MP MetaAnalysis

Christina Bergmann and Katie Von Holzen

10 October, 2018

Contents

Preparation **2**

Methods **2**
Study Selection 2
Data Entry 4
Data analysis 4
Publication Bias 5
Meta-analysis 6

Results **6**
Publication Bias 6
Figure 1 6
Meta-analysis 7
Figure 2 10
Moderator Analyses 15

Discussion **36**

```
## 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 2.0-0). 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

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser
```

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). Figure X plots our PRISMA flowchart.

[Figure X. PRISMA Flowchart.] (figures/PRISMA_MA_Mispronunciation.png)

Study Selection

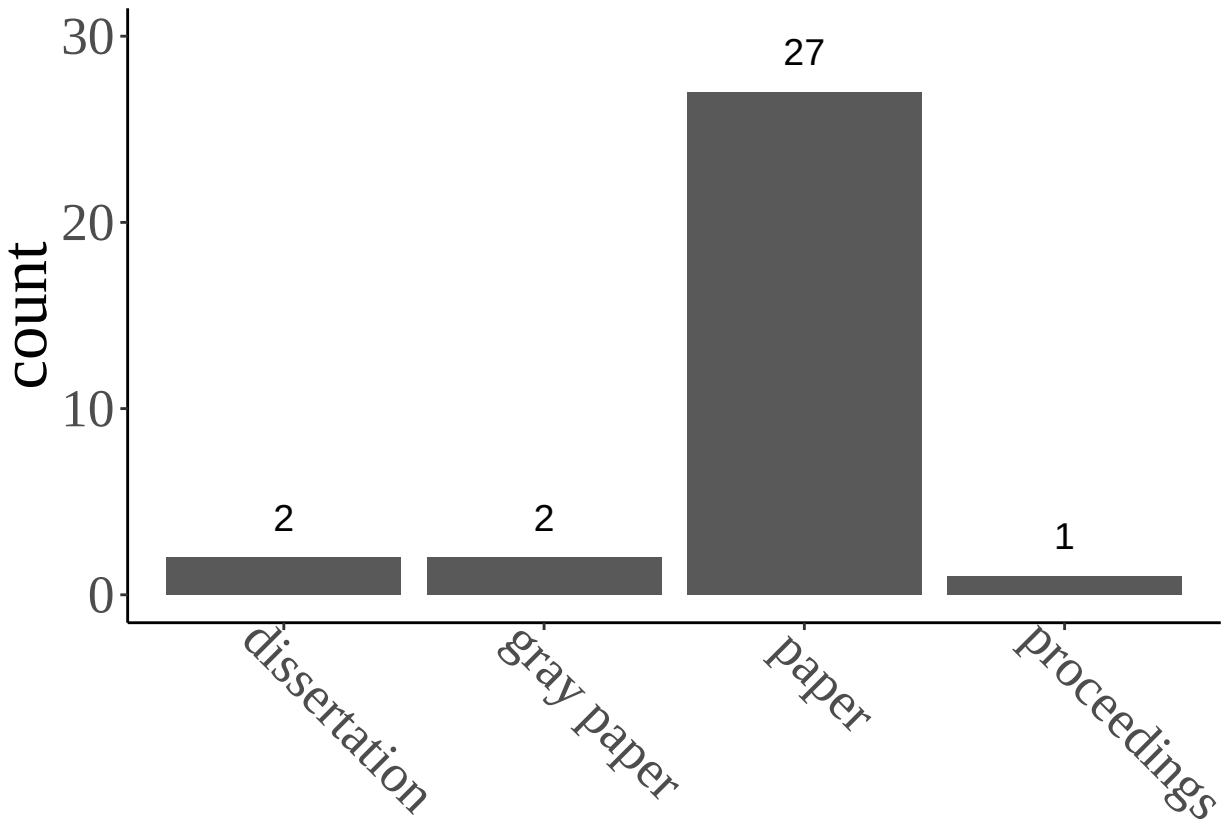
short_cite	Ages_Tested_months	Vocabulary	Number_of_Features	Distrac
Altvater-Mackensen (2010)	22, 25	None	1	Familia
Altvater-Mackensen et al. (2014)	18, 25	None	0, 1	Familia
Bailey & Plunkett (2002)	18, 24	Comprehension	0, 1, 2	Familia

short_cite	Ages_Tested_months	Vocabulary	Number_of_Features	Distrac
Bergelson & Swingley (2017)	7, 9, 12, 6	None	0, NA	Familia
Bernier & White 2017	21	None	0, 1, 2, 3	Unfam
Delle Luche et al. (2015)	20, 19	None	0, 1	Familia
Durrant et al. (2014)	19, 20	None	0, 1	Familia
Hoehle et al. 2006	18	None	1	Familia
Hojen et al.	20	Comprehension/Production	2_3	Familia
Mani & Plunkett 2007	15, 18, 24, 14, 21	Comprehension/Production	0, 1_2, 1	Familia
Mani & Plunkett 2010	12	Comprehension	0, 1	Familia
Mani & Plunkett 2011	23, 17	None	0, 1_2_3, 1, 2, 3	Unfam
Mani, Coleman, & Plunkett (2008)	18	Comprehension/Production	0, 1	Familia
Ramon-Casas & Bosch 2010	24, 25	None	NA	Familia
Ramon-Casas et al. 2009	21, 20	Production	NA	Familia
Ren & Morgan, in press	19	None	0, 1	Unfam
Skoruppa et al. 2013	24	None	0, 1	Unfam
Swingley (2009)	17	Comprehension/Production	0, 1	Familia
Swingley (2016)	27, 28	Production	0, 1	Unfam
Swingley & Aslin (2000)	20	Comprehension	0, 1	Familia
Swingley & Aslin (2002)	15	Comprehension/Production	0, 1, 2	Familia
Swingley 2003	19	Comprehension/Production	0, 1	Familia
Tamasi (2016)	30	None	0, 1, 2, 3	Unfam
Tao & Qinmei 2013	12	None	NA	Familia
Tao et al. 2012	16	Comprehension	NA	Familia
van der Feest & Fikkert, 2015	24, 20	None	0, 1	Familia
van der Feest & Johnson, 2016	24	None	0, 1	Familia
Wewalaarachchi et al. 2017	24	None	0, 1	Unfam
White & Aslin (2011)	18	None	1	Unfam
White & Morgan (2008)	18, 19	None	0, 1, 2, 3	Unfam
Zesiger & Johr 2011	14	None	0, 1	Familia
Zesiger et al. (2012)	12, 19	Comprehension/Production	0, 1, 2	Familia

[KATIE] THIS TABLE IS DEFINITELY NOT FINISHED!

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

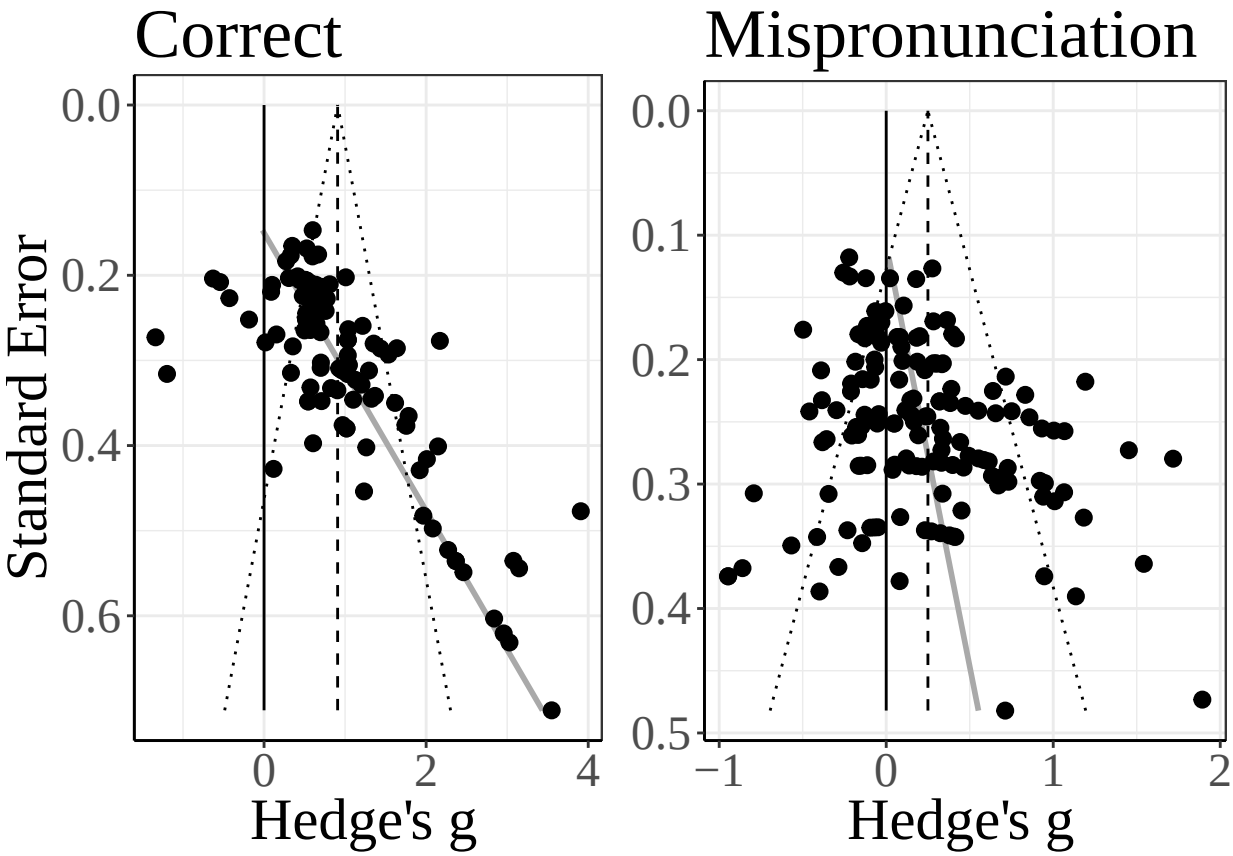
Figure 1 plots the funnel plots for object identification for both correct pronunciations and mispronunciations (code adapted from Sakaluk, 2016). Funnel plot assymmetry was significant for both correct pronunciations (Kendall's $\tau = 0.53$, $p < .0001$) and mispronunciations (Kendall's $\tau = 0.16$, $p < .01$). 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).

```
## Error in select(., cite, expt_num, same_infant, mean_age_1, yi, vi): unused arguments (cite, expt_nu
```

```
## Error in select(., cite, expt_num, same_infant, mean_age_1, yi, vi): unused arguments (cite, expt_nu
```

Figure 1

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




```
## [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, ]
```

