A Meta-Analysis of Infants' Mispronunciation Sensitivity Development

Katie Von Holzen
(katie.m.vonholzen@gmail.com)
Department of Hearing and Speech Science, University of Maryland
7251 Preinkert Dr., College Park, MD 20742 USA

Christina Bergmann
(christina.bergmann@mpi.nl)
Language Development Department, Max Planck Institute for Psycholinguistics
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

Abstract

Before infants become mature speakers of their native language, they must acquire a robust word-recognition system which allows them to strike the balance between allowing some variation (mood, voice, accent) and recognizing variability that potentially changes meaning (e.g. cat vs hat). The current meta-analysis quantifies how the latter, termed mispronunciation sensitivity, changes over infants' first three years, testing competing predictions of mainstream language acquisition theories. Our results show that infants were sensitive to mispronunciations, but accepted them as labels for target objects. Interestingly, and in contrast to predictions of mainstream theories, mispronunciation sensitivity was not modulated by infant age, suggesting that a sufficiently flexible understanding of native language phonology is in place at a young age.

Keywords: language acquisition; mispronunciation sensitivity; word recognition; meta-analysis; lexicon

Introduction

In everyday language processing, we usually do not notice that we make thousands of judgements as to which variation is acceptable (e.g. mood, accented speech), and when variation may actually change the meaning of a word (for example cat vs hat in English). How and at what age does a language learner build such skills? This question has been addressed in a productive line of research studying what is commonly called "mispronunciation sensitivity," which is the sensitivity to a small, but potentially meaning-altering change in the acoustic word form. If infants are sensitive to mispronunciations it would mean that they have an understanding of the sound-level (i.e. phonological) information that distinguishes words in their native language. This knowledge varies cross-linguistically and therefore has to be acquired, a process that commences in the first year (Kuhl, 2004) and continues through toddlerhood. Examining mispronunciation sensitivity, therefore, probes whether infants have learned an important part of their native language and offers unique insight into the developing (and mature) lexicon.

The first evidence of an emerging mispronunciation sensitivity in infants came from Swingley and Aslin (2000). Using the Preferential Looking Procedure (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987), the target-looking behavior of 18-23-month-old American English-learning children was examined when they heard the correct label for a familiar object (e.g. "baby" when seeing an image of a young child next to a distractor, such as a dog) or when the label was mispronounced (e.g. "vaby"). Looks to the target were significantly greater in correct compared to mispronounced trials. This initial finding of a mispronunciation sensitivity has launched almost two decades of research examining a wide age range of infants learning many different languages on their sensitivity to different kinds of mispronunciations.

Considering that infants are sensitive to mispronunciations and that, in general, their processing matures with development, we examine the shape of mispronunciation sensitivity from six to 30 months. On the basis of theoretical accounts and empirical studies, we explore three possible trajectories for the development of mispronunciation sensitivity. We first review the concrete predictions by current competing theories and then describe how we test those predictions with a meta-analytic approach.

The Perceptual Attunement account describes a shift from specific native sound patterns to a more mature understanding of the abstract phonological (i.e. sound category) structure of words (Best, 1994, 1995). This shift is predicted to coincide with the vocabulary spurt around 18 months, and is therefore related to vocabulary growth. In this case, we would expect the size of mispronunciation sensitivity to *decrease* with development, with young infants initially rejecting any phonological variation in familiar words and only later learning to accept some variability. This prediction was empirically borne out in a study showing that between 18 and 24 months sensitivity to small mispronunciations decreased (Mani & Plunkett, 2011).

PRIMIR (Processing Rich Information from Multidimensional Interactive Representations; Curtin & Werker, 2007; Werker & Curtin, 2005), in contrast, describes vocabulary growth as promoting more detailed representations of words due to a re-analysis of learned generalistic word forms into their constituent sound patterns. In this case, we would expect mispronunciation sensitivity to *increase* with development as infants mature and add more words to their growing lexicon. An increase in mispronunciation sensitivity has been shown in a number of studies comparing two or three different ages (Altvater-Mackensen & Mani, 2013; Mani & Plunkett, 2007; van der Feest & Fikkert, 2015).

