

Resubmit Bin Analysis

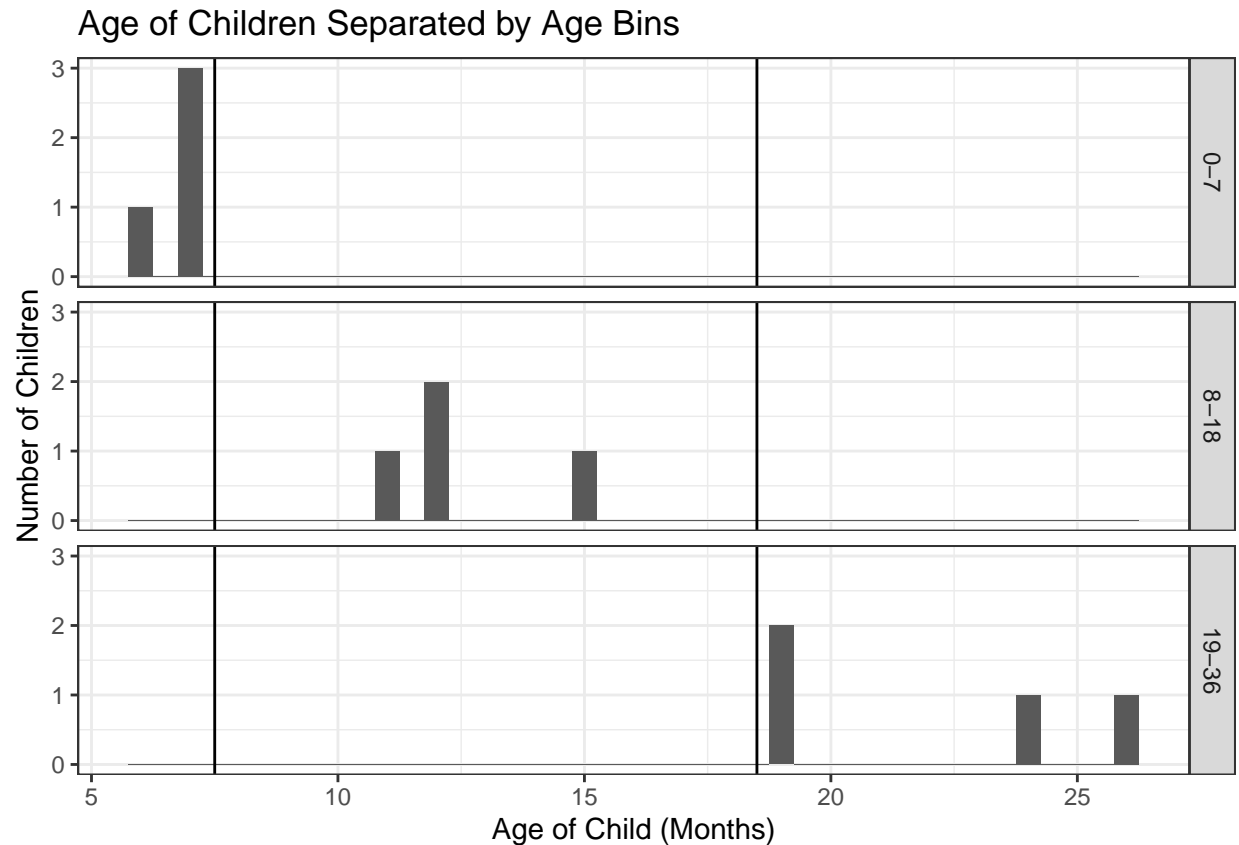
This document responds to some questions raised by reviewers during the review process. You will need to get the private metadata csv file from the BabbleCor authors (and our data) in order to reproduce these results.

Firstly, the reviewers expressed interest in whether babies in the age condition were more likely to be misclassified in the neighboring bin (and therefore “close” in their incorrect guesses). We examined this graphically.