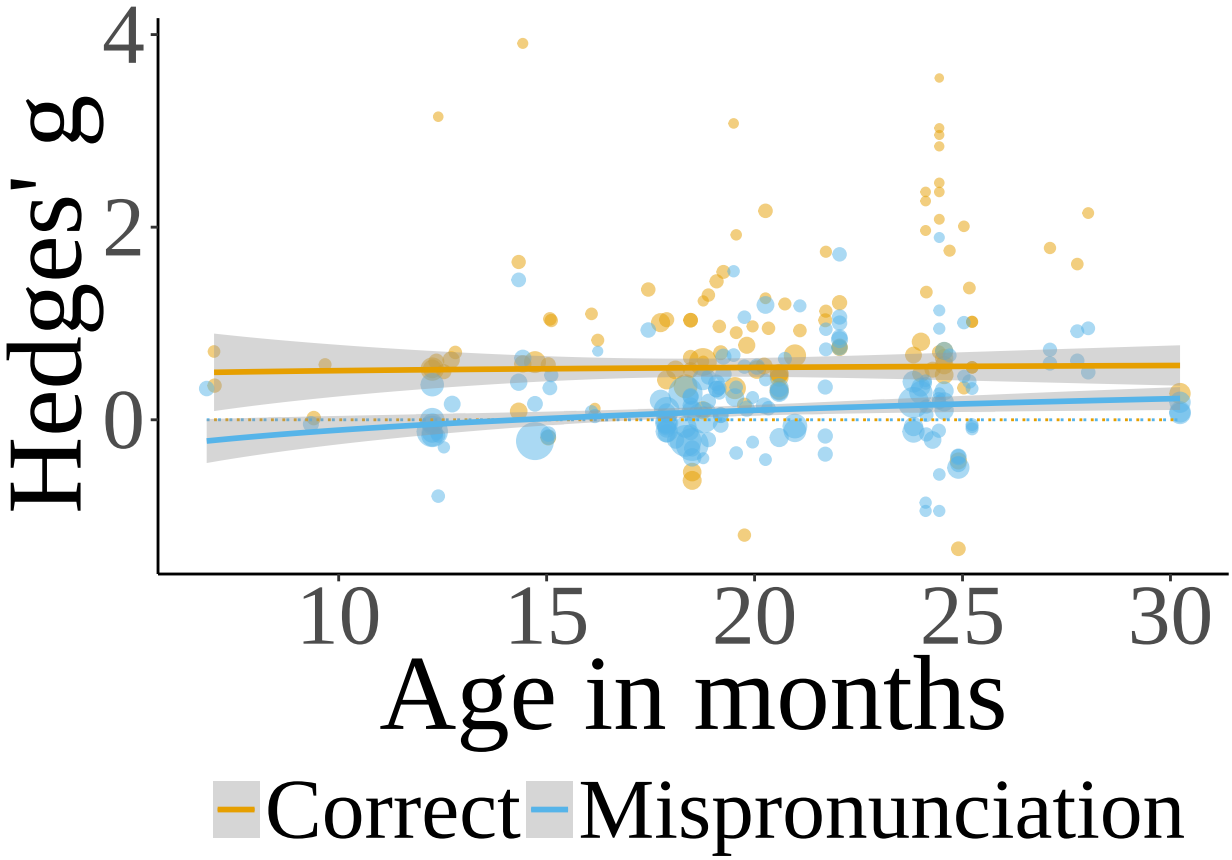
```
rma_MP_age = rma.mv(g_calc, g_var_calc, mods = ~age.C, data = db_ET_MP, random = ~collapse |
    short_cite)

# 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.

[KATIE] I DON'T KNOW HOW TO INTERPRET THESE RESULTS… THE FIXED EFFECTS MODEL ISN'T SIGNIFICANT FOR ANY OF THESE. IS THAT MEANINGFUL?

```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
## Hojen et al.                      0.2220 [-0.2591; 0.7031]       4.3
## Hojen 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
## Hojen et al.                            4.8
## Hojen 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
```

```r
# 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
## Hojen et al.                       0.2220 [-0.2591; 0.7031]       3.7
## Hojen 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
## Hojen et al.                             3.7
## Hojen 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
## Hojen et al.             -0.0510 [-0.5557;  0.4537]       1.6        1.8
## Hojen et al.              0.0510 [-0.4537;  0.5557]       1.6        1.8
## Hojen et al.              0.0600 [-0.4442;  0.5642]       1.6        1.8
## Hojen 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
## Hojen et al.            -0.0510 [-0.5557;  0.4537]       1.6        1.8
## Hojen et al.             0.0510 [-0.4537;  0.5557]       1.6        1.8
## Hojen et al.             0.0790 [-0.4239;  0.5819]       1.6        1.9
## Hojen 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.

[KATIE] CAN A CONCLUSION BE DRAWN FROM THE VOCABULARY RESULTS? IF SO, THIS IS THE PARAGRAPH

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, in addition to investigating the role that these moderators play in mispronunciation sensitivity.

Moderator Analyses

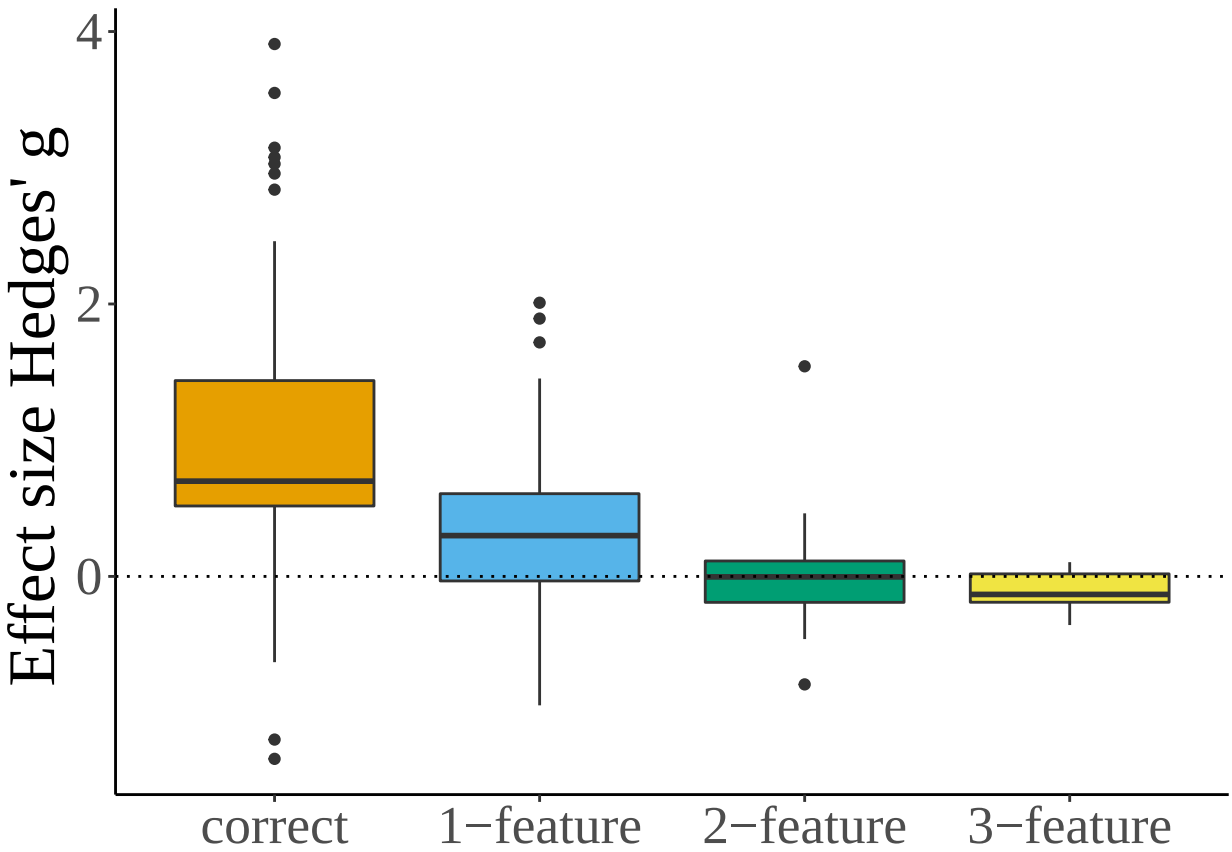
Size of mispronunciation

```
dat.f <- subset(dat, n_feature == "0" | n_feature == "1" | n_feature == "2" |
    n_feature == "3")

nfeature_info <- dat.f %>% group_by(n_feature) %>% summarize(n_exp_conditions = n())
```

To assess whether the number of features changed modulates mispronunciation sensitivity, we calculated the meta-analytic effect for object identification in response to words that were pronounced correctly and mispronounced using 1-, 2-, and 3-feature changes. We did not include data for which the number of features changed in a mispronunciation was not specified or the number of features changed was not consistent. The analysis was therefore based on a subset of the overall dataset, with 81 experimental conditions for correct pronunciations, 108 for 1-feature mispronunciations, 16 for 2-feature mispronunciations, and 6 for 3-feature mispronunciations. Each feature change (from 0 to 3; 0 representing correct pronunciations) was considered to have an equal impact on mispronunciation sensitivity, following the argument of graded sensitivity (White & Aslin, 2008; Mani & Plunkett 2011).

```
dat.f <- subset(dat, n_feature == "0" | n_feature == "1" | n_feature == "2" |
    n_feature == "3")

rma_NFeatures <- rma.mv(g_calc, g_var_calc, mods = ~as.numeric(n_feature), data = dat.f,
    random = ~collapse | short_cite)

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

To understand the relationship between size of mispronunciation and mispronunciation sensitivity, we evaluated the effect size Hedges' g with number of features changed as a moderator. The moderator test was significant, QM(1) = 76.1, p < .0001. Hedges' g for number of features changed was -0.12 (SE = 0.01), which indicated that as the number of features changed increased, the effect size Hedges' g significantly decreased (95% CI [-0.15, -0.1], p < .0001). We plot this relationship in Figure X. This confirms previous findings of a graded sensitivity to the number of features changed for both consonant (White & Morgan, 2008) and vowel (Mani & Plunkett, 2011) mispronunciations as well as the importance of controlling for the degree of phonological mismatch in experimental design.




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

```
dat.f <- subset(dat, n_feature == "0" | n_feature == "1" | n_feature == "2" |
    n_feature == "3")

rma_NFeatures <- rma.mv(g_calc, g_var_calc, mods = ~as.numeric(n_feature) *
    age.C, data = dat.f, random = ~collapse | short_cite)

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

Although we did not have any specific predictions about the relationship between infant age and the impact of number of features changed on mispronunciation sensitivity, we included an exploratory analysis to examine this relationship. When age was also included as a moderator, the moderator test was significant, QM(3) = 81.11, p < .0001, but the interaction between age and number of features changed was not significant, β = 0.01, (SE = 0, 95% CI [0, 0.01], p = 0.07). This suggests that there is no relationship between infant age the impact of number of features changed on mispronunciation sensitivity.

```
dat_f <- subset(dat, n_feature == "0" | n_feature == "1" | n_feature == "2" |
    n_feature == "3")

nfeature_ages <- dat.f %>% group_by(n_feature) %>% summarize(min_age = r2(min(mean_age_1,
    na.rm = TRUE)), max_age = r2(max(mean_age_1, na.rm = TRUE)), mean_age = r2(mean(mean_age_1,
    na.rm = TRUE)))

mf <- subset(dat_f, n_feature == "3")
min_age <- min(mf$mean_age_1)
max_age <- max(mf$mean_age_1)
```

```r
dat_fage = dat_f %>% filter(mean_age_1 >= min_age & mean_age_1 <= max_age)

rma_NFeatures_agesub <- rma.mv(g_calc, g_var_calc, mods = ~as.numeric(n_feature),
    data = dat_fage, random = ~collapse | short_cite)

sum_eff1 <- coef(summary(rma_NFeatures_agesub))[2, ]

rma_AGENFeatures_agesub <- rma.mv(g_calc, g_var_calc, mods = ~as.numeric(n_feature) *
    age.C, data = dat_fage, random = ~collapse | short_cite)

sum_eff2 <- coef(summary(rma_AGENFeatures_agesub))[4, ]
```

Although all papers included in the dataset also included correct pronunciations, not all papers included all three types of feature changes (i.e. 1-3). The age range for each type of number of features changed was 372.89 - 920.2 days (M = 637.4) for 1-feature mispronunciations, 377.28 - 920.2 days (M = 612.17) for 2-feature mispronunciations, and 544.48 - 920.2 days (M = 661.64) for 3-feature mispronunciations. Focusing on the ages where all three numbers of features changed were tested (i.e. 544.48 - 920.2 days), however, did not change the pattern of results.