Finally, another possibility is that mispronunciation sensitivity is not modulated by development. Infants' general language processing skills increase, but this may not translate to an increase or decrease in sensitivity to mispronunciations. Instead, mispronunciations would be detected, but the size of this effect would not be related to developmental change. A handful of studies testing multiple ages have found no difference in mispronunciation sensitivity as infants mature and build their lexicon (Bailey & Plunkett, 2002; Zesiger, Lozeron, Levy, & Frauenfelder, 2012). Yet, this pattern of results has not been incorporated in to a mainstream theory of language acquisition; for completeness, we mention it here.

We disentangle these three possible patterns of infants developing a mispronunciation sensitivity using meta-analysis. This approach allows us to aggregate all available evidence, because single studies lend support to all three possible patterns. Theories predicting change over development tie this change with vocabulary growth; in the current meta-analysis we do not analyze infant vocabulary for practical reasons, as vocabulary size of the participant group is rarely reported (seven studies in our sample) and there is no test yielding comparable scores across different languages at the age range we cover (six to 30 months). Instead, we examine general developmental change associated with age, which coincides with increases in lexicon size (Frank, Braginsky, Yurovsky, & Marchman, 2017). Since we examine effects aggregated over groups and languages, age might in fact be the best proxy for vocabulary in lieu of feasible comparable measures (see Bergmann & Cristia, 2016; for a similar line of reasoning).

In order to quantify the size of infants' response to correct pronunciations and mispronunciations across diverse studies using different designs and a variety of dependent measures, we use standardized effect sizes (see next section for details). We examine both correct pronunciations and mispronunciations individually (object identification) as well as the difference between the two (mispronunciation sensitivity). For correct pronunciations, a larger effect size for object identification reflects a stronger recognition response (across the participant group), while for mispronunciations a larger effect size for object identification reflects a stronger acceptance of that label as appropriate for the target image. A larger effect size for mispronunciation sensitivity, in turn, reflects a larger difference in infants' looking behaviors when hearing correct pronunciations and mispronunciations; the larger the mispronunciation sensitivity, the more robust infants' knowledge that a mispronunciation is not a good label for a known target object. We further examine the relationship between these effect sizes and infant age to explore the developmental trajectory of mispronunciation sensitivity. We aggregate over a large number of studies which allows for a dense sampling over a wide age range of six to 30 months. As a result, we are able to capture developmental change through a wider lens.

Methods

Systematic Study Collection

We first generated a list of potentially relevant items (38 contributed by the authors, 63 by experts in the field). This was supplemented with a systematic google scholar search of papers citing Swingley & Aslin (2000), which yielded 400 results. After removing duplicates, we screened the title and abstract of all items and included them following these criteria: An item had to report (1) original data; (2) on familiar word recognition after correct pronunciations and mispronunciation; (3) of infants younger than 31 months; (4) in an eye movement experiment. Three items had to be excluded because they did not report sufficient data to compute effect sizes. The final sample consisted of 27 journal articles, two dissertations, two unpublished reports, and one proceedings paper. We will refer to these 32 items collectively as papers.

Effect Size Calculation

All scripts and raw data are available on Open Science Framework (OSF)[1]. The dependent variable in a typical mispronunciation study compares infants' looks to the target picture upon naming against some baseline, which can be chance or looks in a pre-naming window. Proportion of target looks (PTL)[2] is calculated as the percentage of looks to the target divided by the total percentage of looks to the target and distractor. We used the baseline as reported by the original authors of a paper. The majority of papers (n = 13) subtracted pre-naming PTL from post- naming PTL to achieve a difference score representing the change in PTL once the target word (correct or mispronounced) was named. This difference score is then compared to zero (no change in looks to the target). The remaining papers compared post-naming PTL with pre-naming PTL directly (n = 10) or compared post-naming PTL with chance (50%; n = 9). In all cases, positive values indicate more looks to the target than expected by chance, thereby reflecting some form of object recognition after hearing a label. We calculated effect sizes for infants' looking to the target pictures separated by whether or not words were mispronounced (object identification). The difference in effect sizes between correct and mispronounced trials was used to estimate infants' mispronunciation sensitivity. We report Hedges' g, which corrects for small sample sizes which are common in infant research (Hedges, 1981; Morris, 2000).

