel plot asymmetry. This test did not reach significance (z =
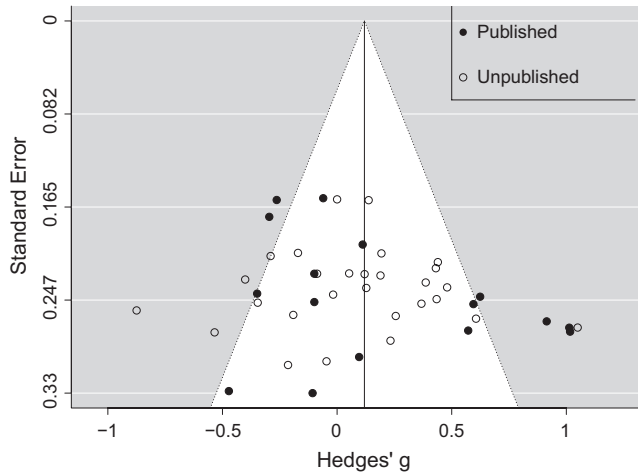
FIGURE 1 Funnel plot of effect sizes by publication status. The vertical line indicates the overall model estimate. The triangle-shaped unshaded region represents a pseudo confidence interval region is with bounds equal to ± 1.96 *SE*

0.812, *p* = .417), indicating that there was no evidence of publication bias. We further explored whether there were differences between the studies published in a peer-reviewed (*N* = 6) journal on the one hand and unpublished studies (*N* = 5) on the other hand. This analysis revealed that there was no evidence for an asymmetry in either the subset of published studies (*z* = 0.650, *p* = .516), or the subset of unpublished studies (*z* = −0.093, *p* = .926).

3.2 | Meta-analytic regression

As a first step, we constructed a linear meta-analytic regression intercept-only model to assess the overall median effect size, assuming the unknown correlation between effect sizes to be the default of rho = 0.8. A sensitivity test revealed that outcomes were not affected by other values of rho between 0 and 1. The model estimate was 0.163 (*SE* = 0.075) and differed significantly from zero [*t*(20.7)

= 2.19, *p* = .040, CI_L = 0.008, CI_U = 0.318]. This shows that children overall were sensitive to sound symbolic matches.[1] The total amount of between-study variance (i.e., τ^2) was 0.14 (calculated using the method-of-moments estimator provided in Hedges et al., 2010). The variability in effect size estimates explained by heterogeneity rather than sampling error was I^2 = 70.79%.

Our second model introduced the three moderators pseudoword type, age, publication status, and the interaction between pseudoword type and age (Figure 2). Note that we excluded the two effect sizes that were based on a combined condition measure from this analysis. The unknown correlation was again set to default rho = 0.8, and replacing this value with others did not affect the results. The intercept in this moderator model was not significant [estimate = 0.085, *SE* = 0.075, *t*(10.84) = 1.14, *p* = .279, CI_L = −0.079, CI_U = 0.249], suggesting that some of the overall sensitivity to sound symbolic matches was taken away by the introduction of moderators. The effect of pseudoword type [estimate = 0.173, *SE* = 0.060, *t*(17.48) = 2.92, *p* = .009, CI_L = 0.048, CI_U = 0.299] reached significance, with sound symbolic matches for bouba-type pseudowords having a significantly higher mean effect size (*M* = 0.270, *SD* = 0.381) than matches for kiki-type pseudowords (*M* = −0.099, *SD* = 0.343). The main effect of age [estimate = 0.117, *SE* = 0.071, *t*(7.70) = 1.64, *p* = .142, CI_L = −0.049, CI_U = 0.283], of publication status [estimate = 0.015, *SE* = 0.083, *t*(10.03) = 0.186, *p* = .856, CI_L = −0.170, CI_U = 0.200], and the pseudoword × age interaction [estimate = −0.035, *SE* = 0.058, *t*(7.66) = −0.603, *p* = .564, CI_L = −0.170, CI_U = 0.100] did not reach significance.[2] These results demonstrate a difference in processing of round and spiky sound symbolic matches, with children being more sensitive to the former.