[KATIE] DOES THE ABOVE PARAGRAPH NEED MORE MOTIVATION? SHOULD IT EVEN BE INCLUDED?

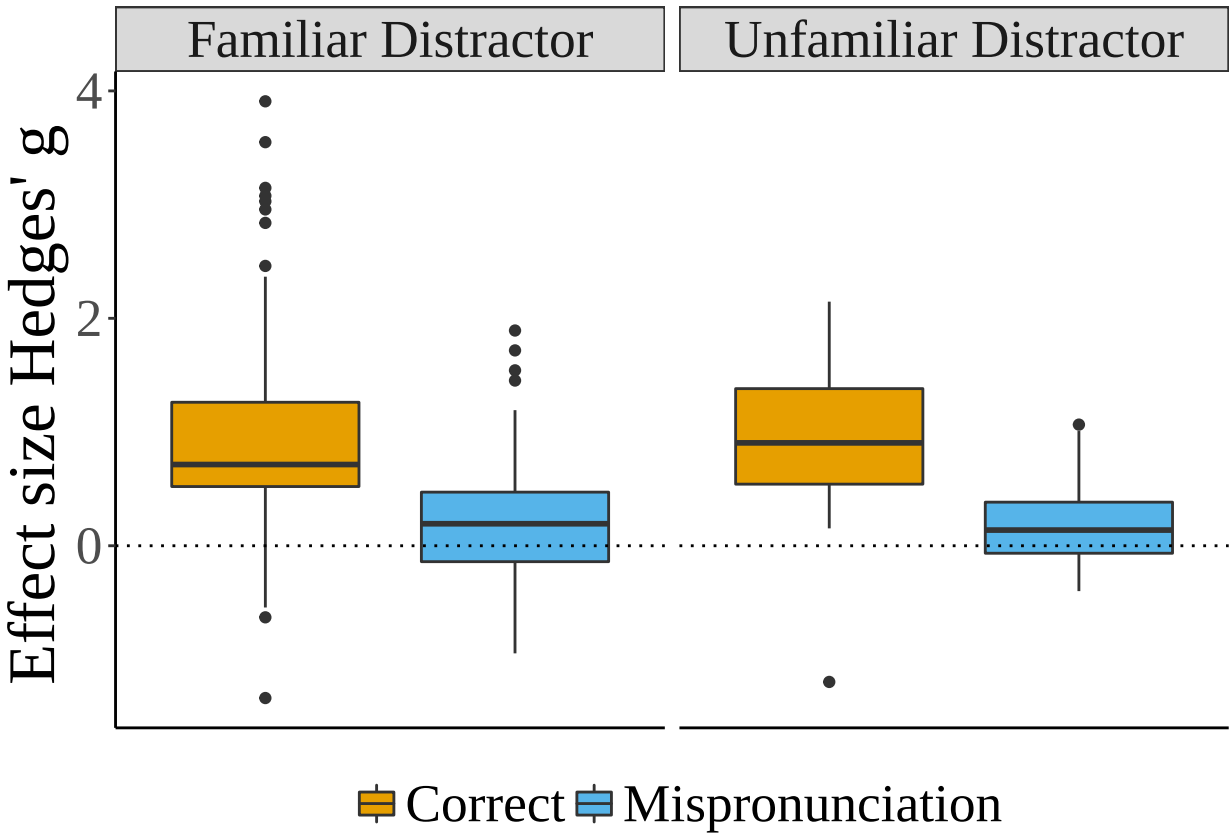
Distractor familiarity

```r
rma_Distractor_MP <- rma.mv(g_calc, g_var_calc, mods = ~as.factor(object_pair),
    data = db_ET_MP, random = ~collapse | short_cite)


rma_Distractor <- rma.mv(g_calc, g_var_calc, mods = ~condition * as.factor(object_pair),
    data = dat, random = ~collapse | short_cite)

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

We next assessed whether distractor familiarity has an impact on the size of mispronunciation sensitivity. First, we calculated the meta-analytic effect for object identification in response to mispronounced target words/images that were paired with either a familiar or an unfamiliar distractor image. The moderator test was not significant QM(1) = 0.43, p = 0.51. This suggests that upon hearing a mispronunciation, infants looks to the target image were similar for when the target image was paired with an image of a familiar or unfamiliar object. We next assessed whether distractor familiarity was related to mispronunciation sensitivity. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as an additional moderator. The moderator test was significant, QM(3) = 219.46, p < .0001, but the interaction between distractor familiarity and condition was not significant $\beta = 0.14$, (SE = 0.08, 95% CI [-0.02, 0.3], p = 0.08). These results suggest that overall, infants' familiarity with the distractor object (familiar or unfamiliar) did not impact their mispronunciation sensitivity.

```r
dat$condition_label = ifelse(dat$condition == 1, "Correct", "Mispronunciation")
dat$dist_code <- ifelse(dat$object_pair == "familiar_familiar", "Familiar Distractor",
    "Unfamiliar Distractor")

# Color Blind palette:
cbPalette <- c("#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00",
    "#CC79A7")
```

```
p <- ggplot(dat, aes(condition_label, g_calc, fill = condition_label)) + geom_boxplot() +
    facet_grid(. ~ dist_code) + # geom_smooth(method = 'lm', formula = y ~ log(x), aes(weight=weights_g
scale_fill_manual(values = cbPalette) + apatheme + theme(text = element_text(size = 25),
    legend.title = element_blank(), legend.position = "bottom", axis.title.x = element_blank(),
    axis.text.x = element_blank(), axis.ticks.x = element_blank()) + # xlab('Number of Features Changed
geom_hline(yintercept = 0, linetype = "dotted") + ylab("Effect size Hedges' g")

p
```




```
jpeg(filename = "figures/Figure_4Distractor_fam.jpg", width = 500, height = 300,
    units = "px")

p

dev.off()
```

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

```
rma_DistractorAge <- rma.mv(g_calc, g_var_calc, mods = ~age.C * as.factor(object_pair),
    data = db_ET_MP, random = ~collapse | short_cite)

sum_eff <- coef(summary(rma_DistractorAge))[4, ]


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

```r
# summary(rma_DistractorAgeMS)

sum_eff1 <- coef(summary(rma_DistractorAgeMS))[8, ]
sum_eff2 <- coef(summary(rma_DistractorAgeMS))[7, ]
```

We next examined whether age modulates object recognition or mispronunciation sensitivity when the distractor image is familiar or unfamiliar. For object recognition in response to a mispronunciation, including age as a moderator resulted in a moderator test that was not significant, QM(3) = 3.24, p = 0.36. This suggests that upon hearing a mispronunciation, infants looks to the target image were similar for when the target image was paired with an image of a familiar or unfamiliar object, regardless of their age. We next assessed whether the relationship between distractor familiarity and mispronunciation sensitivity was modulated by age. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as well as age as additional moderators. The moderator test was significant, QM(7) = 224.96, p < .0001. Although the three-way-interaction between condition, distractor familiarity, and age was not significant (β = -0.02, SE = 0.02, 95% CI [-0.06, 0.02], p = 0.3), here the interaction between condition and distractor familiarity was significant (β = 0.18, SE = 0.09, 95% CI [0, 0.35], p = 0.05).

[KATIE] WHAT DOES IT MEAN THAT IN THIS LAST TEST THERE WAS AN INTERACTION BETWEEN CONDITION AND DISTRACTOR FAMILIARITY, WHEN AGE WAS INCLUDED AS A MODERATOR? DOES THAT MEAN THAT AGE IS CONTROLLED FOR? UNDERSTANDING THIS WILL HELP ME FORMULATE THE CONCLUSION SENTENCE AND MOTIVATE THE NEXT SET OF ANALYSES.

```r
object_pair_ages <- dat %>% group_by(object_pair) %>% summarize(min_age = r2(min(mean_age_1,
    na.rm = TRUE)), max_age = r2(max(mean_age_1, na.rm = TRUE)), mean_age = r2(mean(mean_age_1,
    na.rm = TRUE)))


fn <- subset(dat, object_pair == "familiar_novel")

ff <- subset(dat, object_pair == "familiar_familiar")

age_diff_dist <- t.xtable(t.test(fn$mean_age_1, ff$mean_age_1, paired = F))
```

```
##                        Test                              Results
## 1 Welch Two Sample t-test: t(153.83) = 5.30, p < .001, d = NA
```

```r
# t.test(fn$mean_age_1, ff$mean_age_1)

opa <- as.data.frame(object_pair_ages)
min_age <- max(opa$min_age)
max_age <- min(opa$max_age)


dat_age_MP = db_ET_MP %>% filter(mean_age_1 >= min_age & mean_age_1 <= max_age)

rma_Distractor_MP <- rma.mv(g_calc, g_var_calc, mods = ~as.factor(object_pair),
    data = dat_age_MP, random = ~collapse | short_cite)

# summary(rma_Distractor_MP)

sum_eff <- round(coef(summary(rma_Distractor_MP))[2, ], 2)

# misp sensitivity
dat_age = dat %>% filter(mean_age_1 >= min_age & mean_age_1 <= max_age)
```

```
rma_Distractor_MS <- rma.mv(g_calc, g_var_calc, mods = ~as.factor(object_pair) *
    condition, data = dat_age, random = ~collapse | short_cite)

# summary(rma_Distractor_MS)

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

Although we anticipated that older children may be more impacted by the presence of a unfamiliar compared to familiar distractor image, we found that age and distractor familiarity did not impact mispronunciation sensitivity. Inspection of the ages tested using different kinds of distractors, however, revealed differences. Infants tested using a familiar distractor were younger (M = 588.76 days, SD = 136.47, range = 207.8 - 768 days) than those infants tested using an unfamiliar distractor (M = 678.85 days, SD = 115.47, range = 544.48 - 920.2 days), which a two-sample t-test revealed to be a significant difference, t(153.83) = 5.30, p < .0001.

[KATIE] SO, IS IT OKAY TO DO THE AGE SUBSET ANALYSIS AFTER THIS? SOMETHING LIKE "TO ENSURE THAT THE LACK OF A DIFFERENCE WASN'T DUE TO THE DIFFERENCE IN AGES OF INFANTS TESTED WITH DIFFERENT TYPES OF DISTRACTORS, WE ANALYZED A SUBSET OF PAPERS THAT TESTED AGES WHERE BOTH FAMILIAR AND UNFAMILIAR DISTRACTORS WERE USED." IN THE SUBSET ANALYSIS, FOR MISP SENSITIVITY, THERE IS A SIGNIFICANT INTERACTION BETWEEN DISTRACTOR FAMILIARITY AND CONDITION.

Phonological overlap between target and distractor