Firstly we look at the distribution of ages within each of the bins used.

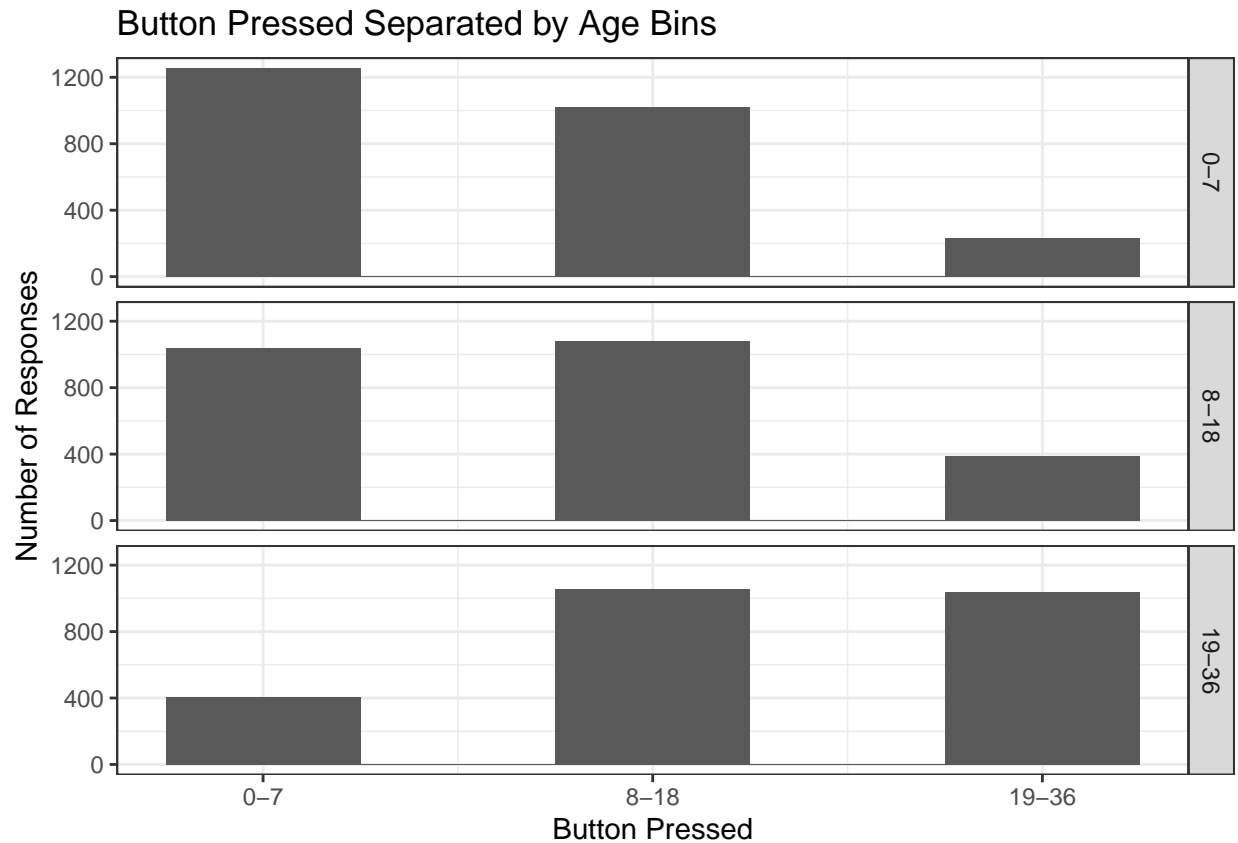
```
# subset to one subjects data to see actual number of babies used in stimuli
oneSub<-subset(d,d$subject_ID==d$subject_ID[1])

# plot age by bin
ggplot(subset(oneSub,oneSub$phase=="Age"), aes(stim_age))+
  geom_histogram(binwidth = .5)+
  facet_grid("stim_ageGroup")+
  geom_vline(xintercept=c(7.5,18.5))+
  xlab("Age of Child (Months)")+
  ylab("Number of Children")+
  labs(title = "Age of Children Separated by Age Bins")+
  theme_bw()
```



