MP MetaAnalysis

Christina Bergmann

Contents

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats
## lag():    dplyr, stats

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':
##
##     expand

## Loading 'metafor' package (version 1.9-9). For an overview
## and introduction to the package please type: help(metafor).

## Loading 'meta' package (version 4.9-0).
## Type 'help(meta)' for a brief overview.

##
## Attaching package: 'meta'

## The following objects are masked from 'package:metafor':
##
##     baujat, forest, funnel, funnel.default, labbe, radial,
##     trimfill

##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

Preparation

Read in data and tidy up dataset

Plotting defaults

Methods

The present meta-analysis was conducted with maximal transparency and reproducibility in mind. To this end, we provide all data and analysis scripts on the supplementary website (https://osf.io/rvbjs/) and open our meta-analysis up for updates (Tsuji, Bergmann, & Cristia, 2014). In addition, we follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and make the corresponding information available as supplementary materials. (Moher, Liberati, Tetzlaff, Altman & PRISMAGroup, 2009).

Study Selection

We first generated a list of potentially relevant items to be included in our meta-analysis by creating an expert list. This process yielded 110 items. We then used the google scholar search engine to search for papers citing the original Swingley & Aslin (2000) publication. This search was conducted on 22 September, 2017 and yielded 288 results. We screened the 398 items, removing 99 duplicate items. We screened remaining 299 items for their title and abstract to determine whether it met the following inclusion criteria: (1) original data was reported; (2) the experiment examined familiar word recognition; (3) infants studied were under 36-months-of-age; (4) the dependent variable was derived from proportion of looks to a target image versus a distractor in a eye movement experiment; 5) the stimuli were auditory speech. The final sample (n = *32*) consisted of 27 journal articles, 1 proceedings paper, 2 thesis, and 2 unpublished reports. We will refer to these items collectively as papers. Table 1 (Summary Table) provides an overview of all papers included in the present meta-analysis.

publication_status	n_unique	count
dissertation	2	17
gray paper	2	14
paper	27	216
proceedings	1	4
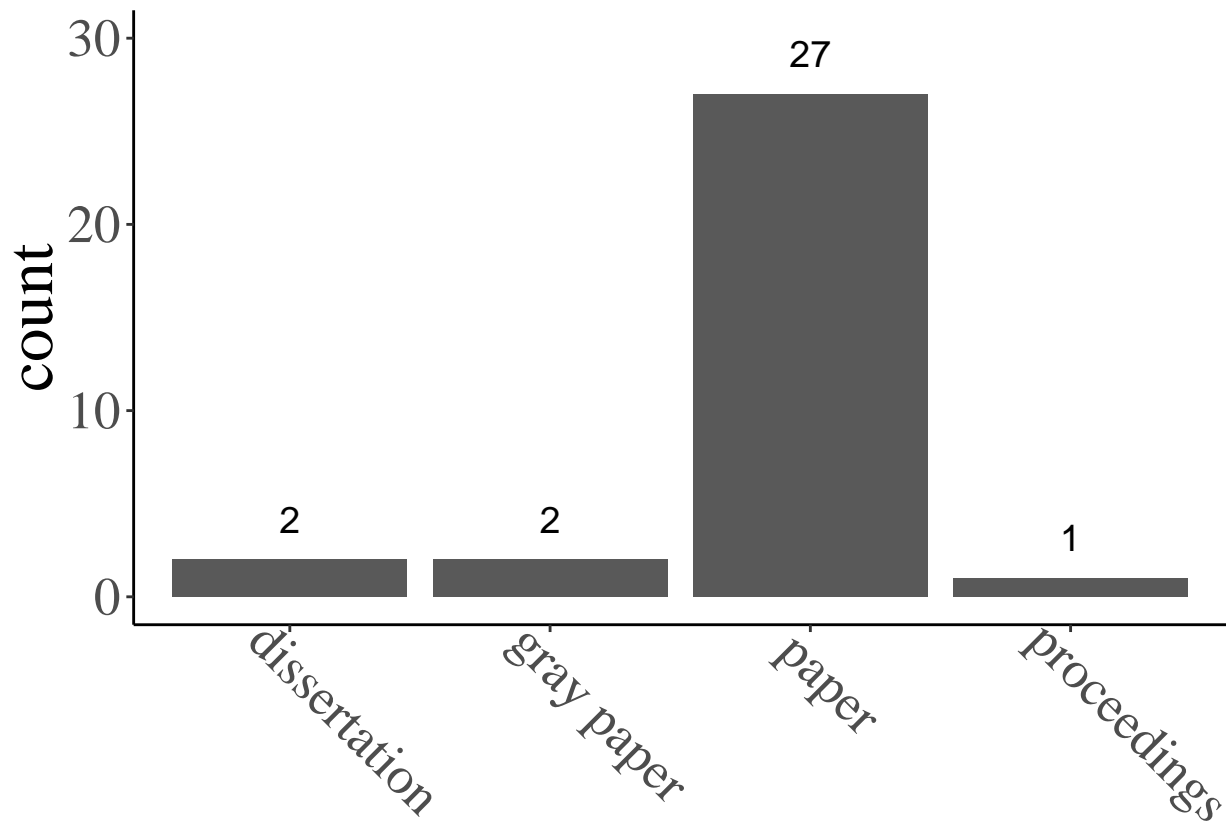
```
## pdf
##   2
```