```
dist_overlap_info <- dat %>% group_by(distractor_overlap) %>% summarize(n_exp_conditions = n())
```

To assess whether phonological overlap between the target and distractor image labels has an impact on the size of mispronunciation sensitivity, we examined the meta-analytic effect for object identification in response to mispronunciations and mispronunciation sensitivity when the target-distractor pairs either shared the same onset phoneme, had no overlap, or where the distractor was an unfamiliar object. We did not include data for which the overlap included both the onset and medial phonemes ($n = 4$) or coda phonemes ($n = 3$). The analysis was therefore based on a subset of the overall dataset, with 104 experimental conditions containing onset phoneme overlap between the target and distractor, 80 containing no overlap between target and distractor, and 60 for targets paired with unfamiliar distractor images. There were therfore three possibilities for distractor overlap: onset phoneme overlap, no overlap, and an unfamiliar distractor. In our analysis, we were interested in the difference between overlap and lack of overlap; therefore, experimental conditions containing onset phoneme overlap were coded as the reference condition and compared with responses to experimental conditions with no overlap or an unfamiliar distractor separately.

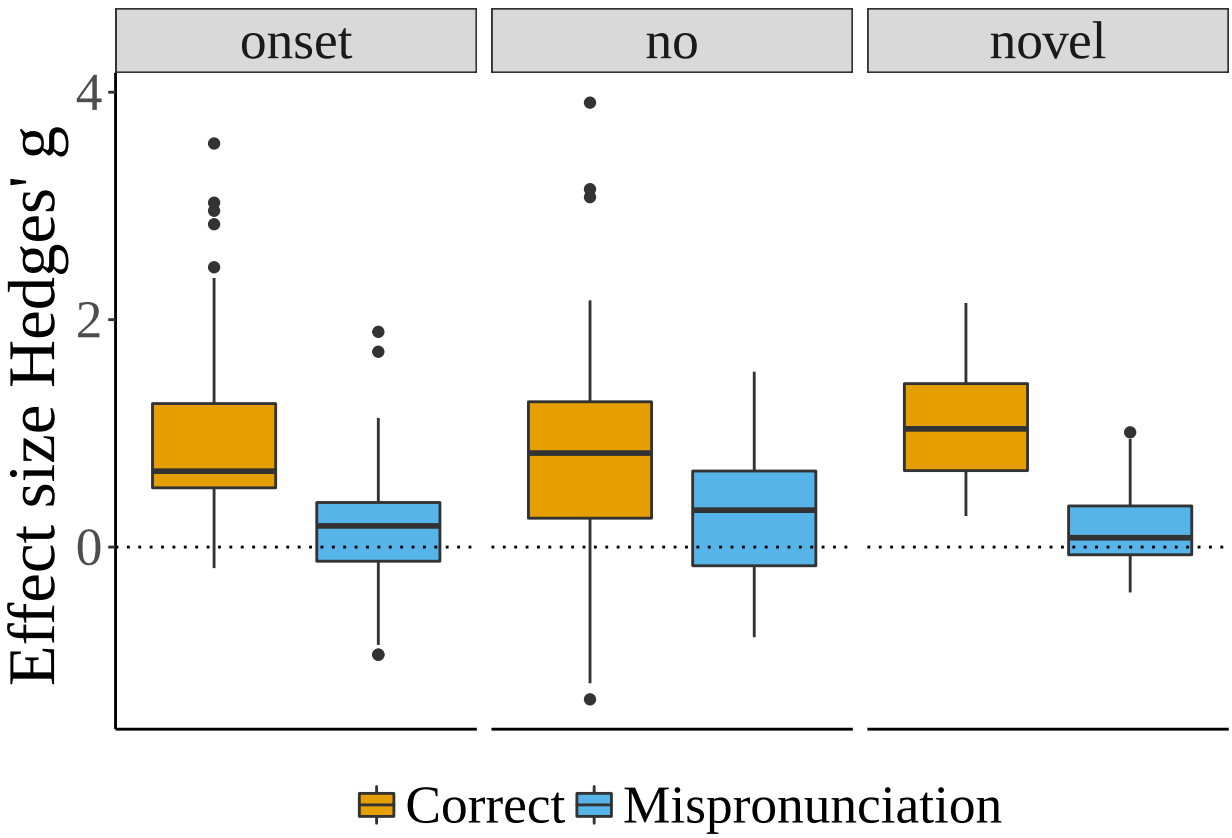
[KATIE] CAN YOU CHECK THE WAY THAT I WROTE THE LAST SENTENCE?

```
db_ET_MPo = db_ET_MP %>% filter(distractor_overlap == "onset" | distractor_overlap ==
    "novel" | distractor_overlap == "no")

db_ET_MPo$distractor_overlap <- factor(db_ET_MPo$distractor_overlap, levels = c("onset",
    "no", "novel"))


rma_DistractorOverlap <- rma.mv(g_calc, g_var_calc, mods = ~distractor_overlap,
    data = db_ET_MPo, random = ~collapse | short_cite)

# summary(rma_DistractorOverlap)


db_ET_MPo_MS = dat %>% filter(distractor_overlap == "onset" | distractor_overlap ==
```

```
    "novel" | distractor_overlap == "no")

db_ET_MPo_MS$distractor_overlap <- factor(db_ET_MPo_MS$distractor_overlap, levels = c("onset",
    "no", "novel"))


rma_DistractorOverlapMS <- rma.mv(g_calc, g_var_calc, mods = ~condition * distractor_overlap,
    data = db_ET_MPo_MS, random = ~collapse | short_cite)

# summary(rma_DistractorOverlapMS)

sum_eff1 <- coef(summary(rma_DistractorOverlapMS))[5, ]
sum_eff2 <- coef(summary(rma_DistractorOverlapMS))[6, ]
```

Regarding object identification in response to mispronunciations, when distractor overlap was included as a moderator, the moderator test was not significant QM(2) = 1.78, p = 0.41. This suggests that upon hearing a mispronunciation, infants looks to the target image were similar for when the target image was paired with a distractor image that contained overlap on the onset phoneme or no overlap with the target word, or was an unfamiliar object. We next assessed whether target-distractor overlap was related to mispronunciation sensitivity. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as an additional moderator. The moderator test was significant, QM(5) = 230.34, p < .0001. The interaction between condition and target-distractor pairs with no overlap was significant (β = -0.25, SE = 0.08, 95% CI [-0.4, -0.1], p < .01), suggesting that mispronunciation sensitivity was greater when target-distractor pairs shared the same onset phoneme compared to when they shared no phonological overlap. The interaction between condition and target-distractor pairs with an unfamiliar distractor was also significant (β = 0.19, SE = 0.09, 95% CI [0.01, 0.37], p = 0.04), suggesting that mispronunciation sensitivity was smaller when the distractor image shared the same onset phoneme as the target image compared to when the distractor image was an unfamiliar object. This can be seen in Figure X.

```
db_ET_MPo = dat %>% filter(distractor_overlap == "onset" | distractor_overlap ==
    "novel" | distractor_overlap == "no")

db_ET_MPo$distractor_overlap <- factor(db_ET_MPo$distractor_overlap, levels = c("onset",
    "no", "novel"))

db_ET_MPo$condition_label <- ifelse(db_ET_MPo$condition == 1, "Correct", "Mispronunciation")


p <- ggplot(db_ET_MPo, aes(condition_label, g_calc, fill = condition_label)) +
    facet_grid(. ~ distractor_overlap) + geom_boxplot() + # geom_smooth(method = 'lm', formula = y ~ lo
scale_fill_manual(values = cbPalette) + apatheme + theme(text = element_text(size = 25),
    legend.title = element_blank(), legend.position = "bottom", axis.title.x = element_blank(),
    axis.text.x = element_blank(), axis.ticks.x = element_blank()) + # xlab('Number of Features Changed
geom_hline(yintercept = 0, linetype = "dotted") + ylab("Effect size Hedges' g")

p
```




```
jpeg(filename = "figures/Figure_5_Distractor_overlap.jpg", width = 500, height = 300,
    units = "px")

p

dev.off()
```

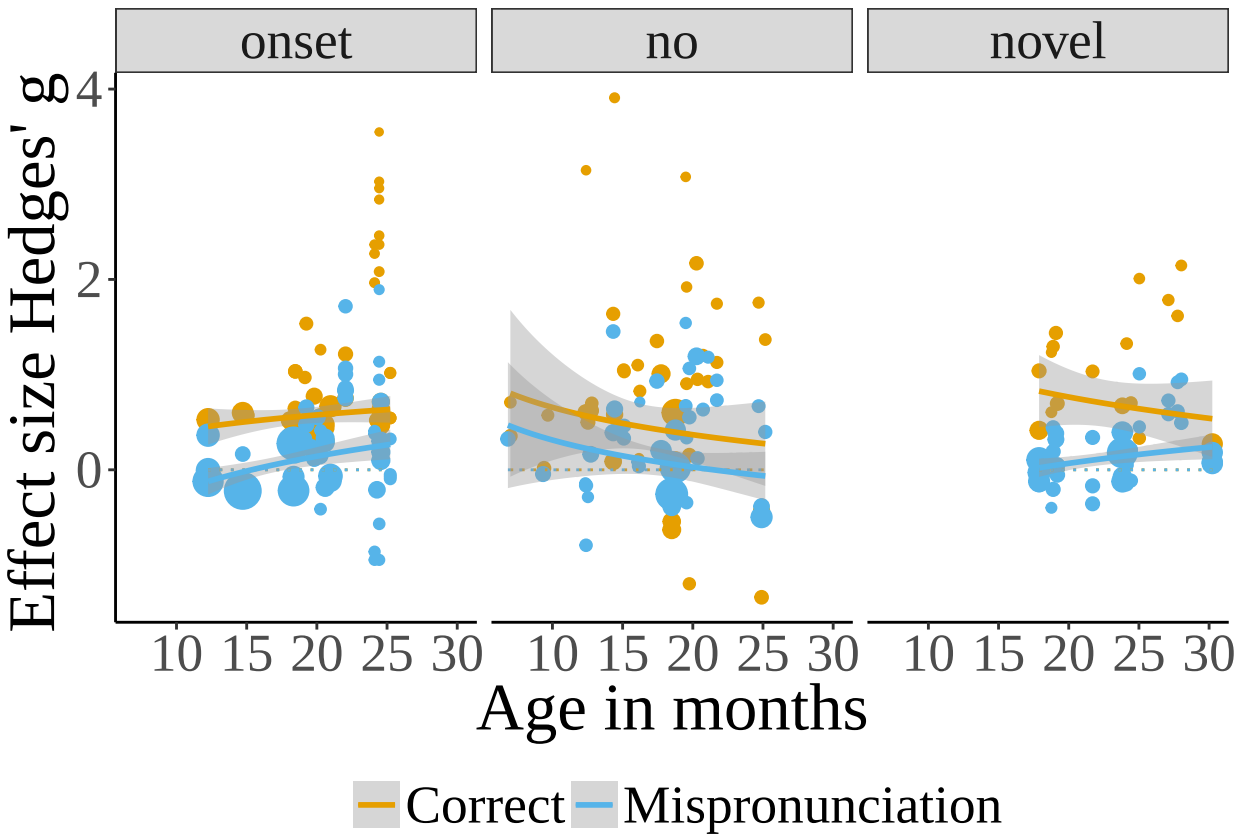
```
## pdf
##   2
```

```
# object recognition
db_ET_MPo = db_ET_MP %>% filter(distractor_overlap == "onset" | distractor_overlap ==
    "novel" | distractor_overlap == "no")

db_ET_MPo$distractor_overlap <- factor(db_ET_MPo$distractor_overlap, levels = c("onset",
    "no", "novel"))

rma_DistractorOverlap_age <- rma.mv(g_calc, g_var_calc, mods = ~age.C * distractor_overlap,
    data = db_ET_MPo, random = ~collapse | short_cite)