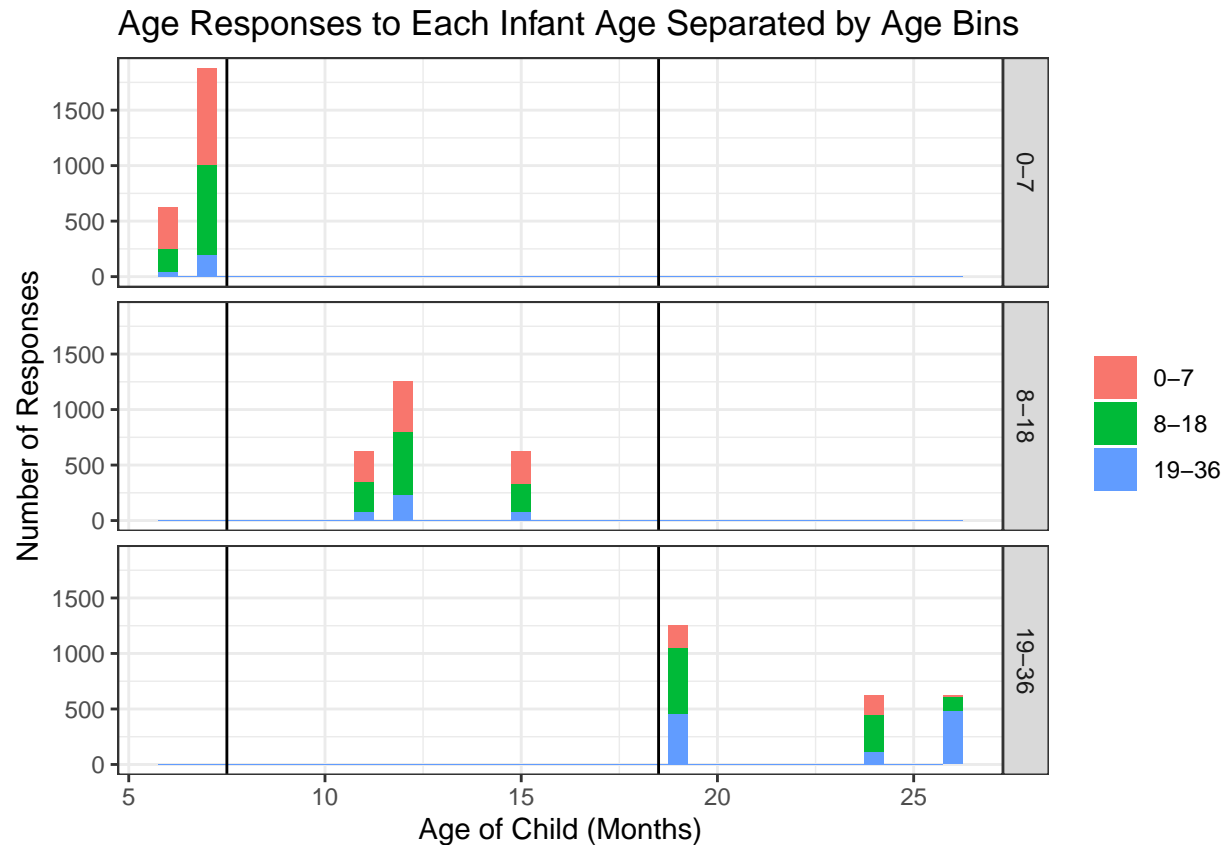
Next we can examine the responses made by participants as a function of the bins that the babies actually belonged to. This shows us that the reviewers' intuitions were correct, participants were more likely to misclassify babies in the neighbouring bins than the distant bins.

```
# plot response by bin
ggplot(subset(d,d$phase=="Age"), aes(button_pressed,fill = button_pressed))+
  geom_histogram(binwidth = .5)+
  scale_x_continuous(name = "Button Pressed", breaks = c(0,1,2),labels = c("0-7","8-18","19-36"))+
  facet_grid("stim_ageGroup")+
  ylab("Number of Responses")+
  labs(title = "Button Pressed Separated by Age Bins")+
  theme_bw()
```