[1] https://osf.io/nvc8m/

[2] Two papers reported longest look (LLK) instead of PTL. Although PTL is now almost exclusively used, LLK was originally thought to be a more sensitive measure of infant comprehension (see Mani & Plunkett, 2007). Both papers compared post-naming LLK with pre-naming LLK, with greater LLK in the post-naming indicating an association or recognition of the label for the target image. When PTL and LLK were reported in the same paper, we only analyzed PTL.

We calculated effect sizes based on reported raw data or, when those were not available, test statistics in the original paper; 25 papers reported raw means and standard deviations, seven papers reported t-values.[3] The formulae we used are standard for effect size calculation in within-participant designs (means and standard deviations: Lipsey & Wilson, 2001; t-values: Dunlap, Cortina, Vaslow, & Burke, 1996).

When two means (i.e. looking during a baseline period and post naming) are reported in a within-participant design, correlations between participant-level results of the two are necessary for calculating effect sizes based on t-values and to obtain effect size variance; only one paper reported this information. For four of the remaining papers, we used means, standard deviations, and t-values to compute correlations (following Csibra, et al. 2016, Appendix B; see also Rabagliati, Ferguson, & Lew-Williams, submitted). For the remaining papers, correlations were imputed (see Black & Bergmann, 2017, for the same procedure). In total, we could compute 104 effect sizes for correct pronunciations and 147 for mispronunciations.

Publication Bias

It is possible that our data contain bias due to a general tendency to value and publish significant over non-significant results (see Ferguson & Heene, 2012). To ensure that our conclusions are not based on heavily biased data, we conduct two tests. First, we use the rank correlation test of funnel plot asymmetry to assess whether effect sizes are distributed as would be expected based on sampling error, that is more precise, low-variance effect sizes are closer to the estimated mean and high-variance effect sizes show increased, evenly-distributed spread around the mean effect.

The second analysis is a p-curve of all significant results, which tests for evidential value (whether the p-values are distributed in a way to be expected when the null hypothesis is true or not) and whether there is a larger proportion of p-values just below the typical alpha threshold of .05 (indicative of questionable research practices). For this analyses, we rely on the p-curve app (v4.0, p-curve.com; Simonsohn, Simmons, & Nelson, 2014). Hearing a correct compared to a mispronounced label is expected to lead to different looking behaviors; therefore, we conduct the two analyses separately for both conditions.

Meta-Analysis

We report hierarchical random-effects models (infant groups nested within papers) of variance-weighted effect sizes with the R (R Core Team, 2016) package metafor (Viechtbauer, 2010). To investigate the impact of development, we intro- duce age (centered; transformed into months for readability by dividing days by 30.44) as a moderator.

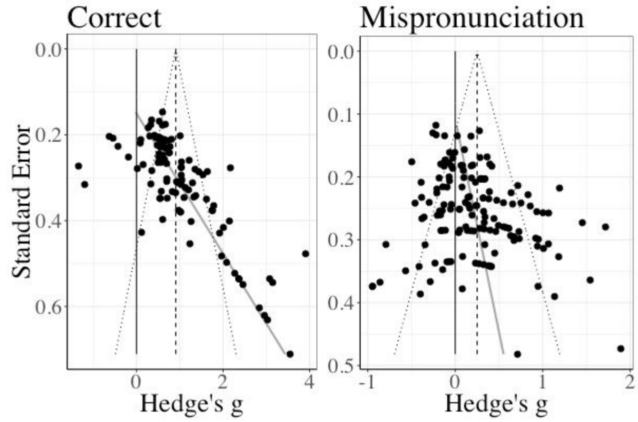

Figure 1: Funnel plots for object identification, plotting the standard error of the effect size in relation to effect size. The black line marks zero, the dashed grey line marks the effect estimate, and the grey line marks funnel plot asymmetry.

Results

Publication Bias

We first analyzed funnel plot asymmetry, which was significant for both correctly pronounced words and mispronuncations, (correct pronunciation: Kendall's $\tau = .53$, $p < .0001$; mispronunciation: Kendall's $\tau = .16$, $p = .004$). These results indicate that there is some bias in the literature, and Figure 1 (based on code adapted from Sakaluk, 2016) underlines this impression. Particularly for correctly pronounced words, we see large effect sizes paired with greater variance (bottom right corner) and more precise effect sizes (i.e. with smaller variance) being smaller than expected (top left, outside the triangle).