# mispronunciation sensitivity
dat_MPo = dat %>% filter(distractor_overlap == "onset" | distractor_overlap ==
    "novel" | distractor_overlap == "no")

dat_MPo$distractor_overlap <- factor(dat_MPo$distractor_overlap, levels = c("onset",
    "no", "novel"))
```

```r
rma_DistractorOverlap_age_MS <- rma.mv(g_calc, g_var_calc, mods = ~age.C * distractor_overlap *
    condition, data = dat_MPo, random = ~collapse | short_cite)

# summary(rma_DistractorOverlap_age_MS)

sum_eff1 <- coef(summary(rma_DistractorOverlap_age_MS))[11, ]
sum_eff2 <- coef(summary(rma_DistractorOverlap_age_MS))[12, ]
```

Although we did not have any specific predictions about the relationship between infant age and the impact of distractor overlap on mispronunciation sensitivity, we included an exploratory analysis to examine this relationship. First, for object recognition in response to mispronunciations, when age in addition to distractor overlap was also included as a moderator, the moderator test was not significant, QM(5) = 5.99, p = 0.31, suggesting that upon hearing a mispronunciation, infants looks to the target image were similar for the three types of overlap, regardless of infant age. We next assessed whether the relationship between distractor overlap and mispronunciation sensitivity was modulated by age. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as well as age as additional moderators. The moderator test was significant, QM(11) = 243.4, p < .0001. The interaction between age, condition, and target-distractor pairs with no overlap was significant (β = -0.04, SE = 0.02, 95% CI [-0.08, 0], p = 0.03). As can be seen in Figure X, the difference between correct pronunciations and mispronunciations (mispronunciation sensitivity) stays steady across infant ages for both target words paired with distractors containing onset overlap with the target word as well as distractors containing no overlap. As infants aged, however, overall recognition (regardless of condition) increased for target-distractor pairs containing onset overlap, whereas for overall recognition decreased for target-distractor pairs containing no overlap. The interaction between age, condition, and target-distractor pairs with an unfamiliar distractor was also significant (β = -0.05, SE = 0.02, 95% CI [-0.09, 0], p = 0.04). As can also be seen in Figure X, mispronunciation sensitivity for target words paired with unfamiliar distractors decreased with age, while it remained steady across a similar range of ages for target words paired with distractors containing onset overlap with the target word.

```r
db_ET_MPo = dat %>% filter(distractor_overlap == "onset" | distractor_overlap ==
    "novel" | distractor_overlap == "no")

db_ET_MPo$distractor_overlap <- factor(db_ET_MPo$distractor_overlap, levels = c("onset",
    "no", "novel"))

db_ET_MPo$condition_label <- ifelse(db_ET_MPo$condition == 1, "Correct", "Mispronunciation")


p <- ggplot(db_ET_MPo, aes(mean_age_1/30.44, g_calc, color = condition_label)) +
    facet_grid(. ~ distractor_overlap) + geom_point(aes(size = weights_g), show.legend = FALSE) +
    geom_line(y = 0, linetype = "dotted") + geom_smooth(method = "lm", formula = y ~
    log(x), aes(weight = weights_g)) + scale_colour_manual(values = cbPalette) +
    apatheme + theme(legend.title = element_blank(), legend.position = "bottom") +
    xlab("Age in months") + ylab("Effect size Hedges' g")

p
```




```
jpeg(filename = "figures/Figure_6_Distractor_overlap_age.jpg", width = 500,
    height = 300, units = "px")

p

dev.off()
```

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

Position of mispronunciation

```
misp_location_info <- dat %>% group_by(mispron_location) %>% summarize(n_exp_conditions = n())
```

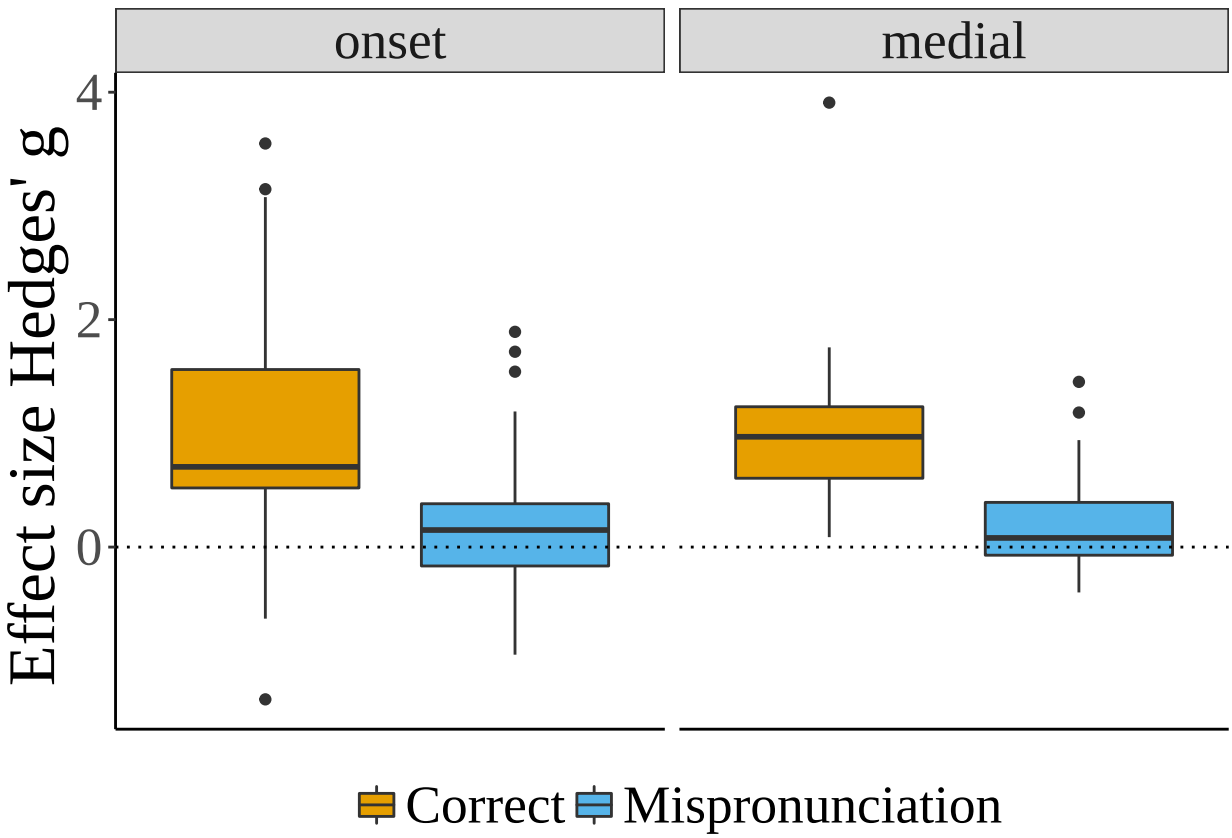
To assess whether the position of the mispronunciation has an impact on mispronunciation sensitivity, we calculated the meta-analytic effect for object identification in response to mispronunciations on the onset and medial phonemes. We did not include data for which the mispronunciation was located on the coda (n = 10, NA), varied in regard to position (n = 3, 29, 8, NA), or was not reported (n = 10). The analysis was therefore based on a subset of the overall dataset, with 143, NA experimental conditions comparing a mispronunciation on the onset phoneme and 48, NA experimental conditions comparing a mispronunciation on the medial phoneme.

```
# table(db_ET_MP$mispron_location)

db_ET_MPl = db_ET_MP %>% filter(mispron_location == "onset" | mispron_location ==
    "medial")
```

```
db_ET_MPl$mispron_location <- factor(db_ET_MPl$mispron_location, levels = c("onset",
    "medial"))

# rma_NFeatures <- rma.mv(g_calc, g_var_calc, mods = ~as.ordered(n_feature),
# data = db_ET_MP, random = ~collapse | short_cite)
rma_Location_MP <- rma.mv(g_calc, g_var_calc, mods = ~mispron_location, data = db_ET_MPl,
    random = ~collapse | short_cite)


db_ET_MPl = dat %>% filter(mispron_location == "onset" | mispron_location ==
    "medial")

db_ET_MPl$mispron_location <- factor(db_ET_MPl$mispron_location, levels = c("onset",
    "medial"))

rma_LocationCondition <- rma.mv(g_calc, g_var_calc, mods = ~mispron_location *
    condition, data = db_ET_MPl, random = ~collapse | short_cite)

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

Regarding object identification in response to mispronunciations, when mispronunciation location was included as a moderator, the moderator test was not significant QM(1) = 0.04, p = 0.84. This suggests that upon hearing a mispronunciation, infants looks to the target image were similar for when the mispronunciation was located on the onset or medial phonemes. We next assessed whether mispronunciation location was related to mispronunciation sensitivity. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as an additional moderator. The moderator test was significant, QM(3) = 180.76, p < .0001, but the interaction between mispronunciation location and condition was not significant $\beta = 0.05$, (SE = 0.09, 95% CI [-0.12, 0.22], p = 0.56). These results suggest that overall, the location of the mispronunciation (onset, medial) did not impact mispronunciation sensitivity.

```
# dat.p <- subset(dat, mispron_location == 'onset' | mispron_location ==
# 'medial' | mispron_location == 'offset')

dat$condition_label = ifelse(dat$condition == 1, "Correct", "Mispronunciation")

dat.p = dat %>% filter(mispron_location == "onset" | mispron_location == "medial")

dat.p$mispron_location <- factor(dat.p$mispron_location, levels = c("onset",
    "medial"))


p <- ggplot(dat.p, aes(condition_label, g_calc, fill = condition_label)) + facet_grid(. ~
    mispron_location) + geom_boxplot() + # geom_smooth(method = 'lm', formula = y ~ log(x), aes(weight=
scale_fill_manual(values = cbPalette) + apatheme + theme(text = element_text(size = 25),
    legend.title = element_blank(), legend.position = "bottom", axis.title.x = element_blank(),
    axis.text.x = element_blank(), axis.ticks.x = element_blank()) + # xlab('Number of Features Changed
geom_hline(yintercept = 0, linetype = "dotted") + ylab("Effect size Hedges' g")

p
```




```
jpeg(filename = "figures/Figure_7_Mispronunciation_position.jpg", width = 500,
    height = 300, units = "px")

p

dev.off()
```

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

```
# table(db_ET_MP$mispron_location)

db_ET_MPl = db_ET_MP %>% filter(mispron_location == "onset" | mispron_location ==
    "medial")

db_ET_MPl$mispron_location <- factor(db_ET_MPl$mispron_location, levels = c("onset",
    "medial"))

# rma_NFeatures <- rma.mv(g_calc, g_var_calc, mods = ~as.ordered(n_feature),
# data = db_ET_MP, random = ~collapse | short_cite)
rma_Location_MP_age <- rma.mv(g_calc, g_var_calc, mods = ~mispron_location *
    age.C, data = db_ET_MPl, random = ~collapse | short_cite)


db_ET_MPl = dat %>% filter(mispron_location == "onset" | mispron_location ==
    "medial")

db_ET_MPl$mispron_location <- factor(db_ET_MPl$mispron_location, levels = c("onset",
```

```
    "medial"))

rma_LocationCondition_age <- rma.mv(g_calc, g_var_calc, mods = ~mispron_location *
    condition * age.C, data = db_ET_MPl, random = ~collapse | short_cite)

sum_eff <- coef(summary(rma_LocationCondition))[8, ]
```

Although we did not have any specific predictions about the relationship between infant age and the impact of mispronuncition location on mispronunciation sensitivity, we included an exploratory analysis to examine this relationship. First, for object recognition in response to mispronunciations, when age in addition to mispronunciation location was also included as a moderator, the moderator test was not significant, QM(3) = 1.22, p = 0.75, suggesting that upon hearing a mispronunciation, infants looks to the target image were similar for both onset and medial mispronunciations, regardless of infant age. We next assessed whether the relationship between mispronunciation location and mispronunciation sensitivity was modulated by age. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as well as age as additional moderators. The moderator test was significant, QM(7) = 185.34, p < .0001, but the interaction between mispronunciation location, age, and condition was not significant β = NA, (SE = NA, 95% CI [NA, NA], p NA). These results provide further evidence that location of the mispronunciation (onset, medial) did not impact mispronunciation sensitivity.