Data Entry

The 32 papers we identified as relevant were then coded with as much detail as possible (Tsuji, Bergmann, & Cristia, 2014; Bergmann et al., in press). For each experiment (note that a paper typically has multiple experiments), we entered variables describing the publication, population, experiment design and stimuli, and results. For the present analyses, we focus on the following characteristics: 1 Condition: Were words mispronounced or not; 2 Mean age reported per group of infants, in days; 3 Vocabulary size:
4 Size of mispronunciation, measured in features changed; 5 Distractor familiarity: familiar, unfamiliar; 6 Phonological overlap between target and distractor: onset, onset/medial, rhyme, none, novel word; 7 Position of mispronunciation: onset, medial, offset; 8 Type of mispronunciation: consonant, vowel;

We separated out conditions according to whether or not the target word was mispronounced to be able to investigate both infants' looking to the target picture separated by whether or not words were mispronounced and their mispronunciation sensitivity, which is the difference between looking in correct and mispronounced trials. When the same infants were further exposed to multiple mispronunciation conditions and the results were reported separately in the paper, we also entered each condition as a separate row (e.g., consonant versus vowel mispronunciations; Mani & Plunkett, 2007). The fact that the same infants have contributed data to multiple rows (minimally those containing information on correct and mispronounced trials) leads to shared variance across effect sizes, which we account for in our analyses (see next section). We will call each row a record; in total there were 251 records in our data.

Data analysis

within_measure_descriptive	n_unique	count
post-naming compared to pre-naming phase	10	29
post-naming phase compared with chance (=50%)	9	23
post-pre difference score compared with chance (=0)	13	52

es_method	n_unique	count
group_means_one	18	120
group_means_two	7	57
t_one	4	39
t_two	5	35

All scripts and raw data are available on Open Science Framework (OSF). INSERT LINK AS FOOTNOTE HERE. Mispronunciation sensitivity studies typically examine infants' proportion of target looks (PTL) in comparison to a baseline measurement. PTL is calculated by dividing the percentage of looks to the target by the total percentage of looks to both the target and distractor images. Across papers the baseline comparison varied; we used the baseline reported by the authors of each paper. Most papers ($n = 13$) subtracted the PTL score for the pre-naming phase from the PTL score for the post-naming phase. When interpreting this difference score, a positive value indicates that infants increased their looks to the target after hearing the naming label (correct or mispronounced). Other papers either compared post- and pre-naming PTL with one another ($n = 10$) or compared post-naming PTL with chance (50%, ($n = 9$)). For these comparisons, a positive difference score or a post-naming phase PTL score that is greater than the pre-naming phase PTL or chance indicate target looks that indicate object recognition after hearing the naming label.

We report effect sizes for infants' looks to target pictures after hearing a correctly pronounced or mispronounced label (object identification) as well as the difference between effect sizes for correct and mispronounced trials (mispronunciation sensitivity). The effect size we report in the present paper are based on comparison of means, standardized by their variance. The most well-known effect size from this group is Cohen's d (Cite cohen). To correct for the small sample sizes common in infant research, however, we use as a dependent variable Hedges' g instead of Cohen's d (Hedges, 1981; Morris, 2000).

No authors made their raw data available. Instead, we calculated Hedges' g using the raw means and standard deviations reported in the paper ($n = 25$) or using reported t-values ($n = 9$). Raw means and standard deviations were extracted from figures for 3 papers. In a within-participation design, when two means are compared (i.e. looking during pre- and post-naming) it is necessary to obtain correlations between the two measurements at the participant level to calculate effect sizes and effect size variance based on t-values. One paper reported this correlation, while we were able to compute correlations using means, standard deviations, and t-values (following Csibra, et al. 2016, Appendix B; see also Rabagliati, Ferguson, & Lew-Williams, submitted). Correlations were imputed for the remaining papers (see Black & Bergmann, 2017, for the same procedure). We computed a total of 104 effect sizes for correct pronunciations and 147 for mispronunciations.

To take into account the fact that the same infants contributed to multiple datapoints, we analyze our results in a multilevel approach using the R (R citation) package metafor (Viechtbauer, 20XX, Multulevel reference). This means we model as random effect that effect sizes from the same paper share are based on more similar studies than those across papers and that nested therein effects can stem from the same infants.