A p-curve based on 72 statistically significant values for correct pronunciations indicates that the data contain evidential value ($Z = -17.93$, $p < .0001$) and there is no excess of "just significant" p-values. For mispronunciation, a p-curve based on 36 statistically significant values indicates that the data contain evidential value ($Z = -6.18$, $p < .0001$), and we again find no excess of p-values near .05.

These results suggest a tendency towards publication bias, which might lead to us systematically over-estimating effects. This is concerning, and we interpret all effect size estimate consequently with caution. A second cause for funnel plot asymmetry might be heterogeneity in the data and/or because some studies examined more subtle effects than others did. At the same time, the p-curves suggest that the literature overall contains evidential value (i.e. a "real" effect) and we therefore continue our meta-analysis.

Meta-analysis

Object Identification for Correct and Mispronounced Words For correctly pronounced words the

[3] We extracted means and standard deviations from figures for four papers.

variance-weighted meta-analytic effect size Hedges' *g* was

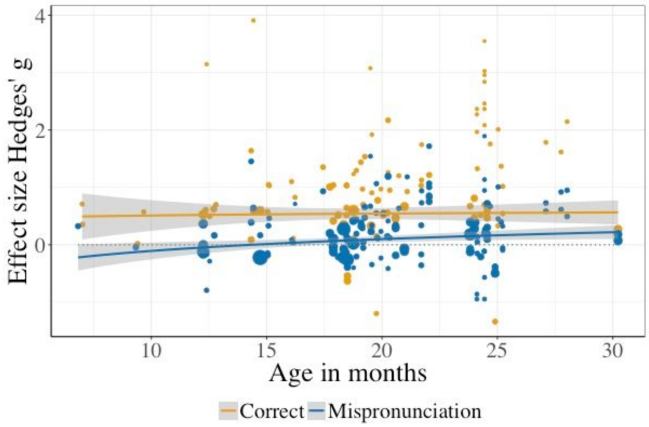

Figure 2: Effect sizes for correct pronunciations (yellow) and mispronunciations (blue) by participant age. Point size depicts inverse variance. The dashed line indicates zero.

0.91 (SE = 0.12, 95% CI [0.63, 1.14], $p < .0001$). This is a large effect size (according to the criteria set by Cohen, 1988; see also Bergmann, et al., 2018; for comparative meta-analytic effect sizes in language acquisition research). The effect is estimated to be significantly above zero, which suggests that when presented with the correctly pronounced label, infants reliably and robustly fixated the corresponding object. We remind the reader, however, that we found evidence for publication bias and this might be an overestimation. Yet, based on the p-curve results and the CI lower bound being at 0.63, we expect this result to be robust even when correcting for publication bias.

For mispronounced words, Hedges' *g* was 0.25 (SE = 0.06, 95% CI [0.13, 0.37], $p < .0001$). This is considered a small effect size (Cohen, 1988), but significantly above zero, which suggests that even when presented with a mispronounced label, infants still fixated the target object. Again, we note the publication bias (which was smaller in this condition), and point out that this might be an overestimation. But since the p-curve indicated evidential value, we are confident in this result as well.

Heterogeneity was significant for both correct (Q(103) = 425.63, $p < .0001$) and mispronounced words, (Q(146) = 426.51, $p < .0001$). This indicated that the sample contains unexplained variance leading to significant difference across the included studies beyond what is to be expected based on random chance.

Mispronunciation Sensitivity The above two analyses considered conditions with mispronounced and correctly pronounced words separately. To evaluate mispronunciation sensitivity, we then compared the effect size Hedges' *g* for correct pronunciations with mispronunciations, merging the two datasets and introducing condition as moderator. The moderator test was significant, QM(1) = 215.76, $p < .0001$. The estimate for the difference, in other words the effect size of infants' mispronunciation sensitivity, was 0.5 (SE = 0.03), which indicated that the two types of conditions elicited responses that significantly differed (95% CI [0.43, 0.56], $p < .0001$). This confirms that although infants fixate the target object when hearing both correct pronunciations and mispronunciations, the observed fixations to target (expressed in effect sizes) are more systematic for correct pronunciations. We thus can now quantify the modulation of fixation behavior in terms of standardized effect sizes.

