

# Flip: Data analysis

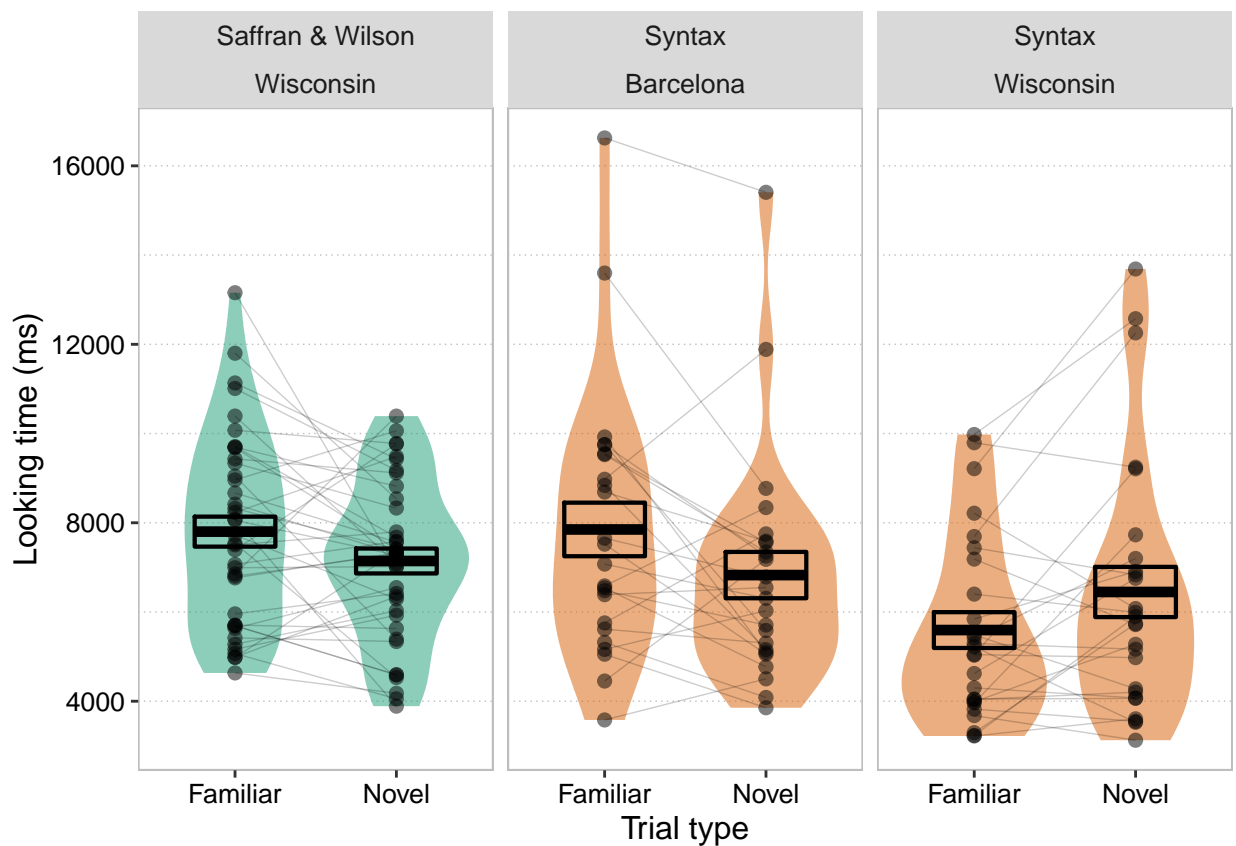
## Contents

Compare distributions of Syntax and Saffran & Wilson . . . . .	1
Linear (Correlational) Analyses . . . . .	2
Linear mixed-effects analyses . . . . .	9
How to interpret the distribution of absolute values of novelty preference scores? . . . . .	12

First, we read in the data and compute both the novelty preference (total looking time to the novel stimuli minus total looking time to familiar stimuli), as well as the novelty preference **corrected for total looking time**, i.e. the “novelty preference score” described in the summary from 10 July 2019.