Publication Bias

In the psychological sciences, there is a documented reluctance to publish null results. As a result, there is a potential for significant results to be valued over non-significant results (see Ferguson & Heene, 2012). To examine whether this is also the case in the mispronunciation sensitivity literature, which would bias the data analyzed in this meta-analysis, we conduct two tests. We first examine whether effect sizes are

distributed as expected based on sampling error using the rank correlation test of funnel plot asymmetry with the R (R Core Team, 2016) package metafor (Viechtbauer, 2010). Effect szes with low-variance are closer to the estimated mean, while effect sizes with high-variance show an increased, evenly-distributed spread around the estimated mean. Second, we analyze all of the significant results in the dataset using a p-curve from the p-curve app (v4.0, p-curve.com; Simonsohn, Simmons, & Nelson, 2014). This tests for evidential value by examining whether the p-values have an expected distribution, regardless of whether the null hypothesis is true or not, as well as whether there is a larger proportion of p-values just below the typical alpha threshold of .05, which may indicate questionable research practices. Responses to correctly pronounced and mispronounced labels are predicted to show different patterns of looking behavior; as a result, we conduct these two analyses to assess publication bias separately for both conditions.

Meta-analysis

The models reported are 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 how development impacts mispronunciation sensitivity, our core theoretical investigation, we introduce age (centered; in days but transformed into months for ease of reading by dividing by 30.44) as a moderator to our main model. For the subsequent investigations of experimental characteristics, we introduce each characteristic as a moderator (more detail below).

Results

Publication Bias

Figure 1 plots the funnel plots for object identification for both correct pronunciations and mispronunciations (code adapted from Sakaluk, 2016). Funnel plot assymmetry was significant for both correct pronunciations (Kendall's $\tau = 0.53$, $p < .0001$) and mispronunciations (Kendall's $\tau = 0.16$, $p < .01$). These results, in conjunction with the assymmetric funnel plot (Figure 1) indicate bias in the literature. This is particularly evident for correct pronunciations, where larger effect sizes have greater variance (bottom right corner) and there are a smaller number of more precise effect sizes (i.e. smaller variance) than expected (top left, outside the triangle).

Figure 1

```
## pdf
##   2
```

Correct

Mispronunciation

Standard Error — Hedge's g
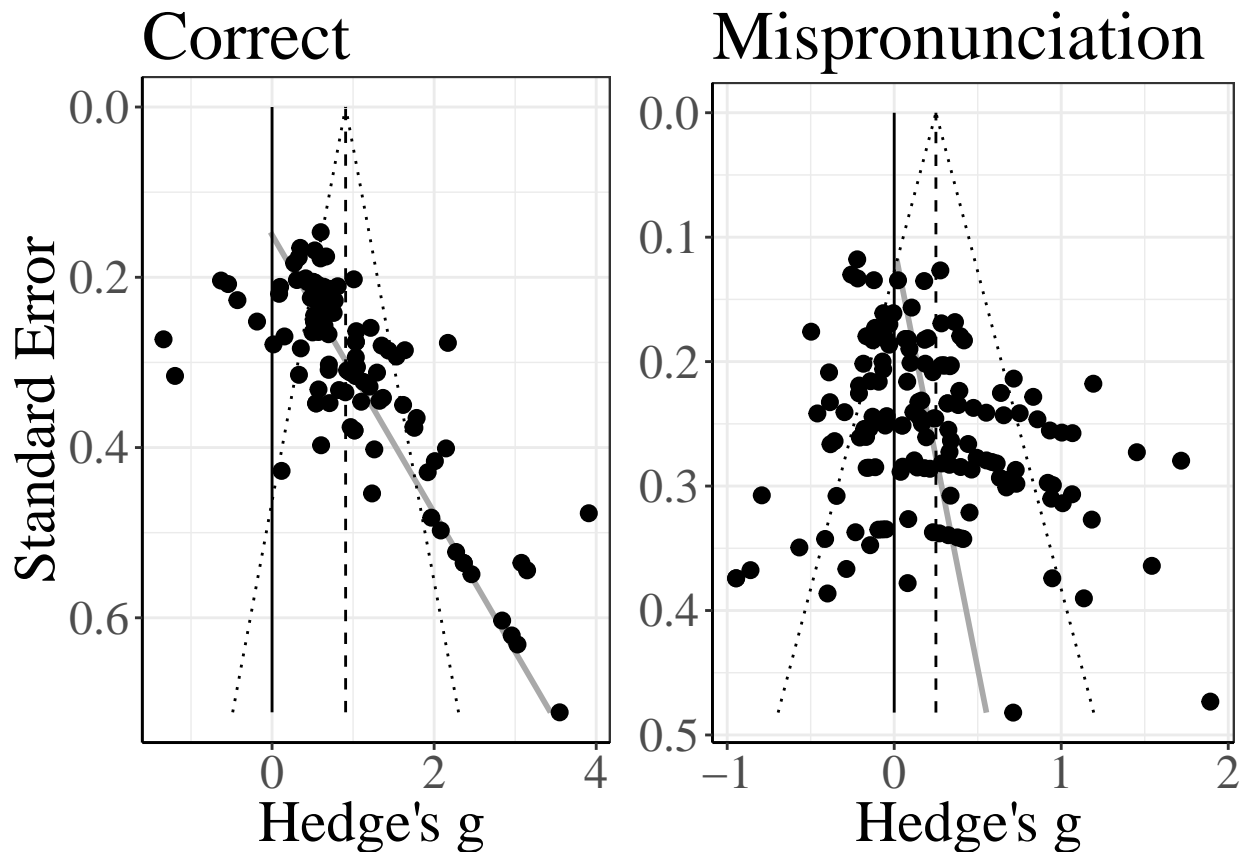
```
## [1] TRUE
```

```
## [1] TRUE
```

We next examined the p-curves for significant values from the correctly pronounced and mispronounced conditions. The 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 evidence of a large proportion of p-values just below the typical alpha threshold of .05. The p-curve based on 36 statistically significant values for mispronunciations indicates that the data contain evidential value (Z = -6.81, p < .0001) and there is no evidence of a large proportion of p-values just below the typical alpha threshold of .05.