To follow up on this difference, we constructed two separate models for the round and spiky sound symbolic matching conditions. The intercept in the model for round matches approached significance [estimate = 0.220, *SE* = 0.106, *t*(10.89) = 2.08, *p* = .062, CI_L = −0.013, CI_U = 0.454]. This tendency was reflected in the analyses based on resampled datasets, which all yielded a *p*-value around .05,
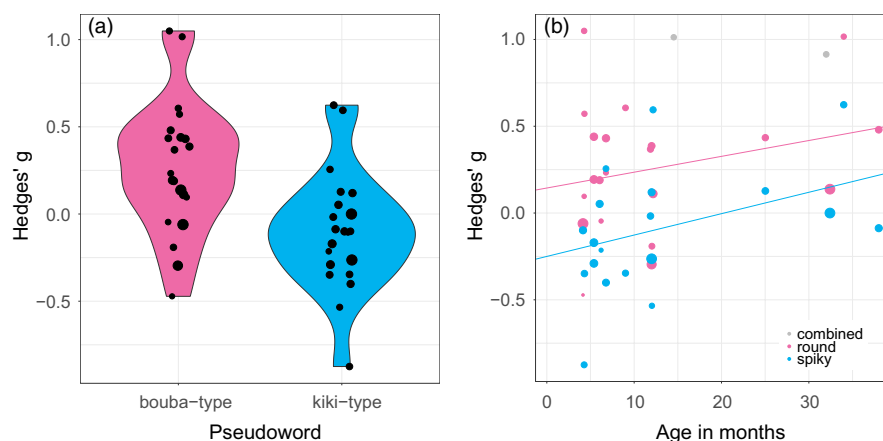

FIGURE 2 (a) Effect sizes by pseudoword type. Violins represent density estimates over underlying data. Dot size represents effect size precision, with larger dots representing more precise effect sizes. (b) Effect sizes by age and pseudoword type. Age is indicated in months rather than log-scaled as for the analysis. Dot size represents effect size precision. Regression lines are based on the intercept and slope values of meta-analytic regressions for bouba-type and kiki-type pseudowords, respectively

with 34 datasets having a *p*-value just below and 67 datasets having a *p*-value just above threshold. There was no effect of age [estimate = 0.061, *SE* = 0.120, *t*(7.69) = 0.506, *p* = .627, CI_L = −0.219, CI_U = 0.341] or publication status [estimate = −0.107, *SE* = 0.120, *t*(10.07) = −0.89, *p* = .394, CI_L = −0.373, CI_U = 0.160].[3] The intercept in the model for spiky matches was not significant [estimate = −0.054, *SE* = 0.085, *t*(11.4) = −0.68, *p* = .511, CI_L = −0.228, CI_U = 0.121]. This model did, however, show a significant effect of age [estimate = 0.168, *SE* = 0.062, *t*(7.73) = 2.71, *p* = .027, CI_L = 0.024, CI_U = 0.311].[4] In descriptive terms, we can observe that studies with children in the group of younger children (between 4 and 15 months of age) had lower effect sizes[5] (*M* = −0.161, *SD* = 0.327) than older children (between 25 and 28 months of age) (*M* = 0.166, *SD* = 0.318). The effect of publication status was not significant [estimate = 0.133, *SE* = 0.087, *t*(9.57) = 1.54, *p* = .156, CI_L = −0.061, CI_U = 0.327]. These follow-up analyses underline the difference in sensitivity to round and spiky matches. In addition, they suggest that sensitivity to spiky—but not round—matches increases with age. As will be taken up in the general discussion, these additional results might help explaining previous divergences in the literature.

4 | DISCUSSION

The present meta-analysis provided an in-depth investigation of whether children before 38 months of age, across a wide range of experimental paradigms and languages, are sensitive to the bouba-kiki effect. More precisely, we assessed whether this effect is present very early on or rather emerges later in language development, and whether sensitivity to round and to spiky sound–shape associations develop simultaneously. Some theories posit that sensitivity to non-arbitrary sound symbolic relationships between a speech sound and its corresponding meaning (e.g., pseudoword–shape associations as in the bouba-kiki effect) is one of the language learning mechanisms present at birth (Ramachandran & Hubbard, 2001; Spector & Maurer, 2009). Others claim that the ability to detect cross-modal correspondences rather emerges with experience (Fernández-Prieto et al., 2015; Lewkowicz & Ghazanfar, 2009; Lewkowicz & Minar, 2014). By means of a meta-analysis, we found overall a greater sensitivity to sound symbolism for bouba-type pseudowords (i.e., round sound–shape correspondences) than for kiki-type pseudowords (i.e., spiky sound–shape correspondences). Detailed analyses showed that for the kiki-type pseudowords, the effect emerges with age. In the following, we elaborate on why these findings (1) challenge the view that sensitivity to sound symbolism is a robust phenomenon in the earliest stages of life; (2) are mostly compatible with the idea that sensitivity to sound symbolism emerges gradually during (language) development; and (3) could result both from innate and learned mechanisms.