Age Effects Figure 2 shows all effect sizes, with color encoding whether the target was pronounced with a correct or mispronounced label. To evaluate the different predictions for how mispronunciation sensitivity will change as infants develop, we next added age (centered, in months), as a moderator. First, we investigate the impact of age separately on conditions where words were either pronounced correctly or not. Age did not significantly modulate object identification in response to correctly pronounced (QM(1) = 0.68, $p = .41$) nor mispronounced words (QM(1) = 1.72, $p = .19$). As age increased, there was no evidence that target looks in response to a correctly pronounced or mispronounced label also increased.

We then examined the interaction between age and mispronunciation sensitivity (correct vs. mispronounced words) in our whole dataset. The moderator test was significant (QM(3) = 218.62, $p < .0001$). The interaction between age and mispronunciation sensitivity, however, was not significant ($\beta = 0.002$, SE = 0.008, 95% CI [-0.01, 0.02]), suggesting that infants' mispronunciation sensitivity remains constant across development and is not driven by specific age groups (see also Figure 2).

Discussion

Using a meta-analytic approach, we examined the developmental trajectory of infants' mispronunciation sensitivity. Previous literature and theoretical accounts predicted three different developmental patterns: mispronunciation sensitivity decreases, increases, or does not change with infant age.

Before investigating which of the three possibilities is supported by all aggregated experimental data, we first examined object identification separately for correct and mispronounced labels, which revealed a significant and reliable effect for both pronunciations across all ages. Despite the altered phonological form, a mispronounced label was still considered a better match with the target image than a distractor image.

Next, we quantified infants' mispronunciation sensitivity by comparing correct and mispronounced words with one another. Indeed, there was a sizable, significant difference between conditions: Correct words had a significantly larger meta-analytic effect size than mispronounced words, indicating a sensitivity to mispronunciations.

Finally, we assessed the three possible patterns of development modulating infants' mispronunciation sensitivity, which we established in the preceding analyses. When age was considered as a moderator, we observed no significant main effects of age or interactions. Surprisingly, this result was not predicted by any mainstream theory of language acquisition that we are aware of; both the Perceptual Attunement account (Best 1994, 1995) and PRIMIR (Curtin & Werker, 2007; Werker & Curtin, 2005) predict a *change* of mispronunciation sensitivity with development (brought about by vocabulary growth). In the current meta-analysis, however, few papers report estimates of infants' vocabulary (n = 7). We can however observe that overall and in a large enough sample, vocabulary will increase dramatically in the age range covered here (see wordbank.stanford.edu; Frank et al., 2017), and since our analysis concerns groups of infants, we are confident that this pattern holds over the age range covered here. Further support our conjecture that there is no relation between vocabulary size and mispronunciation sensitivity comes from some of the previous empirical studies examining vocabulary size (e.g. Mani & Plunkett, 2007; Swingley & Aslin, 2000; but see Mani & Plunkett, 2010).

Acceptance of a mispronunciation (object identification) while maintaining that mispronunciation and correct pronunciation are distinct (mispronunciation sensitivity) could indicate a somewhat abstract understanding of the phonological structure of words. Although infants demonstrate an understanding that the phonological form of a mispronunciation is not an exact match with the correct pronunciation, they still accept this label as more appropriate for the target compared with the distractor image. The lack of age effects suggest that this understanding coincides with the learning of the earliest words and persists throughout early lexical development. Indeed, the youngest ages in our database were below one year (Bergelson & Swingley, 2017; Mani & Plunkett, 2007; Zesiger et al., 2012) and even at that age sensitivity to mispronunciations is in place according to our meta-analytic regression models (see also Figure 2). Upon learning their first words, it would appear that infants already have a sense of the information that builds words in their native language. Updated theoretical accounts should take this lack of a developmental change into account.