Taken together, these results suggest a tendency in the literature towards publication bias. As a result, our meta-analysis may systematically overestimate effect sizes and we therefore interpret all estimates with caution. The funnel plot asymmetry may also reflect heterogeneity in the data, perhaps due to some studies investigating more subtle effects than other studies. Yet, the p-curve analysis suggests that overall, the literature contains evidential value, reflecting a "real" effect. We therefore continue our meta-analysis.

Meta-analysis

Object Identification for Correct and Mispronounced Words

```
rma_correct = rma.mv(g_calc, g_var_calc, data = db_ET_correct, random = ~collapse |
    short_cite)

# summary(rma_correct) kable(round(coef(summary(rma_correct)), 2))

sum_eff <- coef(summary(rma_correct))[1, ]
```

We first calculated the meta-analytic effect for object identification in response to correctly pronounced words. The variance-weighted meta-analytic effect size Hedges' *g* was 0.91 (SE = 0.12) which was significantly different from zero (95% CI [0.67, 1.14], p < .0001). This is a rather 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) and that it is significantly above zero suggests that when presented with the correctly pronounced label, infants fixated the corresponding object. Our analysis of funnel plot asymmetry, however, found evidence for publication. Although the effect size Hedges' *g* may be overestimated for object identification in response to correctly pronounced words, the p-curve results and a CI lower bound of 0.67 suggests that this result is robust even when correcting for publication bias.

```
rma_MP = rma.mv(g_calc, g_var_calc, data = db_ET_MP, random = ~collapse | short_cite)

# summary(rma_MP)

sum_eff <- coef(summary(rma_MP))[1, ]
```

We then calculated the meta-analytic effect for object identification in response to mispronounced words. In this case, the variance-weighted meta-analytic effect size Hedges' g was 0.25 (SE = 0.06) which was also significantly different from zero (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 fixated the correct object. We again note the publication bias (which was smaller in this condition), and the possibility that the effect size Hedges' *g* may be overestimated. But, as the p-curve indicated evidential value, we are confident in this result as well.

Heterogeneity was significant for both correctly pronounced (Q(103) = 625.63, p < .0001) and mispronounced words, (Q(146) = 462.51, p < .0001). This indicated that the sample contains unexplained variance leading to significant difference across our studies beyond what is to be expected based on random sampling error.

Mispronunciation Sensitivity Meta-analytic Effect

```
rma_MPeffect <- rma.mv(g_calc, g_var_calc, mods = ~condition, data = dat, random = ~collapse |
    short_cite)

# summary(rma_MPeffect)

# rma_MPeffect_1 <- rma.mv(g_calc, g_var_calc, mods = ~condition-1, data =
# dat, random = ~ collapse | short_cite)

# summary(rma_MPeffect_1)

sum_eff <- coef(summary(rma_MPeffect))[2, ]
```

The above two analyses considered the data from 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. The moderator test was significant, QM(1) = 215.76, p < .0001. Hedges' *g* for mispronunciation sensitivity was 0.5 (SE = 0.03), which indicated that the two type of responses were significantly different from one another (95% CI [0.43, 0.56], p < .0001). This confirms that although infants fixate the correct object for both correct pronunciations and mispronunciations, the observed fixations to target (as measured by the effect sizes) were significantly greater for correct pronunciations. In other words, we observe a significant difference between the two conditions and can now quantify the modulation of fixation behavior in terms of standardized effect sizes.

Object Recognition and Mispronunciation Sensitivity Modulated by Age

```
rma_correct_age = rma.mv(g_calc, g_var_calc, mods = ~age.C, data = db_ET_correct,
    random = ~collapse | short_cite)

# summary(rma_correct_age) kable(round(coef(summary(rma_correct_age)), 2))
# aov.type <- anova(rma_correct_age)

Csum_eff <- coef(summary(rma_correct_age))[2, ]
```

```
rma_MP_age = rma.mv(g_calc, g_var_calc, mods = ~age.C, data = db_ET_MP, random = ~collapse |
    short_cite)

# summary(rma_MP_age) aov.type <- anova(rma_MP_age)

Msum_eff <- coef(summary(rma_MP_age))[2, ]
```