The first central finding of the present work is the overall above-chance effect size of sound symbolic matches in infants' and young children's perception. Thus, despite including a number of unpublished studies that did not themselves find evidence for a bouba-kiki effect, a quantitative examination of all available results shows overall evidence for such an effect, albeit modest in size. Second, when examining the effect of moderator variables on the bouba-kiki effect, we found that the effect of sound symbolism in young children was overall significantly higher for bouba-type pseudowords than for kiki-type pseudowords, as has also been found in adults (Jones et al., 2014). Third, the analyses show that at least part of the bouba-kiki effect emerges postnatally. Our follow-up analyses indeed showed that when considering the kiki-type pseudowords in a separate model, the effect of sound symbolism becomes stronger with age, while no such effect of age was observed for the bouba-type pseudowords. This could either mean that only a "bouba effect" exists early on, or that the sensitivity for spiky pseudoword-shape correspondences (i.e., "kiki" effect) is likewise present but is masked by another type of (possibly methodologically induced) unisensory processing bias. In the first case scenario, the absence of a "kiki" effect could be due to the fact that infants prefer listening to (and thus pay more attention to) bouba-type over kiki-type pseudowords for their acoustic or phonetic features (e.g., low- vs. high-frequency speech sounds, respectively: Trainor, Clark, Huntley, & Adams, 1997) that differentiate these two categories (see Spence, 2011, for a review). The absence of a kiki effect might also be due to the fact that infants prefer watching curved (round) objects over angular (spiky) objects), a preference consistently reported from 3 to 24 months of age in infants (Javda et al., 2010; Quinn et al., 1997), and also in adults (cf. Palumbo, Ruta & Bertamini, 2015). This unisensory bias could have created a floor effect on the kiki-type pseudowords, masking the cross-modal sound symbolic association in that condition. In the present data, however, neither such auditory (mean looking times while listening to bouba-type pseudowords: *M* = 6.92 s, *SD* = 5.3; kiki-type pseudowords: *M* = 6.57 s, *SD* = 4.81, *t* < 1) nor visual preference (mean looking times while listening to round shapes: *M* = 6.86 s, *SD* = 5.1; spiky shapes: *M* = 6.62 s, *SD* =5.1, *t* < 1) was present. Another possibility for the lack of a significant sound symbolic effect for kiki-type pseudowords may come from a discrepancy in the degree of prototypicality of the round and spiky shapes across different experiments. Spiky shapes might be more variable overall or less representative in their degree of spikiness than their round counterparts. The latter could make the recognition of the shape feature more difficult, preventing the establishment of the cross-modal correspondence with the spiky pseudowords. To test this hypothesis, we asked five naïve adult participants to rate each of the shapes provided in the stimuli examples of each of the experiments on a scale going from 1 (very spiky) to 7 (very round) (see Table 1 for more details). We then averaged these results, obtaining one score for each shape type (round shapes: *M* = 5.34, *SD* = 0.65; spiky shapes: *M* = 1.57, *SD* = 0.43). Both of these measures were significantly different from the extreme values (i.e., 1 and 7: all *p* < .05, respectively) and the mean value (i.e., 3.5: both *p* < .05) of the scale, with similar between-experiment variability for each shape, making it unlikely that such a difference in the visual stimuli could account for the absence of a significant kiki effect in younger infants. In other words, the present results demonstrate that infants only

show a "bouba" effect in their first stages of development, the "kiki" effect emerging later on.