Although this meta-analysis suggests that overall there is no developmental change in mispronunciation sensitivity, other factors, which pertain to experimental sample, design, and analysis, may play a role. For example, our meta-analysis included studies that varied in the native language of infants tested, type of mispronunciation (e.g. consonant or vowel), and time window of analysis. Concretely, most studies used (presumably) known objects as distractors (n = 22), but roughly a third of papers (n = 10) presented novel objects, which could be a possible referent of the mispronounced word form, creating a minimal pair (e.g. cat-hat). Could it be that infants fixate the correct object upon hearing a mispronounced label, because the distractor is equally well-known? Although a preliminary moderator test was significant (QM(3) = 219.46, $p < .0001$), the interaction between distractor familiarity (known or novel) and mispronunciation sensitivity was not(β = 0.14, SE = 0.08, 95% CI [-0.02, 0.30]). Future analyses of our dataset will further elucidate how experimental planning and analytic decisions might mask or amplify any effects or changes in infants' behavior we may expect to see.

Statistical Power

The median sample size across studies in our dataset was 26. Assuming that the main goal of researchers was to study infants' mispronunciation sensitivity, which we estimate to have an effect size of 0.5, 26 participants would be required in a within-participant design using a one-sided t-test to conduct a study which has 80% power (i.e. 20% of studies show a false null result implying there is no difference between conditions). This suggests that the studies in the current literature are well powered when investigating mis pronunciation sensitivity per se. While this is encouraging, the required sample size for studies on factors modulating mispronunciation sensitivity (e.g. the type of the mispronunciation) would have to be much larger. We thus reiterate our remark that future investigations have to quantify if and how differences in design affect infants' mispronunciation sensitivity.

Conclusion

Our meta-analysis revealed that although infants fixate the target image for both correct and mispronounced labels, they are more likely to do so for correct labels, indicating sensitivity to mispronunciations. Interestingly, this pattern was not modulated by infants' age. We suggest that already at a young age and with just a small lexicon, infants use their understanding of their native language phonological structure during word recognition. These results have important implications for existing theories of early phono-lexical developments, which predict developmental changes in mispronunciation sensitivity that were not borne out in this meta-analysis.

Acknowledgments

Katie Von Holzen was supported by the Agence Nationale de la Recherche (ANR-13-BSH2-0004) and by training grant DC-00046 from the National Institute of Deafness and Communicative Disorders of the National Institutes of Health. Christina Bergmann was supported by the European Horizon 2020 programme (Marie Skłodowska-Curie grant No 660911), the Agence Nationale de la Recherche (ANR-10-IDEX-0001-02 PSL*, ANR-10-LABX-0087 IEC) and the Fondation de France.

References

Altvater-Mackensen, N., & Mani, N. (2013). The impact of mispronunciations on toddler word recognition: Evidence for cascaded activation of semantically related words from mispronunciations of familiar words. *Infancy*, *18*(6), 1030–1052.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, *17*(2), 1265–1282.

Bergelson, E., & Swingley, D. (2017). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*.

Bergmann, C., & Cristia, A. (2016). Development of infants' segmentation of words from native speech: A meta-analytic approach. *Developmental Science*, *19*(6), 901–917.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M. B., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*.

Best, C. T. (1994). *The emergence of native-language phonological influences in infants: A perceptual assimilation model*. *Haskins Laboratories Status Report on Speech Research* (Vol. 107/108).

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*.

Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A meta-analysis. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, Inc.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521–36.

Curtin, S., & Werker, J. F. (2007). The perceptual foundations of phonological development. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics*. New York: Oxford University Press.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*(2), 170–177.

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*(6), 555–561.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(1), 23–45.

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*(2), 107–128.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews. Neuroscience*, *5*(11), 831–43.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272.

Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy*, *15*(5), 445–470.

Mani, N., & Plunkett, K. (2011). Does size matter? Subsegmental cues to vowel mispronunciation detection. *Journal of Child Language*, *38*(3), 606–627.

Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical and Statistical Psychology*, *53*, 17–29.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rabagliati, H., Ferguson, B. F., & Lew-Williams, C. (submitted). The profile of abstract rule learning in infancy. Evidence from meta-analysis and a cross-lab experiment.

Sakaluk, J. (2016). 7. Make it pretty: Forest and funnel plots for meta-analysis using ggplot2. [Blog post].

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166.

van der Feest, S. V. H., & Fikkert, P. (2015). Building phonological lexical representations. *Phonology*, *32*(2), 207–239.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234.

Zesiger, P., Lozeron, E. D., Levy, A., & Frauenfelder, U. H. (2012). Phonological specificity in 12- and 17-month-old French-speaking infants. *Infancy*, *17*(6), 591–609.