To evaluate the different predictions for how mispronunciation sensitivity will change as infants develop, we next added the moderator age (centered, in days). In the first analyses, 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 $(QM(1) = 0.68, p = 0.41)$ or mispronounced words $(QM(1) = 1.72, p = 0.19)$. The lack of a significant modulation indicates that there was no relationship between age and target looks in response to a correctly pronounced or mispronounced label. This relationship is plotted in Figure 2.

```
rma_MPeffect_age <- rma.mv(g_calc, g_var_calc, mods = ~age.C * condition, data = dat,
    random = ~collapse | short_cite)

# summary(rma_MPeffect_age) aov.type <- anova(rma_MPeffect_age)

sum_eff <- coef(summary(rma_MPeffect_age))[4, ]
```

We then examined the interaction between age and mispronunciation sensitivity (correct vs. mispronounced words) in our whole dataset. The moderator test was significant $(QMQM(3) = 218.62, p < .0001)$. The interaction between age and mispronunciation sensitivity, however, was not significant $\beta = 0$, $(SE = 0.01$, 95% CI [-0.01, 0.02], p = 0.73), suggesting that as infants age, their mispronunciation sensitivity remains the same.

Figure 2
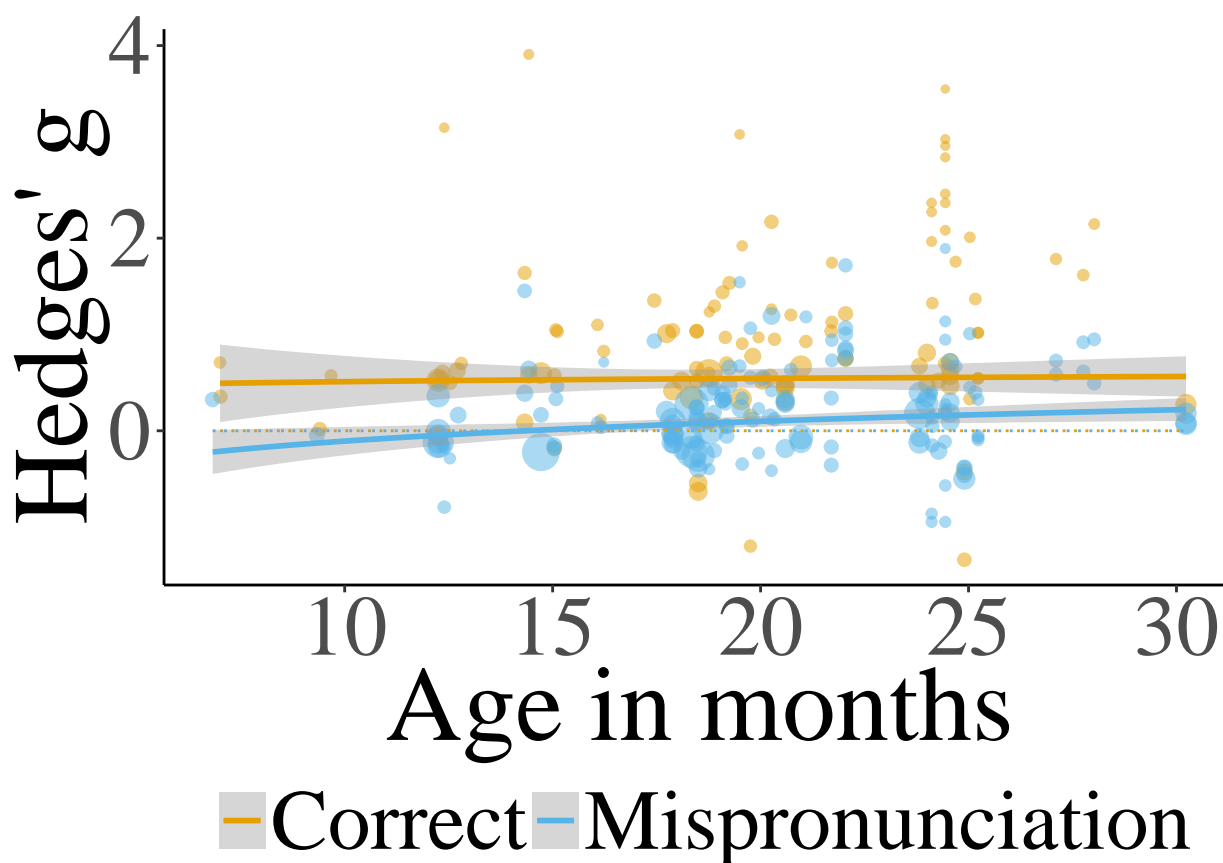

```
## pdf
##   2
```

Vocabulary Size: Correlation Between Mispronunciation Sensitivity and Vocabulary

Of the 32 papers included in the meta-analysis, 8 (comprehension = 7 papers; production = 1) analyzed the relationship between vocabulary scores and mispronunciation sensitivity. There is reason to believe that production data are different from comprehension data (the former being easier to estimate for parents in the typical questionnaire-based assessment), so we analyze this data separately.

```r
# we're relying on the library meta function metacor
compr <- subset(db_ET_correct, !is.na(db_ET_correct$r_comprehension) & r_comprehension >
    -1)

metacor(cor = r_comprehension, n = n_1, studlab = short_cite, data = compr,
    sm = "COR")
```

```
##                                 COR        95%-CI %W(fixed)
## Zesiger et al. (2012)          0.0610 [-0.3553; 0.4773]     5.8
## Zesiger et al. (2012)         -0.1590 [-0.5663; 0.2483]     6.1
## Mani, Coleman, & Plunkett (2008)  0.0300 [-0.2271; 0.2871]    15.2
## Swingley & Aslin (2000)        0.1050 [-0.1564; 0.3664]    14.7
## Mani & Plunkett 2007          -0.1700 [-0.5234; 0.1834]     8.0
## Mani & Plunkett 2007          -0.1700 [-0.5175; 0.1775]     8.3
## Swingley & Aslin (2002)        0.1410 [-0.2432; 0.5252]     6.8
```

```
## Swingley & Aslin (2002)            0.1410 [-0.2596; 0.5416]       6.3
## Swingley 2003                       0.3400 [ 0.0470; 0.6330]      11.7
## Swingley 2003                       0.0600 [-0.3472; 0.4672]       6.1
## H\xbfjen et al.                     0.2220 [-0.2591; 0.7031]       4.3
## H\xbfjen et al.                     0.4820 [ 0.0935; 0.8705]       6.7
##                                  %W(random)
## Zesiger et al. (2012)                    6.2
## Zesiger et al. (2012)                    6.5
## Mani, Coleman, & Plunkett (2008)        13.7
## Swingley & Aslin (2000)                 13.4
## Mani & Plunkett 2007                     8.3
## Mani & Plunkett 2007                     8.5
## Swingley & Aslin (2002)                  7.2
## Swingley & Aslin (2002)                  6.7
## Swingley 2003                           11.2
## Swingley 2003                            6.5
## H\xbfjen et al.                          4.8
## H\xbfjen et al.                          7.0
##
## Number of studies combined: k = 12
##
##                          COR           