Before proceeding to our discussion of the consequences of this finding, we want to caution that one alternative explanation for the difference in "kiki" effect between age groups is a difference in testing method: Indeed, the age range for central fixation studies, the method that showed the lowest average effect sizes, was 4–12 months, while studies where children were instructed to point to a shape, the method that showed the highest average effect sizes, were conducted with 32–38-month-old children. Follow-up regression analyses did not, however, confirm such a method effect.[6] While this result is in support of our age effect reported, due to our relatively small sample sizes we cannot exclude the possibility that testing method affects measurable outcomes, and we suggest that future studies carefully consider this possible confounding factor. One experimental way to address this question would be to test prelexical infants (4–5 months of age) using more sensitive and direct measures of perception such as ERPs, like the ones used in 11-month-olds by Asano et al. (2015). In particular, preferential looking paradigms in the present meta-analysis often involve an open behavioral response (e.g., looking more time at the object matching a given speech sound) on top of the perceptual processing per se (e.g., finding the matching/mismatching object), which increases the cognitive demands of the task as compared to paradigms with direct measures of perceptual processing (e.g., N400 event-related potentials and time frequency analysis of the EEG signal, as in Asano et al., 2015). In sum, the higher cognitive demands of preferential looking and/or central fixation paradigms mostly used in the youngest group of infants might have masked the perceptual sensitivity to sound symbolism. For these reasons it might be interesting to consider that such direct measures of infants' perception could provide new insights into the ontological roots of sensitivity to sound symbolism and to its underlying neural mechanisms.

What, then, do these results mean for the role of sound symbolism in language learning? According to the neonatal synesthesia theory, sensitivity to sound symbolic associations is an inherent property of the infant's brain, rooted in an exuberant anatomical connectivity between contiguous sensory and/or sensory-motors areas (Ramachandran & Hubbard, 2001; Spector & Maurer, 2009). This theory predicts that children show a *strong* sensitivity to cross-modal correspondences such as the bouba-kiki effect in their earliest stages of development, and that this sensitivity decreases over time through selective synaptic pruning. The present results are clearly not in accordance with these predictions. While the lack of an age effect for the bouba-type stimuli leaves open the possibility that the bouba effect is indeed innate, the overall modest effect size in the present data and the absence of a kiki effect in the earliest stages of development demonstrate that the overall bouba-kiki effect is *weak* before the age of 3. Moreover, the discrepancy observed between bouba-type and kiki-type associations cannot be explained by the neonatal synesthesia theory as it is formulated by Spector and Maurer (2009). Indeed, there is so far no mechanism in this theory that can account for the pace difference in the *emergence* of a kiki

and a bouba effect, respectively. Instead, the present data suggest that sensitivity to sound symbolism is weak in the first months of life, and gradually emerges during (language) development. This is in line with very recent experimental data, showing that sensitivity to sound symbolic associations exists before 3 years of age but further develops during childhood (until at least 7 years of age), suggesting that basic sensitivity to sound symbolism is present early in life and facilitates children's word–referent mappings, while sensitivity to subtler sound symbolic cues requires greater exposure to the language environment (Tzeng, Nygaard, & Namy, 2017). Our conclusion also gives rise to the possibility that the bouba-type and kiki-type associations are not entirely equivalent, and have different underlying mechanisms. In line with this hypothesis, some research already found that the phonetico-acoustic properties of some pseudowords (i.e., vowel backness) influence adults' pairing with round-like shapes, but not with spiky ones (D'Onofrio, 2014). Further research should be done in order to clearly document the developmental trajectory of this sensitivity to sound symbolism and how it influences the way infants and children learn their native language.

One possibility is that the different developmental timing between the bouba effect and the kiki effect could be rooted in the fact that infants' and adults' visual attention is mostly drawn by round or curved shapes rather than spiky and angular ones (Javda et al., 2010; Palumbo et al., 2015; Quinn et al., 1997). Ontologically, this perceptual bias—that could be in itself innate or learned—could have consequences for the emergence of their sensitivity, favoring bouba-like over kiki-type associations. However, the present data do not entirely support this assumption, as there was no perceptual bias for round over spiky shapes (see below). Another and non-mutually exclusive hypothesis is that infants have more experience with round objects and/or bouba-type words than spiky objects and/or kiki-type words in their direct perceptual environment. While intuitively this seems true, there is no evidence in the literature accounting for such difference. Further research should thus address these questions by conducting corpora analyses, contrasting "bouba-type" vs. "kiki-type" words in infants' and young children's language environment and their use, as well as by measuring the proprioceptive features of objects that are the most frequent in their environment.