## Compare distributions of Syntax and Saffran & Wilson

The first thing I think that it makes sense to do is just to look at the distribution of (average) looking time for familiar and novel stimuli in the two studies. This seems like a good first step before combining the data from the two studies.



Looking at the average looking times, it seems from visual inspection that the distributions look somewhat similar. Perhaps the looking times for Syntax in Wisconsin are somewhat shorter, mainly because the looking times to familiar items look substantially shorter in Wisconsin compared to Barcelona.

In general, there's nothing that looks too dramatic here to me suggesting that one distribution looks radically different than the others (in shape or in the range of values).

## Linear (Correlational) Analyses

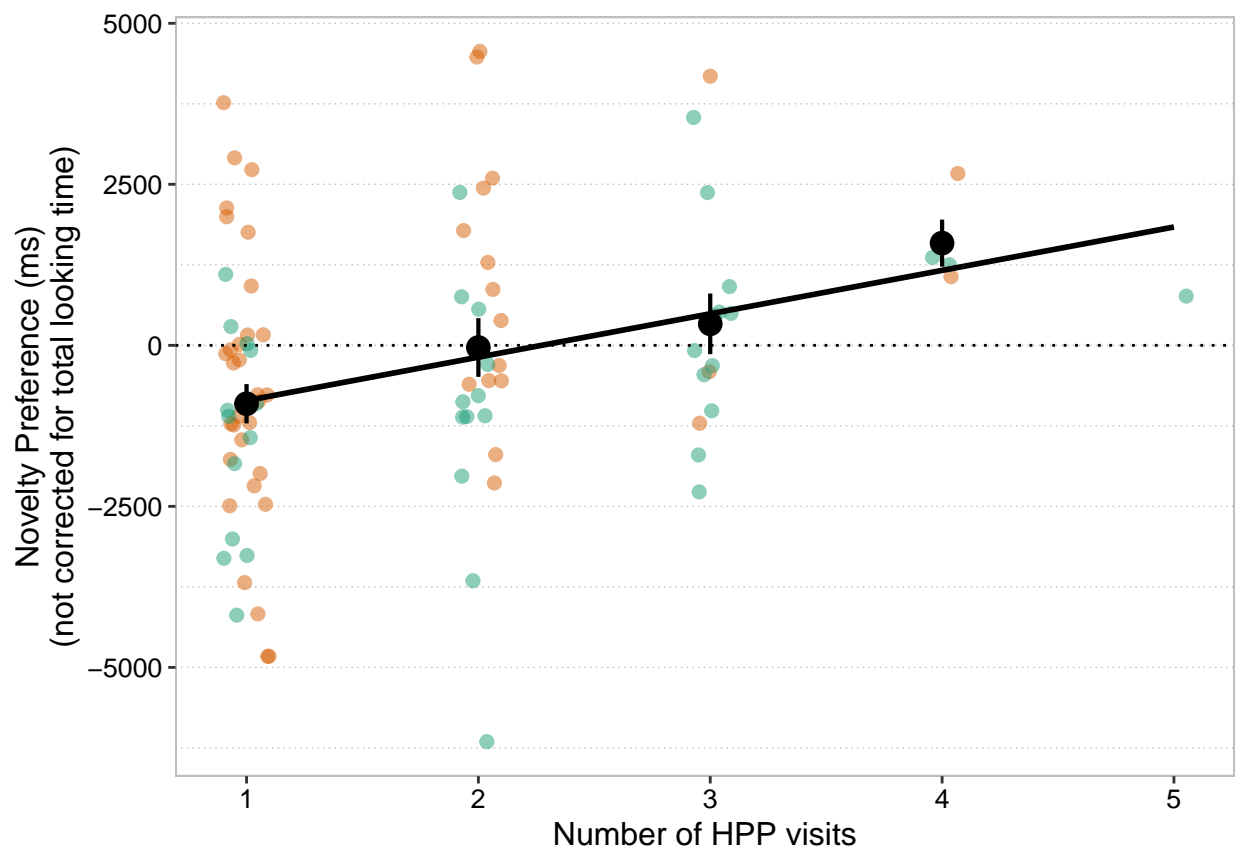
As a first step, I'm just going to look at the linear relationship between visit number and headturn preference, not accounting for non-independence due to other factors (e.g., study). This helps for getting a sense of the simple effect and how robust it is to different sources of non-independence/ analytical decisions. I'll also look at how much the effect relies on high leverage points (children with higher visit numbers) and visualize the effect for each study and location.