95%-CI    z p-value
## Fixed effect model    0.0897 [-0.0105; 0.1900] 1.75  0.0795
## Random effects model 0.0893 [-0.0212; 0.1999] 1.58  0.1132
##
## Quantifying heterogeneity:
## tau^2 = 0.0060; H = 1.09 [1.00; 1.50]; I^2 = 15.7% [0.0%; 55.4%]
##
## Test of heterogeneity:
##      Q d.f. p-value
##  13.05   11  0.2899
##
## Details on meta-analytical method:
## - Inverse variance method
## - DerSimonian-Laird estimator for tau^2
## - Untransformed correlations
# we're relying on the library meta function metacor
prodr <- subset(db_ET_correct, !is.na(db_ET_correct$r_production) & r_production <
    1)

metacor(cor = r_production, n = n_1, studlab = short_cite, data = prodr, sm = "COR")

##                                     COR          95%-CI %W(fixed)
## Zesiger et al. (2012)            -0.0090 [-0.4268; 0.4088]      5.0
## Zesiger et al. (2012)            -0.1720 [-0.5775; 0.2335]      5.3
## Mani, Coleman, & Plunkett (2008)  0.0700 [-0.1861; 0.3261]     13.2
## Mani & Plunkett 2007             -0.1100 [-0.4696; 0.2496]      6.7
## Mani & Plunkett 2007             -0.1100 [-0.4635; 0.2435]      6.9
## Swingley & Aslin (2002)           0.1820 [-0.1970; 0.5610]      6.0
## Swingley & Aslin (2002)           0.1820 [-0.2131; 0.5771]      5.6
## Swingley 2003                     0.1800 [-0.1406; 0.5006]      8.4
## Swingley 2003                     0.0700 [-0.3367; 0.4767]      5.2
## Ramon-Casas et al. 2009           0.0980 [-0.3068; 0.5028]      5.3
## Ramon-Casas et al. 2009          -0.1470 [-0.5468; 0.2528]      5.4
```

```
## Ramon-Casas et al. 2009          -0.2300 [-0.6171; 0.1571]       5.8
## Ramon-Casas et al. 2009           0.2400 [-0.1451; 0.6251]       5.9
## Ramon-Casas et al. 2009           0.4350 [ 0.1037; 0.7663]       7.9
## H\xbfjen et al.                   0.2220 [-0.2591; 0.7031]       3.7
## H\xbfjen et al.                  -0.1480 [-0.6430; 0.3470]       3.5
##                                   %W(random)
## Zesiger et al. (2012)                    5.0
## Zesiger et al. (2012)                    5.3
## Mani, Coleman, & Plunkett (2008)        13.2
## Mani & Plunkett 2007                     6.7
## Mani & Plunkett 2007                     6.9
## Swingley & Aslin (2002)                  6.0
## Swingley & Aslin (2002)                  5.6
## Swingley 2003                            8.4
## Swingley 2003                            5.2
## Ramon-Casas et al. 2009                  5.3
## Ramon-Casas et al. 2009                  5.4
## Ramon-Casas et al. 2009                  5.8
## Ramon-Casas et al. 2009                  5.9
## Ramon-Casas et al. 2009                  7.9
## H\xbfjen et al.                          3.7
## H\xbfjen et al.                          3.5
##
## Number of studies combined: k = 16
##
##                      COR          95%-CI    z p-value
## Fixed effect model   0.0601 [-0.0331; 0.1533] 1.26  0.2061
## Random effects model 0.0601 [-0.0331; 0.1533] 1.26  0.2061
##
## Quantifying heterogeneity:
## tau^2 = 0; H = 1.00 [1.00; 1.42]; I^2 = 0.0% [0.0%; 50.7%]
##
## Test of heterogeneity:
##      Q d.f. p-value
##  14.51   15  0.4870
##
## Details on meta-analytical method:
## - Inverse variance method
## - DerSimonian-Laird estimator for tau^2
## - Untransformed correlations
```

```r
# we're relying on the library meta function metacor
compr <- subset(db_ET_MP, !is.na(db_ET_MP$r_comprehension) & r_comprehension >
    -1)

metacor(cor = r_comprehension, n = n_1, studlab = short_cite, data = compr,
    sm = "COR")
```

```