Importantly, the present findings preclude any conclusion regarding the ontological origin (innate or learned through experience) of the sensitivity to sound symbolism. One first possibility is that sensitivity to iconic relationships is mostly biologically endowed: early experience with sound symbolic words in development would only act as a "revelator" of this sensitivity. Accordingly, research in visual cognition has shown that while early visual experience is critical for infants to have normal vision, much of the visual system is hard-wired (see Kellman & Arterberry, 2007, for an overview). Here, a weak disposition to map some sounds onto visual properties could be enough to scaffold the association of particular sounds to certain aspects of visual stimuli (e.g., shape), and repeated experience of this sort could develop into the insight that speech sounds have meanings (see Imai & Kita, 2014, for such an hypothesis). In line with this idea, Walker (2016) argues

that visual properties (such as size) could be intrinsically coded with auditory information (such as acoustic frequencies), making the cross-modal correspondence between these two dimensions obvious to the perceiver. For instance, formant frequencies are more spatially dispersed from the fundamental frequency in low- rather than in high-frequency sounds. This could explain another sensitivity to sound symbolism, that is the tendency to associate high-frequency sounds—with small spatial dispersion—with small objects and low-frequency sounds—with greater spatial disper- sion—with larger objects (see, for instance, Walker, 2016, for an overview of pitch-size correspondences). If this hypothesis were true, this would mean that sensitivity to sound symbolism emerges from a direct perceptual association between visual and auditory dimensions, and that this association does not necessarily need to be learned. Of course, this explanation remains hypothetical and further evidence for such dimensions being intrinsically coded un- imodally needs to found, especially in the present case of sound– shape correspondence (i.e., bouba-kiki effect). However, the idea of experience acting as a "revelator" is in line with what we ob- serve in other aspects of early language acquisition. Indeed, one could examine whether this experience of sound symbolic associ- ations is required at a specific critical period in development, as is the case for phonetic learning, both for auditory (Werker & Tees, 1984), visual (Weikum et al., 2007) and audiovisual speech (Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés 2009).

Another potential mechanism is that sensitivity to sound symbol- ism is mostly a learned aspect of perception, resulting from sensory and sensorimotor experiences to cross-modal statistical regularities present in their (linguistic) environment (e.g., Lewkowicz & Minar, 2014; but see Walker et al., 2010; Walker et al., 2014, for more dis- cussion). This idea relies on several assumptions. First, these statis- tical regularities need to be actually present in the infant perceptual input. As previously pointed out, there is no study that determines whether such sound–shape iconicity is present in children's early speech, and whether this is specifically the case for the bouba-type words. However, Monaghan et al., (2012) found that this specific sound-shape iconicity is weak but exists in the adults' English lexi- con. They did not report any discrepancy between bouba-type and kiki-type words, but found that round or curvy objects are slightly more often associated with labels containing voiced consonants with low spectral frequency (e.g., /b/ in "bouba") rather than unvoiced high spectral frequency ones (e.g., /k/ in "kiki"). Another (and per- haps more likely) possible co-occurrence in infants' perceptual en- vironment is that round or curvy objects with soft edges (e.g., ball) are usually made of soft materials and thus tend to often produce low-spectral frequency sounds when they hit a surface (e.g., when they fall on the floor). Conversely, spiky-like objects with sharp edges (e.g., keys) are usually made of hard materials and tend to produce, in the same situation, high-spectral frequency sounds. If, as intuited above, infants indeed get more sensory-motor daily experience with round rather than with spiky-like objects, this could explain why they start noticing round shape–sound correspondences in their linguis- tic and non-linguistic environment before the spiky ones. Second,