### Visit number predicts magnitude of novelty/ familiarity preference

#### Uncorrected novelty preference

Fit a linear model predicting novelty preference from number of HPP visits (equivalent to a correlation).

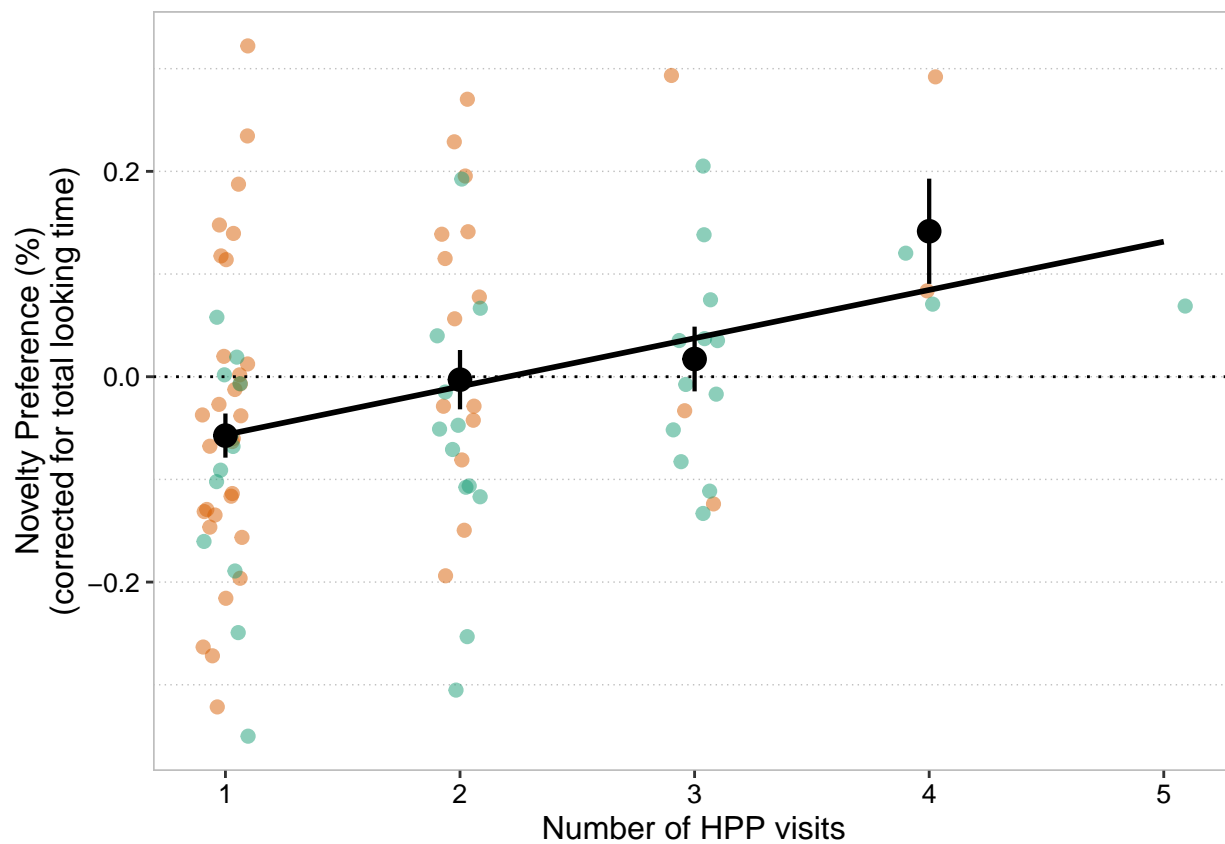
	Coefficient	SEM	F	p-value
(Intercept)	-1530.987	459.778	-3.330	0.001
hpp	673.657	226.151	2.979	0.004