##                              COR          95%-CI %W(fixed) %W(random)
## Bailey & Plunkett (2002)  0.0630 [-0.3441;  0.4701]      2.5        2.7
## Bailey & Plunkett (2002) -0.0660 [-0.4729;  0.3409]      2.5        2.7
## Bailey & Plunkett (2002)  0.0630 [-0.3441;  0.4701]      2.5        2.7
## Bailey & Plunkett (2002) -0.0660 [-0.4729;  0.3409]      2.5        2.7
## Bailey & Plunkett (2002)  0.0630 [-0.3441;  0.4701]      2.5        2.7
## Bailey & Plunkett (2002) -0.0660 [-0.4729;  0.3409]      2.5        2.7
```

```
## Bailey & Plunkett (2002)   0.0630 [-0.3441;  0.4701]      2.5      2.7
## Bailey & Plunkett (2002)  -0.0660 [-0.4729;  0.3409]      2.5      2.7
## Swingley (2009)           -0.1200 [-0.3715;  0.1315]      6.4      5.6
## Swingley (2009)            0.1300 [-0.1957;  0.4557]      3.8      3.8
## Swingley & Aslin (2000)    0.4210 [ 0.2036;  0.6384]      8.6      6.7
## Mani & Plunkett 2007      -0.0590 [-0.4349;  0.3169]      2.9      3.0
## Mani & Plunkett 2007      -0.0590 [-0.4349;  0.3169]      2.9      3.0
## Mani & Plunkett 2007      -0.1700 [-0.5234;  0.1834]      3.3      3.3
## Mani & Plunkett 2007      -0.1700 [-0.5175;  0.1775]      3.4      3.4
## Mani & Plunkett 2007       0.1000 [-0.2667;  0.4667]      3.0      3.2
## Mani & Plunkett 2007      -0.2800 [-0.6214;  0.0614]      3.5      3.5
## Mani & Plunkett 2007      -0.0700 [-0.3354;  0.1954]      5.8      5.2
## Mani & Plunkett 2007       0.1000 [-0.1641;  0.3641]      5.8      5.2
## Swingley & Aslin (2002)    0.1750 [-0.2050;  0.5550]      2.8      3.0
## Swingley & Aslin (2002)    0.1750 [-0.2212;  0.5712]      2.6      2.8
## Mani & Plunkett 2010      -0.3500 [-0.6640; -0.0360]      4.1      4.0
## Swingley 2003              0.1500 [-0.1738;  0.4738]      3.9      3.8
## Swingley 2003              0.1500 [-0.1738;  0.4738]      3.9      3.8
## Swingley 2003              0.0600 [-0.3472;  0.4672]      2.4      2.7
## Swingley 2003              0.0600 [-0.3472;  0.4672]      2.4      2.7
## H\xbfjen et al.           -0.0510 [-0.5557;  0.4537]      1.6      1.8
## H\xbfjen et al.            0.0510 [-0.4537;  0.5557]      1.6      1.8
## H\xbfjen et al.            0.0600 [-0.4442;  0.5642]      1.6      1.8
## H\xbfjen et al.           -0.0600 [-0.5642;  0.4442]      1.6      1.8
## Tao et al. 2012            0.5700 [ 0.1710;  0.9690]      2.6      2.7
##
## Number of studies combined: k = 31
##
##                      COR            95%-CI    z p-value
## Fixed effect model   0.0377 [-0.0260; 0.1014] 1.16  0.2465
## Random effects model 0.0322 [-0.0399; 0.1043] 0.87  0.3818
##
## Quantifying heterogeneity:
## tau^2 = 0.0079; H = 1.11 [1.00; 1.39]; I^2 = 19.3% [0.0%; 48.4%]
##
## Test of heterogeneity:
##     Q d.f. p-value
##  37.18   30  0.1719
##
## Details on meta-analytical method:
## - Inverse variance method
## - DerSimonian-Laird estimator for tau^2
## - Untransformed correlations
```

```r
# we're relying on the library meta function metacor
prodr <- subset(db_ET_MP, !is.na(db_ET_MP$r_production) & r_production < 1)

metacor(cor = r_production, n = n_1, studlab = short_cite, data = prodr, sm = "COR")
```

```
##                          COR           95%-CI %W(fixed) %W(random)
## Swingley (2009)       0.0700 [-0.1839;  0.3239]      6.4      5.4
## Swingley (2009)       0.1600 [-0.1628;  0.4828]      3.9      3.8
## Mani & Plunkett 2007 -0.2820 [-0.6292;  0.0652]      3.4      3.4
## Mani & Plunkett 2007 -0.2820 [-0.6292;  0.0652]      3.4      3.4
## Mani & Plunkett 2007 -0.1100 [-0.4696;  0.2496]      3.2      3.3
```

```