infants need to be able to learn to associate visual and auditory infor- mation through these co-occurrences. While co-occurrence has not actually been demonstrated to result in sound symbolic associations in the language domain, evidence exists for it creating cross-modal correspondences in adults, with non-linguistic stimuli outside of the language domain. Ernst (2007) trained participants to discriminate a square that differed from two other squares on some perceptual aspect (luminance, stiffness, or both). During the whole duration of this task, participants were also consistently presented with artificially created cross-modal associations between two sensory dimensions (visual luminance and haptic stiffness of an object) that are usually not present in our sensory environment (e.g., the brighter the object, the stiffer it was). Their results showed that at the end of the test phase, participants' discrimination performance was influenced by the newly learned cross-modal association, when comparing before and after training. This shows that exposure to a new distribution of statistical co-occurrences created sensitivity to a novel cross-modal association, subsequently influencing perceptual judgements. This Bayesian learn- ing mechanism, if it applies to linguistic stimuli, could then explain how infants start noticing such sound–shape correspondences in their perceptual environment, as soon as they are able to form concrete object–label associations (i.e., as early as 6 months of age; Bergelson & Swingley, 2012). Importantly, these cross-modal associations do not need to be experienced in a deterministic way (i.e., with 100% consistency) to be noticed by infants and to support word learning later on. A wide range of literature has shown that from birth, infants are sensitive to statistical regularities in both their auditory and visual environment (see Lany & Saffran, 2013, for a review). More recent results even demonstrate that when given the opportunity to choose, 12-month-old infants prefer learning from probabilistic rather than from deterministic data (Téglás & Bonatti, 2016). For language learn- ing, this would mean that the presence of counter-examples to sound symbolic associations (e.g., round objects associated with "kiki-type" labels or making high-frequency sounds) should not prevent infants from noticing that the opposite pattern (e.g., round objects associated with "bouba-type" labels or producing low-frequency sounds) is the most frequent one. Of course, this idea is speculative, but it suggests that sensitivity to sound symbolic relations can emerge as a result of exposure to linguistic or non-linguistic statistical regularities early on in development, and play a bootstrapping role (Imai & Kita, 2014) in infants' word learning.

5 | CONCLUSIONS AND FUTURE DIRECTIONS

Some readers might be left wondering what to take away from the complexity of these results. One important merit of meta-analyses is their focus on effect sizes rather than significance, thus put- ting emphasis on examining how strong an effect is rather than on whether it exists or not. We argue that the present results, precisely because of their moderate nature, take us an important step fur- ther in the study of sound symbolism. Previous published studies in

infant research have clearly shown that an early sensitivity to sound symbolism does exist, and our meta-analysis has re-emphasized this finding by showing that, even when adding unpublished studies and studies that did not find evidence for an early sensitivity to sound symbolism, the effect is significantly above chance. Given that, a central next step will be to understand *to what extent* and *under which circumstances* this sensitivity influences language learning and usage. To that end, a recent study has investigated adults' sensitivity to a large array of sound symbolic pairings that were more naturalistic than many of the classical stimuli used in previous studies (Dingemanse, Schuerman, Reinisch, Tufvesson, & Mitterer, 2016). Results showed a moderate, but consistent, sensitivity to sound symbolic patterns, where sound symbolic associations were detected with above-chance probability, but not in a deterministic way. As such, these results suggest that sound symbolic matches can be thought of as cues that are used for interpretation when available. Our meta-analytic approach, including studies from a variety of laboratories, using a variety of stimuli, and expressing study results in effect sizes rather than based on *p*-values, also contributes to assessing the extent to which sound symbolic associations play a role in language acquisition. Our present conclusion is indeed consistent with those drawn by Dingemanse et al. (2016). That is, young children will show sensitivity to sound symbolism if presented with it and their language acquisition could be supported by the presence of sound symbolic associations in their native language; however, they do not require sound symbolic vocabulary in their input to acquire language. Apart from providing an estimate of a mean true effect size, our meta-analysis is also a starting point for further study, for instance by focusing on gaps in the literature, especially on the relationship between the structure and the lexicon, perceptual sensitivity to sound symbolism and language acquisition performances. On top of the missing evidence spotted in the previous part of the discussion, researchers might wonder whether children growing up with a language with a large sound symbolic vocabulary would show specific learning advantages or show a stronger or earlier sensitivity to sound symbolism. Another idea would be to test whether languages exist that could serve as counterfactuals (e.g., if there is a language that would yield a kiki effect before the bouba effect, is there a language that would yield neither effect, etc.). In addition, researchers might want to look more closely into the different trajectories for bouba-type and kiki-type associations by investigating the relation between an overall preference for looking at round shapes or listening to bouba-type words on the one hand and developing sound symbolic associations on the other hand. While we do not find evidence for such a preference in the present meta-analysis, the data that were available were restricted; for instance, we did not have access to looking times over the whole trial, but only to specific time-windows. We hope that future research will find our online resource helpful for exploring further questions. Finally, we want to re-emphasize that any future work can be added to our online repository, such that it is not only valuable in itself, but also contributes cumulatively to a quantitative overarching picture of the development of sensitivity to sound symbolism.