Type of mispronunciation (consonant or vowel)

```
CV_info <- dat %>% group_by(type_feature) %>% summarize(n_exp_conditions = n())

CV_location_info <- dat %>% group_by(type_feature, mispron_location) %>% summarize(n_exp_conditions = n
```

To assess whether the type of mispronunciation impacts sensitivity to mispronunciations, we calculated the meta-analytic effect for object identification in response to the type of mispronunciation. Although most theoretical discussion of mispronunciation type has focused on consonants and vowels, our dataset also included tone mispronunciations. In our analysis, we were interested in the difference between consonants and vowels, but also include an exploratory analysis of responses to tones, consonants, and vowels. We therefore conducted two sets of analyses, one analyzing consonants and vowels alone and a second comparing responses to tones with that of consonants and vowels, separately. For the latter analysis, tones were coded as the reference condition. We did not include data for which mispronunciation type varied within experiment and was not reported separately (n = 21, 2). The analysis was therefore based on a subset of the overall dataset, with 145 experimental conditions comparing a consonant mispronunciation, 71 experimental conditions comparing a vowel mispronunciation, and 12 experimental conditions comparing a tone mispronunciation. Below, we first report the set of analyses comparing consonants with vowels before moving on to the second set of exploratory analyses comparing tones with that of consonants and vowels.

[KATIE] WHAT DO YOU THINK ABOUT THIS? WE HAVE THE TONES AND ITS A NOVEL, INTERESTING THING, I THINK, AND PERHAPS WORTH IT TO INCLUDE A COMPARISON OF TONES ALONGSIDE THE MORE THEORETICALLY IMPORTANT COMPARISON BETWEEN CONSONANTS AND VOWELS.

```
# C vs. V
db_MP_type <- subset(db_ET_MP, type_feature == "consonant" | type_feature ==
    "vowel")

db_MP_type$type_feature <- factor(db_MP_type$type_feature, levels = c("consonant",
    "vowel"))

rma_TypeFeaturesMP <- rma.mv(g_calc, g_var_calc, mods = ~type_feature, data = db_MP_type,
    random = ~collapse | short_cite)
```

```r
# C vs. V with condition

db_type <- subset(dat, type_feature == "consonant" | type_feature == "vowel")

db_type$type_feature <- factor(db_type$type_feature, levels = c("consonant",
    "vowel"))

rma_TypeFeaturesMP_Condition <- rma.mv(g_calc, g_var_calc, mods = ~type_feature *
    condition, data = db_type, random = ~collapse | short_cite)

sum_eff1 <- coef(summary(rma_TypeFeaturesMP_Condition))[4, ]
```

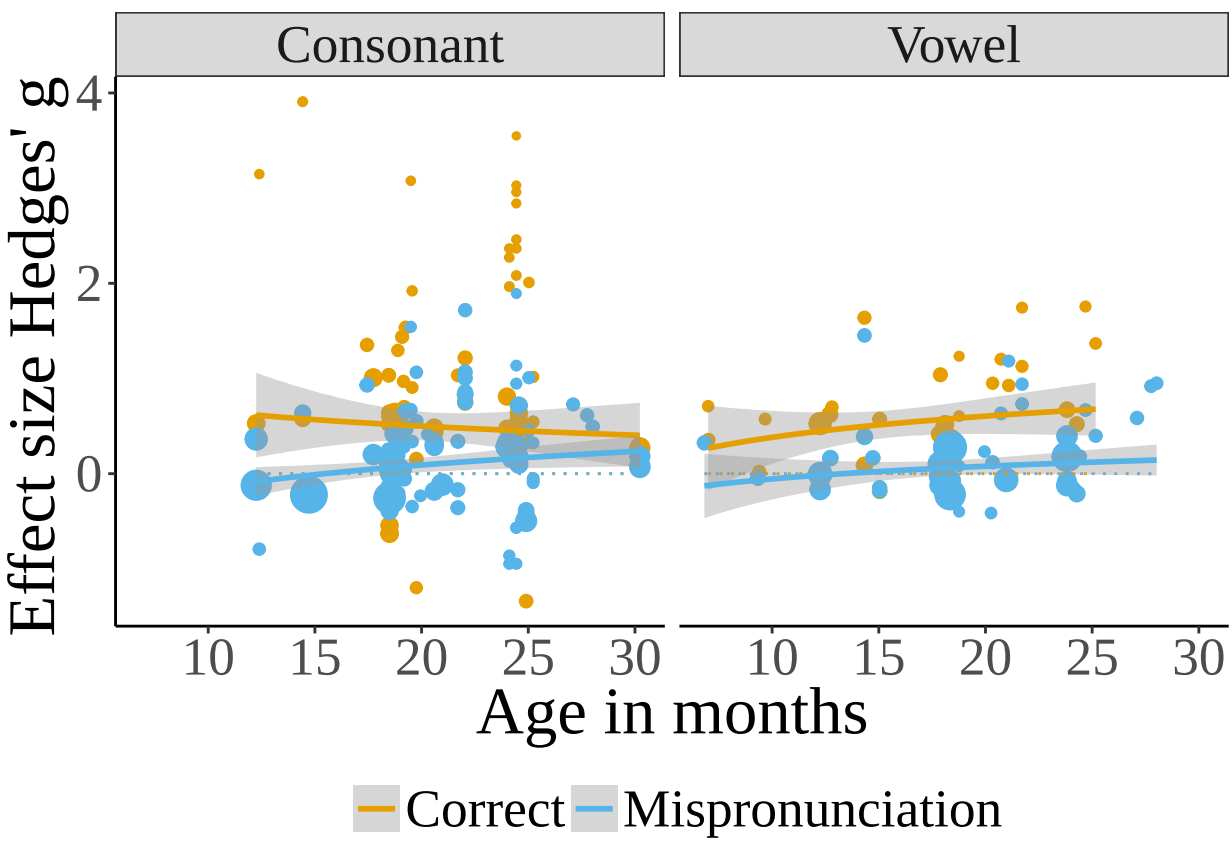
We first analyzed experimental conditions where mispronunciation type was either a consonant or vowel. Regarding object identification in response to mispronunciations, when mispronunciation type was included as a moderator, the moderator test was not significant QM(1) = 0.15, p = 0.7. This suggests that upon hearing a mispronunciation, infants looks to the target image were similar for when the mispronunciation was a consonant or a vowel. We next assessed whether type of mispronunciation (consonant or vowel) was related to mispronunciation sensitivity. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as an additional moderator. The moderator test was significant, QM(3) = 145.14, p < .0001, but the interaction between mispronunciation type and condition was not significant $\beta = 0.06$, (SE = 0.08, 95% CI [-0.1, 0.21], p = 0.48). These results suggest that overall, the type of mispronunciation (consonant vs. vowel) did not impact mispronunciation sensitivity.

```r
db_MP_type <- subset(db_ET_MP, type_feature == "consonant" | type_feature ==
    "vowel")

db_MP_type$type_feature <- factor(db_MP_type$type_feature, levels = c("consonant",
    "vowel"))

rma_TypeFeaturesMPage <- rma.mv(g_calc, g_var_calc, mods = ~type_feature * age.C,
    data = db_MP_type, random = ~collapse | short_cite)



# with condition
db_type <- subset(dat, type_feature == "consonant" | type_feature == "vowel")

db_type$type_feature <- factor(db_type$type_feature, levels = c("consonant",
    "vowel"))

rma_TypeFeaturesMP_Condition <- rma.mv(g_calc, g_var_calc, mods = ~type_feature *
    condition * age.C, data = db_type, random = ~collapse | short_cite)

sum_eff <- coef(summary(rma_TypeFeaturesMP_Condition))[8, ]
```

We next examined whether age modulates object recognition or mispronunciation sensitivity when the mispronunciation is a consonant or vowel. For object recognition in response to a mispronunciation, including age as a moderator resulted in a moderator test that was not significant, QM(3) = 1.54, p = 0.67. This suggests that upon hearing a mispronunciation, infants looks to the target image were similar for when the mispronunciation was on a consonant or vowel phoneme, regardless of their age. We next assessed whether the relationship between mispronunciation type (consonant or vowel) and mispronunciation sensitivity was modulated by age. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as well as age as additional moderators. The moderator test was significant, QM(7) = 153.8, p < .0001. The interaction between mispronunciation type, condition, and age was significant ($\beta = 0.04$, SE = 0.02, 95% CI [0.01,

0.08], p = 0.02). As can be seen in Figure X, as infants age, mispronunciation sensitivity grows larger for vowel mispronunciations but becomes smaller for consonant mispronunciations. Noticeably, mispronunciation sensitivity appears greater for consonant compared to vowel mispronunciations at younger ages, but this difference shifts as infants age.

```
dat_type <- subset(dat, type_feature == "consonant" | type_feature == "vowel")
dat_type$type_feature <- ifelse(dat_type$type_feature == "consonant", "Consonant",
    "Vowel")

# db_MP_type <- subset(db_ET_MP, type_feature != 'consonant_and_vowel')

# dat_type_sub <- subset(dat_type, lang_family != 'Sino-Tibetian')


dat_type$condition_label = ifelse(dat_type$condition == 1, "Correct", "Mispronunciation")

p <- ggplot(dat_type, aes(mean_age_1/30.44, g_calc, color = condition_label)) +
    facet_grid(. ~ type_feature) + geom_point(aes(size = weights_g), show.legend = FALSE) +
    geom_line(y = 0, linetype = "dotted") + geom_smooth(method = "lm", formula = y ~
    log(x), aes(weight = weights_g)) + scale_colour_manual(values = cbPalette) +
    apatheme + theme(legend.title = element_blank(), legend.position = "bottom") +
    xlab("Age in months") + ylab("Effect size Hedges' g")

p
```




```
jpeg(filename = "figures/Figure_8_FeatureType_Cond_Age.jpg", width = 500, height = 300,
    units = "px")
```

```r
p

dev.off()
```

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

```r
family <- dat %>% group_by(lang_family) %>% summarize(count = n())

lang_family <- dat %>% group_by(native_lang, lang_family) %>% summarize(count = n())

cv_family <- dat %>% group_by(type_feature, lang_family) %>% summarize(count = n())
```