## Mani & Plunkett 2007      -0.1100 [-0.4635;  0.2435]      3.3      3.3
## Mani & Plunkett 2007       0.2300 [-0.1208;  0.5808]      3.3      3.4
## Mani & Plunkett 2007       0.2200 [-0.1325;  0.5725]      3.3      3.4
## Mani & Plunkett 2007      -0.2700 [-0.5173; -0.0227]      6.7      5.5
## Mani & Plunkett 2007      -0.0590 [-0.3248;  0.2068]      5.8      5.0
## Swingley & Aslin (2002)   0.0290 [-0.3627;  0.4207]      2.7      2.8
## Swingley & Aslin (2002)   0.0290 [-0.3793;  0.4373]      2.5      2.7
## Swingley (2016)          -0.2690 [-0.6326;  0.0946]      3.1      3.2
## Swingley (2016)          -0.2690 [-0.6326;  0.0946]      3.1      3.2
## Swingley (2016)          -0.2690 [-0.6326;  0.0946]      3.1      3.2
## Swingley (2016)          -0.2690 [-0.6326;  0.0946]      3.1      3.2
## Swingley (2016)          -0.2690 [-0.6326;  0.0946]      3.1      3.2
## Swingley (2016)          -0.2690 [-0.6326;  0.0946]      3.1      3.2
## Swingley 2003             0.1500 [-0.1738;  0.4738]      3.9      3.8
## Swingley 2003             0.1500 [-0.1738;  0.4738]      3.9      3.8
## Swingley 2003             0.0700 [-0.3367;  0.4767]      2.5      2.7
## Swingley 2003             0.0700 [-0.3367;  0.4767]      2.5      2.7
## Ramon-Casas et al. 2009   0.0980 [-0.3068;  0.5028]      2.5      2.7
## Ramon-Casas et al. 2009  -0.1470 [-0.5468;  0.2528]      2.6      2.7
## Ramon-Casas et al. 2009  -0.2300 [-0.6171;  0.1571]      2.7      2.9
## Ramon-Casas et al. 2009   0.2400 [-0.1451;  0.6251]      2.8      2.9
## Ramon-Casas et al. 2009   0.4350 [ 0.1037;  0.7663]      3.7      3.7
## H\xbfjen et al.          -0.0510 [-0.5557;  0.4537]      1.6      1.8
## H\xbfjen et al.           0.0510 [-0.4537;  0.5557]      1.6      1.8
## H\xbfjen et al.           0.0790 [-0.4239;  0.5819]      1.6      1.9
## H\xbfjen et al.          -0.0790 [-0.5819;  0.4239]      1.6      1.9
##
## Number of studies combined: k = 31
##
##                           COR          95%-CI     z p-value
## Fixed effect model    -0.0402 [-0.1043; 0.0238] -1.23  0.2181
## Random effects model  -0.0393 [-0.1127; 0.0340] -1.05  0.2931
##
## Quantifying heterogeneity:
## tau^2 = 0.0094; H = 1.13 [1.00; 1.42]; I^2 = 22.0% [0.0%; 50.1%]
##
## Test of heterogeneity:
##      Q d.f. p-value
##  38.45   30  0.1386
##
## Details on meta-analytical method:
## - Inverse variance method
## - DerSimonian-Laird estimator for tau^2
## - Untransformed correlations
```

Interim Discussion

The main goal of this paper was to assess mispronunciation sensitivity and its maturation with age. The results are clear: Although infants consider a mispronunciation as a better match with the target image than a distractor image, there was a consistent effect of mispronunciation sensitivity. This did not change with development. Of the 3 predictions and assumptions about the development of infants' sensitivity to mispronunciations discussed in the Introduction, the present results lend some support for the argument that mispronunciation sensitivity stays consistent as infants develop. This runs counter to existing theories of

phono-lexical development, which predict either an increase (PRIMR ref) or decrease (Assim Model ref) in mispronunciation sensitivity. In sum, it seems that current theories of infants' phono-lexical development cannot fully capture our results and should be reconsidered with all the evidence in mind.

Alternatively, the lack of developmental change in mispronunciation sensitivity could be due to differences in the types of tasks given to infants of different ages. To examine this possibility, we include an exploratory analysis of whether different moderators and experimental design features were included at different ages. We return to this possibility after the Moderator Analyses.