ENDNOTES

[1] For the alternative analysis with reversed effect size directionality for two articles (see Section 2.3), analyses showed qualitatively the same results with a significant intercept [estimate = 0.218, SE = 0.066, $t(20.6)$ = 3.31, p = .003, CL_l = 0.081, CL_u = 0.356].

[2] The alternative analysis did again exhibit the same pattern of results, with a marginally significant intercept [estimate = 0.123, SE = 0.067, $t(10.78)$ = 1.83, p = .095, CL_L = −0.025, CL_U = 0.271, and a significant main effect of condition [estimate = 0.198, SE = 0.059, $t(1741)$ = 3.35, p = .004, CL_L = 0.074, CL_U = 0.323].

[3] The alternative analysis for round matches also showed a significant intercept [estimate = 0.288, SE = 0.089 $t(10.61)$ = 3.26, p = .008, CL_L = 0.093, CL_U = 0.484], with no other significant effects.

[4] The alternative analysis for spiky matches also showed a significant effect of age [estimate = 0.193, SE = 0.073, $t(7.72)$ = 2.66, p = .030, CL_L = 0.024, CL_U = 0.361], with no other significant effects.

[5] Age was not continuously distributed in this dataset, with participants grouping into a younger and an older age group (see Figure 2) that we used for our descriptive split. We did not apply a median split to age for these descriptive statistics, since this would have put the boundary very low, between 6 and 7 months of age. However, a median split reveals a similar descriptive tendency.

[6] We have assembled this additional material on our OSF project page.

REFERENCES

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, *63*, 196–205.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 3253–3258.

Bergmann, C., & Cristia, A. (2016). Development of infants' segmentation of words from native speech: A meta-analytic approach. *Developmental Science*, *19*, 901–917.

Blasi, D.E., Wichmann, S., Hammarström, H., Stadler, P.F., & Christiansen, M.H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 10818–10823.

Cuskley, C., Simner, J., & Kirby, S. (2015). Phonological and orthographic influences in the bouba-kiki effect. *Psychological Research*, *81*, 119–130.

de Saussure, F.D. (1959). *Course in general linguistics*. New York: Philosophical Library.

Dingemanse, M., Blasi, D.E., Lupyan, G., Christiansen, M.H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*, 603–615.

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*, e117–e133.

D'Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the bouba-kiki paradigm. *Language and Speech*, *57*, 367–393.

Egger, M., Davey Smith, G., Schneider, M., Minder, C., Mulrow, C., Egger, M., … Olkin, I. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.)*, *315*, 629–634.

Ernst, M.O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7* (7), 1–14.

Fernández-Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior and Development*, *38*, 77–81.

Fisher, Z., & Tipton, E. (2015). Robumeta: Robust variance meta-regression. R package version 1.6. Retrieved from: https://cran.r-project.org/package=robumeta

Fort, M., Guevara-Rukoz, A., & Peperkamp, S., (2015). Looking for the bouba-kiki effect in prelexical infants vII. Unpublished results.

Fort, M., Weiss, A., Martin, A., & Peperkamp, S., (2013). *Looking for the bouba-kiki effect in prelexical infants*. Paper presented at AudioVisual Speech Processing, AVSP 2013, Interspeech, Annecy, France.

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*, 107–128.

Hedges, L.V., Tipton, E., & Johnson, M.C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 20130298.

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

Imai, M., Miyazaki, M., Yeung, H.H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS ONE*, *10*, e0116494.

Jadva, V., Hines, M., & Golombok, S. (2010). Infants' preferences for toys, colors, and shapes: Sex differences and similarities. *Archives of Sexual Behavior*, *39*, 1261–1273.

Jones, M., Vinson, D., Clostre, N., Zhu, A. L., Santiago, J. & Vigliocco, G. (2014). The bouba effect: Sound-shape iconicity in iterated and implicit learning. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, (pp. 1144–1119. Austin, TX: Cognitive Science Society.

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, *35*, 575–586.

Kellman, P.J., & Arterberry, M.E. (2007). Infant visual perception. In D. Kuhn & R.S. Siegler (Eds.), *Handbook of child psychology* (Vol. *2*, pp. 109–160). Hoboken, NJ: Wiley.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright.

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

Lammertink, I., Tsuji, S., & Fikkert, P. (2015). *Sensitivity to the bouba-kiki effect arises late in Dutch toddlers*. Paper presented at the biennial meeting of the Society for Research in Child Development (SRCD 2015). Philadelphia, USA.