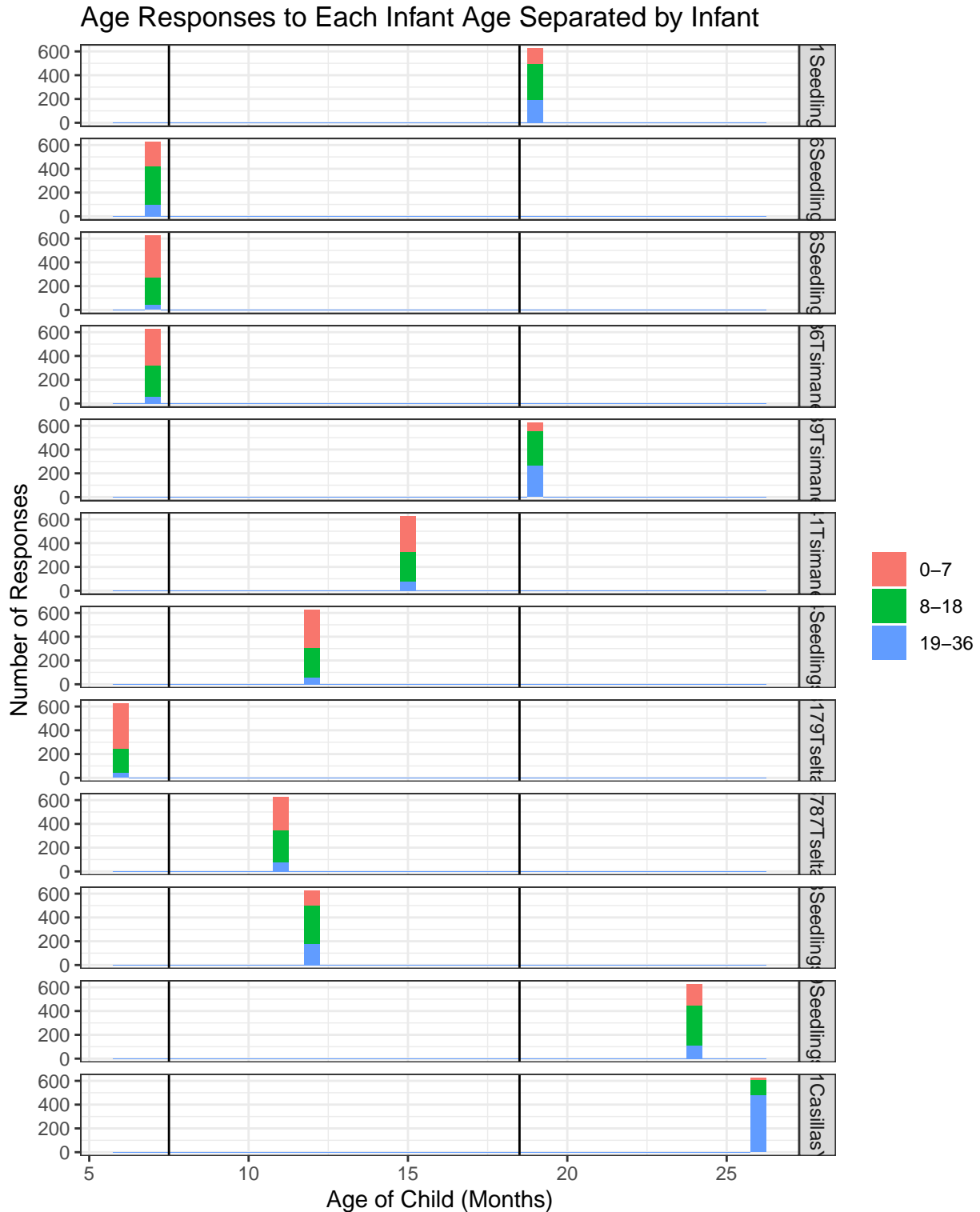
So we can then further examine participants behaviours by plotting the responses at each age of the babies. Visual inspection of this data seems to indicate that the closer babies are to the bin limit, the more likely they are to be misclassified.

```
# plot response by age
ggplot(subset(d,d$phase=="Age"),aes(fill = as.factor(button_pressed), x=stim_age))+
  geom_histogram(binwidth = .5)+
  facet_grid("stim_ageGroup")+
  geom_vline(xintercept=c(7.5,18.5))+
  xlab("Age of Child (Months)")+
  ylab("Number of Responses")+
  labs(title = "Age Responses to Each Infant Age Separated by Age Bins")+
  scale_fill_discrete(labels = c("0-7","8-18","19-36"))+
  guides(fill=guide_legend(title = NULL))+
  theme_bw()
```