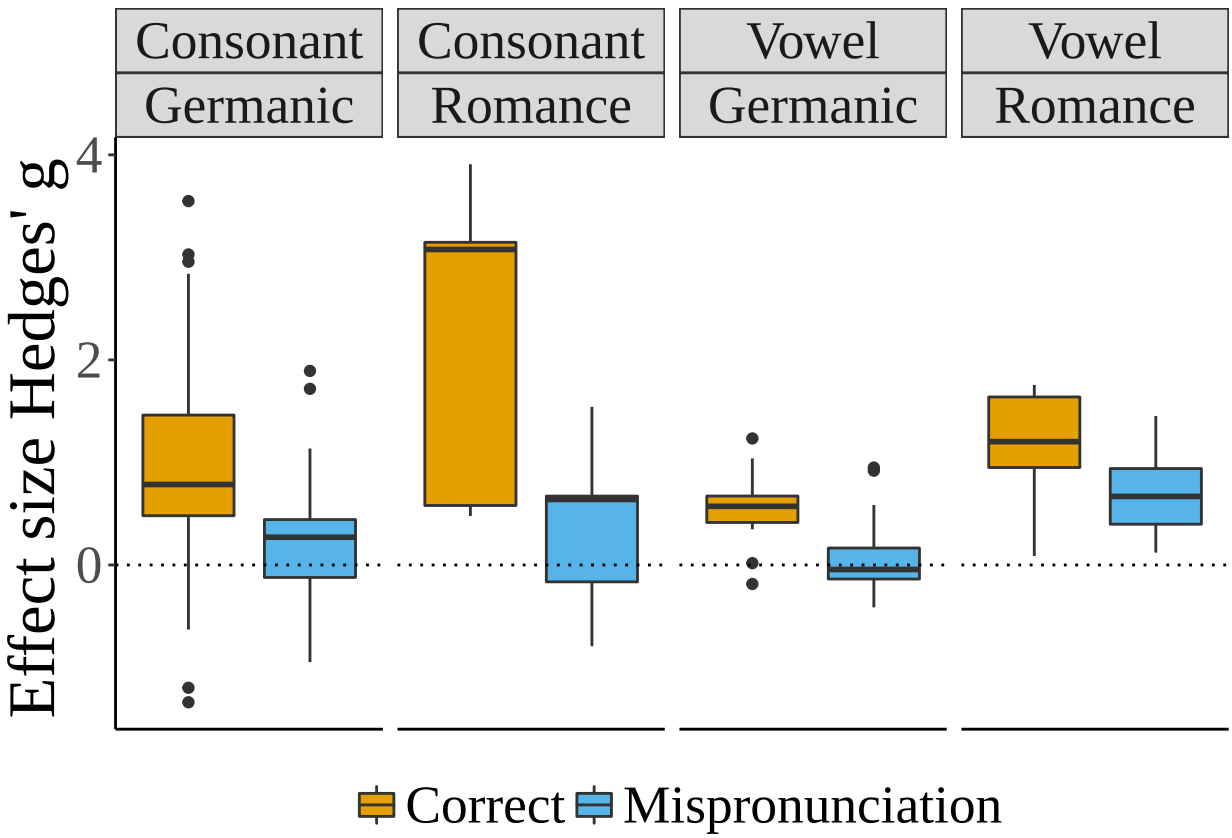
```r
db_MP_type <- subset(db_ET_MP, type_feature == "consonant" | type_feature ==
    "vowel")

db_MP_type$type_feature <- factor(db_MP_type$type_feature, levels = c("consonant",
    "vowel"))

db_MP_type <- subset(db_MP_type, lang_family != "Sino-Tibetian")


rma_TypeFeaturesMPfam <- rma.mv(g_calc, g_var_calc, mods = ~type_feature * lang_family,
    data = db_MP_type, random = ~collapse | short_cite)



# with condition
db_type <- subset(dat, type_feature == "consonant" | type_feature == "vowel")

db_type$type_feature <- factor(db_type$type_feature, levels = c("consonant",
    "vowel"))

db_type <- subset(db_type, lang_family != "Sino-Tibetian")

rma_TypeFeaturesMPC <- rma.mv(g_calc, g_var_calc, mods = ~type_feature * condition *
    lang_family, data = db_type, random = ~collapse | short_cite)

sum_eff1 <- coef(summary(rma_TypeFeaturesMPC))[7, ]
sum_eff2 <- coef(summary(rma_TypeFeaturesMPC))[8, ]

# summary(glht(rma_TypeFeaturesMPC, linfct=cbind(contrMat(rep(1,8),
# type='Tukey'))), test=adjusted('none')) ranef(rma_TypeFeaturesMPC)
```

To examine whether infants' native language impacts sensitivity to consonant and vowel mispronunciations, we classified infants into language families. Infants learning American English (n = 56), British English (n = 66), Danish (n = 6), Dutch (n = 58), and German (n = 21) were classified into the Germanic language family (n = 207). Infants learning Catalan (n = 4), Spanish (n = 4), French (n = 8), Catalan and Spanish simultaneously (i.e. bilinguals; n = 6), and Swiss French (n = 6) were classified into the Romance language family (n = 28). For object recognition in response to a mispronunciation, including language family as a moderator resulted in a moderator test that was not significant, QM(3) = 5.36, p = 0.15. This suggests that upon hearing a mispronunciation, infants looks to the target image were similar for when the mispronunciation was on a consonant or vowel phoneme, regardless of the language family of their native language. We next assessed whether the relationship between mispronunciation type (consonant or vowel) and mispronunciation sensitivity was modulated by language family. We merged the two datasets and included condition (correct

pronunciation, mispronunciation) as well as language family as additional moderators. The moderator test was significant, QM(7) = 158.89, p < .0001. The interaction between condition and language family was significant (β = 0.73, SE = 0.23, 95% CI [0.27, 1.18], p < .01), suggesting XXXXXXXXXXXXXX. The interaction between mispronunciation type, condition, and language family was also significant (β = -0.87, SE = 0.28, 95% CI [-1.42, -0.32], p < .01). As can be seen in Figure X, mispronunciation sensitivity for consonants was similar for Germanic and Romance languages. Mispronunciation sensitivity for vowels, however, was greater for Germanic compared to Romance languages.

[KATIE] I'M NOT REALLY SURE WHAT THE CONDITION BY LANGUAGE FAMILY INTERACTION MEANS. SHOULD WE EVEN INTERPRET IT?

```
dat_type <- subset(dat, type_feature == "consonant" | type_feature == "vowel")
dat_type$type_feature <- ifelse(dat_type$type_feature == "consonant", "Consonant",
    "Vowel")

# db_MP_type <- subset(db_ET_MP, type_feature != 'consonant_and_vowel')

dat_type_sub <- subset(dat_type, lang_family != "Sino-Tibetian")


dat_type_sub$condition_label = ifelse(dat_type_sub$condition == 1, "Correct",
    "Mispronunciation")


p <- ggplot(dat_type_sub, aes(condition_label, g_calc, fill = condition_label)) +
    geom_boxplot() + facet_grid(. ~ type_feature * lang_family) + # geom_line(y= 0, linetype='dotted') +
# y ~ log(x), aes(weight=weights_g)) +
scale_fill_manual(values = cbPalette) + apatheme + theme(text = element_text(size = 25),
    legend.title = element_blank(), legend.position = "bottom", axis.title.x = element_blank(),
    axis.ticks.x = element_blank(), axis.text.x = element_blank()) + # xlab('Number of Features Changed
geom_hline(yintercept = 0, linetype = "dotted") + xlab("Language Family") +
    ylab("Effect size Hedges' g")
p
```




```
jpeg(filename = "figures/Figure_9_FeatureType_Cond_LangFam.jpg", width = 500,
    height = 300, units = "px")

p

dev.off()
```

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

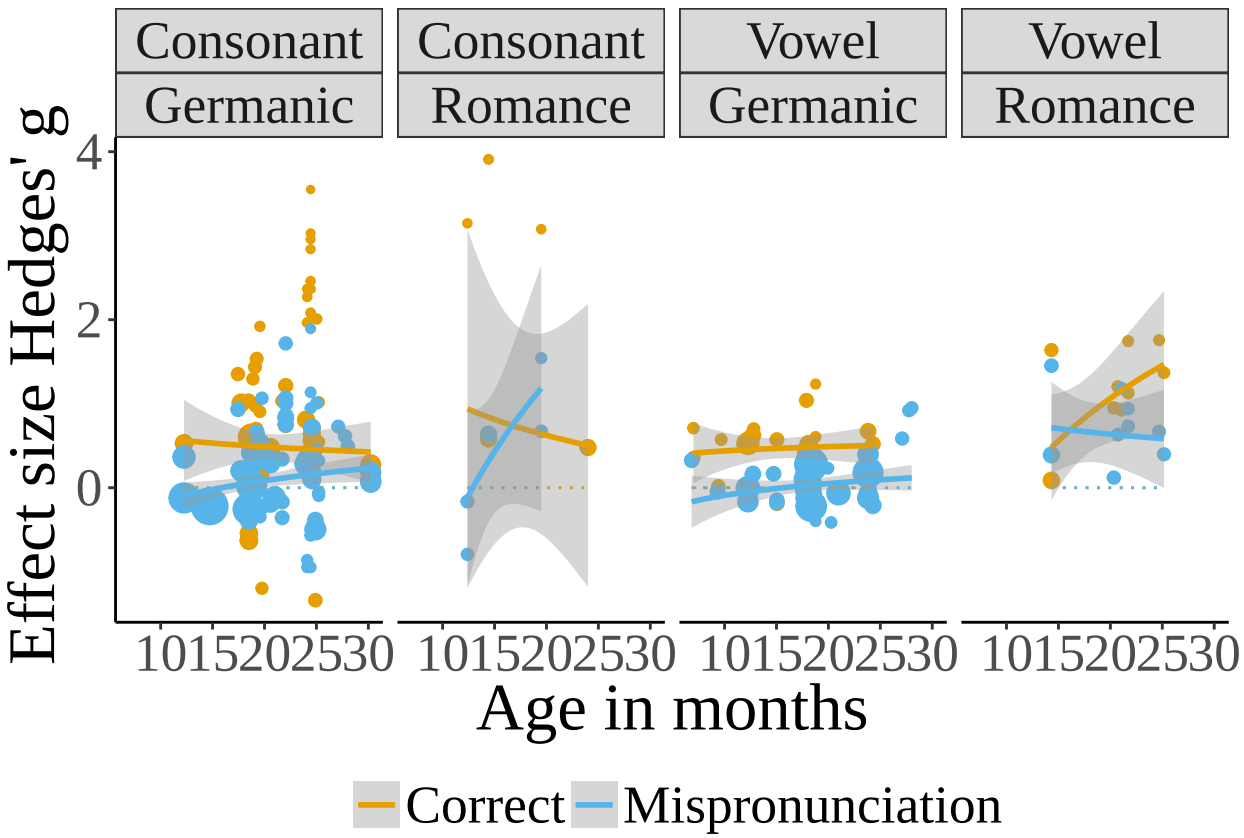
```
db_MP_type <- subset(db_ET_MP, type_feature == "consonant" | type_feature ==
    "vowel")

db_MP_type$type_feature <- factor(db_MP_type$type_feature, levels = c("consonant",
    "vowel"))

db_MP_type <- subset(db_MP_type, lang_family != "Sino-Tibetian")


rma_TypeFeaturesMPfam <- rma.mv(g_calc, g_var_calc, mods = ~type_feature * lang_family *
    age.C, data = db_MP_type, random = ~collapse | short_cite)

sum_eff1 <- coef(summary(rma_TypeFeaturesMPC))[8, ]


# with condition
db_type <- subset(dat, type_feature == "consonant" | type_feature == "vowel")
```

```r
db_type$type_feature <- factor(db_type$type_feature, levels = c("consonant",
    "vowel"))

db_type <- subset(db_type, lang_family != "Sino-Tibetian")

rma_TypeFeaturesMPC <- rma.mv(g_calc, g_var_calc, mods = ~type_feature * condition *
    lang_family, data = db_type, random = ~collapse | short_cite)

sum_eff1 <- coef(summary(rma_TypeFeaturesMPC))[7, ]
sum_eff2 <- coef(summary(rma_TypeFeaturesMPC))[8, ]

# summary(glht(rma_TypeFeaturesMPC, linfct=cbind(contrMat(rep(1,8),
# type='Tukey'))), test=adjusted('none')) ranef(rma_TypeFeaturesMPC)
```

Finally, we examined the relationship between language family and infant age and mispronunciation sensitivity to consonants and vowels. For object recognition in response to a mispronunciation, including language family and infant age as a moderator resulted in a moderator test that was significant, QM(7) = 19.53, p < .01. The interaction between language family, age, and type of mispronunciation (consonant or vowel) was significant, ($\beta = 0.73$, SE = 0.23, 95% CI [0.27, 1.18], p < .01). As can be seen in Figure X, for infants learning Germanic languages, increasing age was related to increasing mispronunciation sensitivity for vowel mispronunciations, but decreasing sensitivity for consonant mispronunciations. In contrast, infants learning Romance languages have an even greater increase with age in sensitivity to vowel mispronunciations. Surprisingly, sensitivity to consonant mispronunciations shows a reversal in infants learning Romance languages: the growth in target looks for consonant mipronunciations increases and surpases that of target looks for correct pronunciations.