Lany, J., & Saffran, J.R. (2013). Statistical learning mechanisms in infancy. In J.L.R. Rubenstein & P. Rakic (Eds.), *Comprehensive developmental neuroscience: Neural circuit development and function in the heathy and diseased brain* (Vol. *3*, pp. 231–248). Amsterdam: Elsevier.

Lewkowicz, D.J., & Ghazanfar, A.A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, *13*, 470–478.

Lewkowicz, D.J., & Minar, N.J. (2014). Infants are not sensitive to synesthetic cross-modality correspondences: A comment on Walker et al. (2010). *Psychological Science*, *25*, 832–834.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, *6*, 1246.

Maurer, D., Pathman, T., & Mondloch, C.J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*, 316–322.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & the PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*, e1000097.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *38*, 1152–1164.

Monaghan, P., Shillcock, R.C., Christiansen, M.H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 20130299.

Nygaard, L.C., Cook, A.E., & Namy, L.L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*, 181–186.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*, 173–186.

Palumbo, L., Ruta, N., & Bertamini, M. (2015). Comparing angular and curved shapes in terms of implicit associations and approach/avoidance responses. *PLoS ONE*, *10*, e0140043.

Pejovic, J., & Molnar, M. (2017). The development of spontaneous sound-shape matching in monolingual and bilingual infants during the first year. *Developmental Psychology*, *53*, 581–586.

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, *22*, 1419–1421.

Perry, L.K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS One*, *10*, e0137147.

Perry, L.K., Perlman, M., Winter, B., Massaro, D.W., & Lupyan, G. (2017). Iconicity in the speech of children and adults. *Developmental Science*, e12572. https://doi.org/10.1111/desc.12572

Pons, F., Lewkowicz, D.J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 10598–10602.

Quinn, P., Brown, C., & Streppa, M. (1997). Perceptual organization of complex visual configurations by young infants. *Infant Behavior and Development*, *20*, 45–46.

Ramachandran, S., & Hubbard, E.M. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies*, *8*, 3–34.

R Core Team. (2016). *R: A language and environment for statistical computing*. http://www.r-project.org

Spector, F. (2009). *The colours and shapes of the world: Testing predictions from synesthesia about the development of sensory associations*. Dissertation, McMaster University, Hamilton (Ontario), Canada.

Spector, F., & Maurer, D. (2009). Synesthesia: A new approach to understanding the development of perception. *Developmental Psychology*, *45*, 175–189.

Spector, F., & Maurer, D. (2013). Early sound symbolism for vowel sounds. *I-Perception*, *4*, 239–241.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, and Psychophysics*, *73*, 971–995.

Starr, A., & Brannon, E.M. (2012). *Sound-shape congruency in preverbal infants*. Poster presented at the biennial meeting of the International Society for Infant Studies. Minneapolis, MN.

Téglás, E., & Bonatti, L.L. (2016). Infants anticipate probabilistic but not deterministic outcomes. *Cognition*, *157*, 227–236.

Trainor, L.J., Clark, E.D., Huntley, A., & Adams, B.A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development*, *20*, 383–396.

Tsuji, S., Bergmann, C., & Cristia, A. (2014). Community-augmented meta-analyses toward cumulative data assessment. *Perspectives on Psychological Science*, *9*, 661–665.

Tzeng, C.Y., Nygaard, L.C., & Namy, L.L. (2017). Developmental change in children's sensitivity to sound symbolism. *Journal of Experimental Child Psychology*, *160*, 107–118.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. http://www.jstatsoft.org/v36/i03

Walker, P. (2016). Cross-sensory correspondences and symbolism in spoken and written language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000253

Walker, P., Bremner, J.G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S.P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*, 21–25.

Walker, P., Bremner, J.G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S.P. (2014). Preverbal infants are sensitive to cross-sensory correspondences: Much ado about the null results of Lewkowicz and Minar (2014). *Psychological Science*, *25*, 835–836.

Weikum, W.M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J.F. (2007). Visual language discrimination in infancy. *Science*, *316*, 1159.

Werker, J.F., & Tees, R.C. (1984). Cross-language speech-perception: Evidence for perceptual reorganization during the 1st year of life. *Infant Behavior and Development*, *7*, 49–63.