And to take it even a step further we can split the graph for every individual infant. Visual inspection seems to indicate that there is notable variability in responses to babies within each age bin.

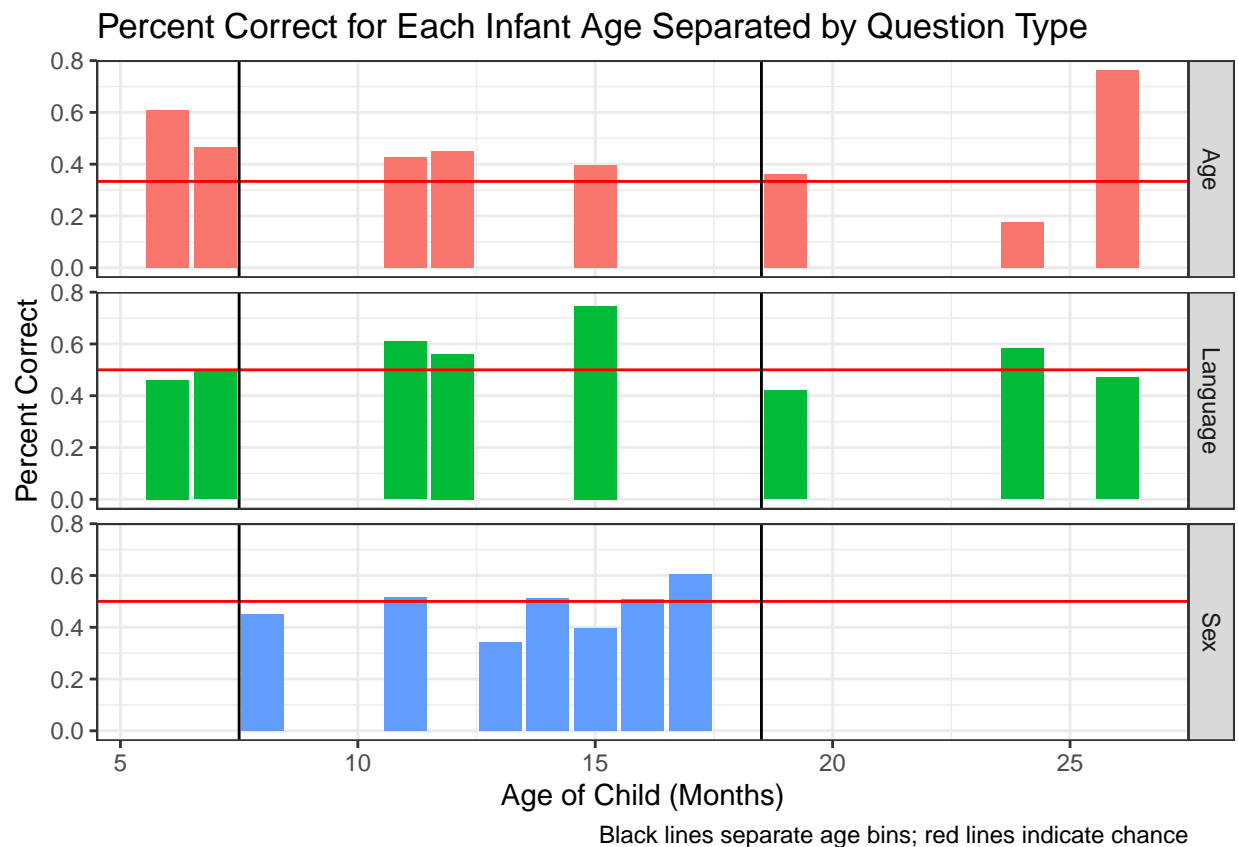
```
# plot response by individual baby
ggplot(subset(d,d$phase=="Age"),aes(fill = as.factor(button_pressed), x=stim_age))+
  geom_histogram(binwidth = .5)+
  facet_grid("stim_ID")+
  geom_vline(xintercept=c(7.5,18.5))+
  xlab("Age of Child (Months)")+
  ylab("Number of Responses")+
  labs(title = "Age Responses to Each Infant Age Separated by Infant")+
  scale_fill_discrete(labels = c("0-7","8-18","19-36"))+
  guides(fill=guide_legend(title = NULL))+
  theme_bw()
```