[KATIE] AGAIN, THERE ARE ADDITIONAL INTERACTIONS THAT ARE SIGNIFICANT... SHOULD THEY BE INTERPRETED? ALSO, WTF ROMANCE LANGUAGES?

```r
dat_type <- subset(dat, type_feature == "consonant" | type_feature == "vowel")
dat_type$type_feature <- ifelse(dat_type$type_feature == "consonant", "Consonant",
    "Vowel")

# db_MP_type <- subset(db_ET_MP, type_feature != 'consonant_and_vowel')

dat_type_sub <- subset(dat_type, lang_family != "Sino-Tibetian")


dat_type_sub$condition_label = ifelse(dat_type_sub$condition == 1, "Correct",
    "Mispronunciation")

p <- ggplot(dat_type_sub, aes(mean_age_1/30.44, g_calc, color = condition_label)) +
    facet_grid(. ~ type_feature * lang_family) + geom_point(aes(size = weights_g),
    show.legend = FALSE) + geom_line(y = 0, linetype = "dotted") + geom_smooth(method = "lm",
    formula = y ~ log(x), aes(weight = weights_g)) + scale_colour_manual(values = cbPalette) +
    apatheme + theme(legend.title = element_blank(), legend.position = "bottom") +
    xlab("Age in months") + ylab("Effect size Hedges' g")

p
```




```
jpeg(filename = "figures/Figure_10_FeatureType_Cond_Age_LangFam.jpg", width = 500,
    height = 300, units = "px")

p

dev.off()
```

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

```
# C vs. V vs. T
db_MP_type <- subset(db_ET_MP, type_feature == "consonant" | type_feature ==
    "vowel" | type_feature == "tone")

db_MP_type$type_feature <- factor(db_MP_type$type_feature, levels = c("tone",
    "consonant", "vowel"))

rma_TypeFeaturesMPT <- rma.mv(g_calc, g_var_calc, mods = ~type_feature, data = db_MP_type,
    random = ~collapse | short_cite)

# C vs. V vs. T with condition

db_type <- subset(dat, type_feature == "consonant" | type_feature == "vowel" |
    type_feature == "tone")

db_type$type_feature <- factor(db_type$type_feature, levels = c("tone", "consonant",
    "vowel"))
```

```r
rma_TypeFeaturesMP_ConditionT <- rma.mv(g_calc, g_var_calc, mods = ~type_feature *
    condition, data = db_type, random = ~collapse | short_cite)

sum_eff2c <- coef(summary(rma_TypeFeaturesMP_ConditionT))[5, ]
sum_eff2v <- coef(summary(rma_TypeFeaturesMP_ConditionT))[6, ]
```

Although we had no predictions regarding mispronunciation sensitivity to tone mispronunciations, we included an exploratory analysis to examine whether responses to tone mispronunciations were different from that of consonants or vowels. Regarding object identification in response to mispronunciations, when mispronunciation type was included as a moderator, the moderator test was not significant QM(2) = 1.16, p = 0.56. This suggests that upon hearing a mispronunciation, infants looks to the target image were similar for tone mispronunciations in comparison with both consonants and vowels. We next assessed whether type of mispronunciation (tone, consonant, vowel) was related to mispronunciation sensitivity. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as an additional moderator. The moderator test was significant, QM(5) = 154.88, p < .0001. The interaction between condition and consonant mispronunciations was not significant β = -0.19, (SE = 0.21, 95% CI [-0.59, 0.21], p = 0.36), suggesting that there was no difference in looks to the target in response to consonant and tone mispronunciations. The interaction between condition and vowel mispronunciations was also not significant β = -0.13, (SE = 0.21, 95% CI [-0.55, 0.28], p = 0.53), suggesting that there was no difference in looks to the target in response to vowel and tone mispronunciations.

```r
db_MP_type <- subset(db_ET_MP, type_feature == "consonant" | type_feature ==
    "vowel" | type_feature == "tone")

db_MP_type$type_feature <- factor(db_MP_type$type_feature, levels = c("tone",
    "consonant", "vowel"))

rma_TypeFeaturesMPage <- rma.mv(g_calc, g_var_calc, mods = ~type_feature * age.C,
    data = db_MP_type, random = ~collapse | short_cite)



# with condition
db_type <- subset(dat, type_feature == "consonant" | type_feature == "vowel" |
    type_feature == "tone")

db_type$type_feature <- factor(db_type$type_feature, levels = c("tone", "consonant",
    "vowel"))

rma_TypeFeaturesMP_Condition <- rma.mv(g_calc, g_var_calc, mods = ~type_feature *
    condition * age.C, data = db_type, random = ~collapse | short_cite)

sum_eff1 <- coef(summary(rma_TypeFeaturesMP_Condition))[11, ]
sum_eff2 <- coef(summary(rma_TypeFeaturesMP_Condition))[12, ]
```

We further included an exploratory analysis of the relationship between infant age and the impact of tone mispronunciations in comparison to consonant and vowel mispronunciations. First, for object recognition in response to mispronunciations, when age in addition to mispronunciation location was also included as a moderator, the moderator test was not significant, QM(5) = 2.78, p = 0.73. This suggests that upon hearing a mispronunciation, infants looks to the target image were not different between tone and vowel or tone and consonant mispronunciations, regardless of their age. We next assessed whether the relationship between mispronunciation type (tone, consonant, vowel) and mispronunciation sensitivity was modulated by age. We merged the two datasets and included condition (correct pronunciation, mispronunciation) as well as age as additional moderators. The moderator test was significant, QM(11) = 163.85, p < .0001, but the interactions

between condition, age, and both consonant mispronciations (β = 0.02, SE = 0.1, 95% CI [-0.19, 0.22], p = 0.87) and vowel mispronunciations (β = 0.06, SE = 0.1, 95% CI [-0.14, 0.27], p = 0.56) were not significant. Infants' sensitivity to tone mispronunciations compared to consonant or vowel mispronunciations did not differ with age.

[KATIE] WORTH IT TO INCLUDE LANGUAGE FAMILY ANALYSES TOO?

Discussion

To Summarize:

** Overall Meta-analytic Effect **

- Accept mispronunciations as labels for targets
- Sensitive to mispronunciations
- lack of change over development

** Vocabulary **

- no relationship?
- talk about how few studies report it

** Size of Mispronunciation **

- graded sensitivity to number of features changed in a mispronunciation
- importance for controlling in experimental design
- Perhaps a call for more studies to include multiple number of features changed, so that this can be assessed? There was a narrow age where this was actually manipulated.

** Distractor Familiarity **

- Not really sure, check the results. A key interaction is significant in one model but not the other.
- Again, ages not matched very well for the two groups here. Decide about whether to include a age subset analysis

** Phonological overlap between target and distractor **

- mispronunciation sensitivity was greater when target-distractor pairs shared the same onset phoneme compared to when they shared no phonological overlap
- this is rather the opposite of what one would expect, right?
- mispronunciation sensitivity was smaller when the distractor image shared the same onset phoneme as the target image compared to when the distractor image was an unfamiliar object
- Maybe it would be useful to have time course analyses to address this issue further
- As infants aged, overall recognition (regardless of condition) increased for target-distractor pairs containing onset overlap, whereas overall recognition decreased for target-distractor pairs containing no overlap.
- mispronunciation sensitivity for target words paired with unfamiliar distractors decreased with age, while it remained steady across a similar range of ages for target words paired with distractors containing onset overlap with the target word.

** Position of mispronunciation **

- really no impact at all

** Type of mispronunciation **

- Overall, no difference between consonants and vowels
- Consonant mispronunciation sensitivity decreases with age
- Vowel mispronunciation sensitivity increases with age
- mispronunciation sensitivity for consonants similar for Germanic and Romance languages

- Mispronunciation sensitivity for vowels greater for Germanic compared to Romance languages
- For Germanic infants, increasing age was related to increasing mispronunciation sensitivity for vowel mispronunciations, but decreasing sensitivity for consonant mispronunciations.
- For Romance infants, an even greater increase with age in sensitivity to vowel mispronunciations.
- For Romance infants, the growth in target looks for consonant mipronunciations increases and surpases that of target looks for correct pronunciations
- exploratory analyses with tone mispronunciations suggest no great difference in sensitivity when compared to consonant and vowel mispronunciations

When it comes to designing studies, best practices and current standards might not always overlap. Indeed, across a set of previous meta-analyses it was shown that particularly infant research does not adjust sample sizes according to the effect in question (Bergmann et al., in press). A meta-analysis is a first step in improving experiment planning by measuring the underlying effect and its variance, which is directly related to the sample needed to achieve satisfactory power in the null hypothesis significance testing framework. Failing to take effect sizes into account can both yield to underpowered research and to testing too many participants, both consequences are undesirable for a number of reasons that have been discussed in depth elsewhere. We will just briefly mention two that we consider most salient for theory building: Underpowered studies will lead to false negatives more frequently than expected, which in turn results in an unpublished body of literature (citationcitation). Overpowered studies mean that participants were tested unnecessarily, which has substantial ethical consequences particularly when working with infants and other difficult to recruit and test populations.

From Christina: let's make a note to put sth in the discussion about our curve being surprisingly flat for correctly pronounced words bc people adapt their analysis windows? Bc if you look at Molly's reaction time paper, there is a steep increase.

Discussing the Moderator Analyses Maybe put them together into the ones that worked out as we predicted and those that didn't? So, here is evidence that supports existing arguments, that doesn't need to be a huge chunk. But then more space devoted to moderator analyses that didn't work out according to predictions.

It should be noted that the majority of consonant mispronunciations were located on the onset phoneme (n = 120; total consonant conditions, n = 145), while the majority of vowel mispronunciations were located on the medial phoneme (n = 44; total vowel conditions, n = 71). In their analysis using TRACE, Mayor and Plunkett (2014) found that the difference between sensitivity to consonant and vowel mispronunciations was due to infants' lexical knowlege consisting of a majority consonant onset words.

[KATIE] COME BACK TO THIS.