And finally we can examine accuracy rates for each age. If participants were more likely to misclassify a baby near the bin limit then there should be a drop in accuracy around that point. Unfortunately there is not a full spread of infant ages for what we would consider a proper analysis but from the data we do have, there doesn't appear to be a notable drop in accuracy near the bin limits. We have included the results for all 3 questions for a complete visualization of the data.

```
# make dummy variable for lines
dummy <- data.frame(phase = c("Age", "Language", "Sex"), Z = c(1/3, 1/2, 1/2))

# plot percent correct for each age and question type
ggplot(d, aes(y=correct, x=stim_age, fill=phase)) +
  geom_bar(position = "dodge", stat = "summary", fun = "mean", show.legend = F) +
  facet_grid("phase") +
  geom_hline(data = dummy, aes(yintercept=Z), col = "red") +
  geom_vline(xintercept=c(7.5, 18.5)) +
  xlab("Age of Child (Months)") +
  ylab("Percent Correct") +
  labs(title = "Percent Correct for Each Infant Age Separated by Question Type", caption = "Black lines separate age bins; red lines indicate chance") +
  scale_fill_discrete(labels = c("0-7", "8-18", "19-36")) +
  guides(fill=guide_legend(title = NULL)) +
  theme_bw()
```



However, to move beyond visual inspection we did test whether there was a relationship between the accuracy of responses and how close the babies were to bin limits. We first calculated the “age error” as the absolute difference between the babies age in months and the closest bin limit; 7.5 or 18.5 (this was calculated earlier in the document). We then tested the correlation between the “age error” and accuracy in responding. We view the reliability of this analysis as somewhat questionable but it was a tractable method of testing the question. Results demonstrated that there was a small but significant correlation.

```
# load in library
library(rmcorr)
```

```
## Warning: package 'rmcorr' was built under R version 4.0.5
```

```
# refactor the subject_ID to avoid an error with this package
d$subject_ID<-factor(d$subject_ID)
# get the repeated measures correlation coefficient
rmcorr(participant = subject_ID,
        measure1 = age_error,
        measure2 = correct,
        dataset = subset(d,d$phase=="Age"))
```

```
##
## Repeated measures correlation
##
## r
## 0.05902455
##
## degrees of freedom
## 6885
##
## p-value
## 9.500552e-07
##
## 95% confidence interval
## 0.03545234 0.08253112
```