#### Corrected novelty preference

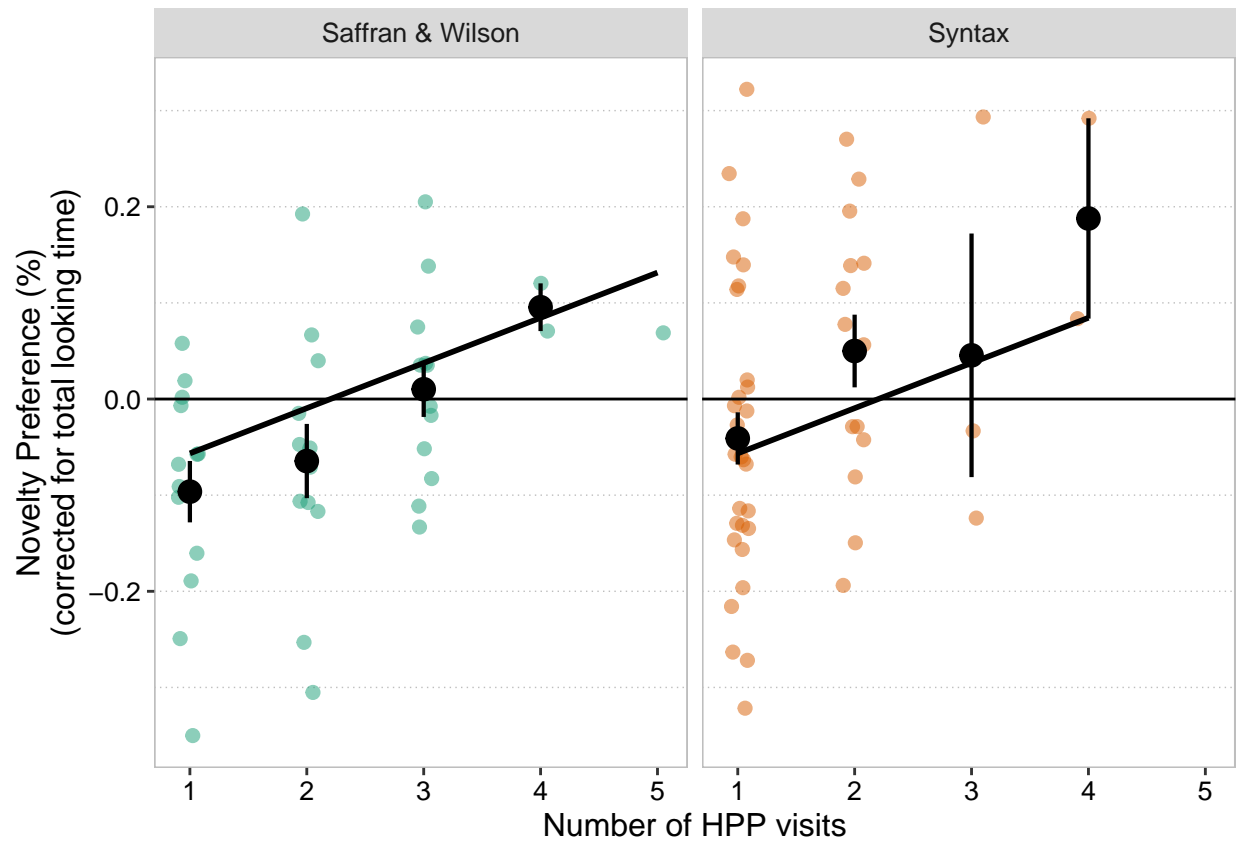
Fit a linear model predicting novelty preference from number of HPP visits (equivalent to a correlation).

	Coefficient	SEM	F	p-value
(Intercept)	-0.104	0.031	-3.306	0.001
hpp	0.047	0.015	3.051	0.003

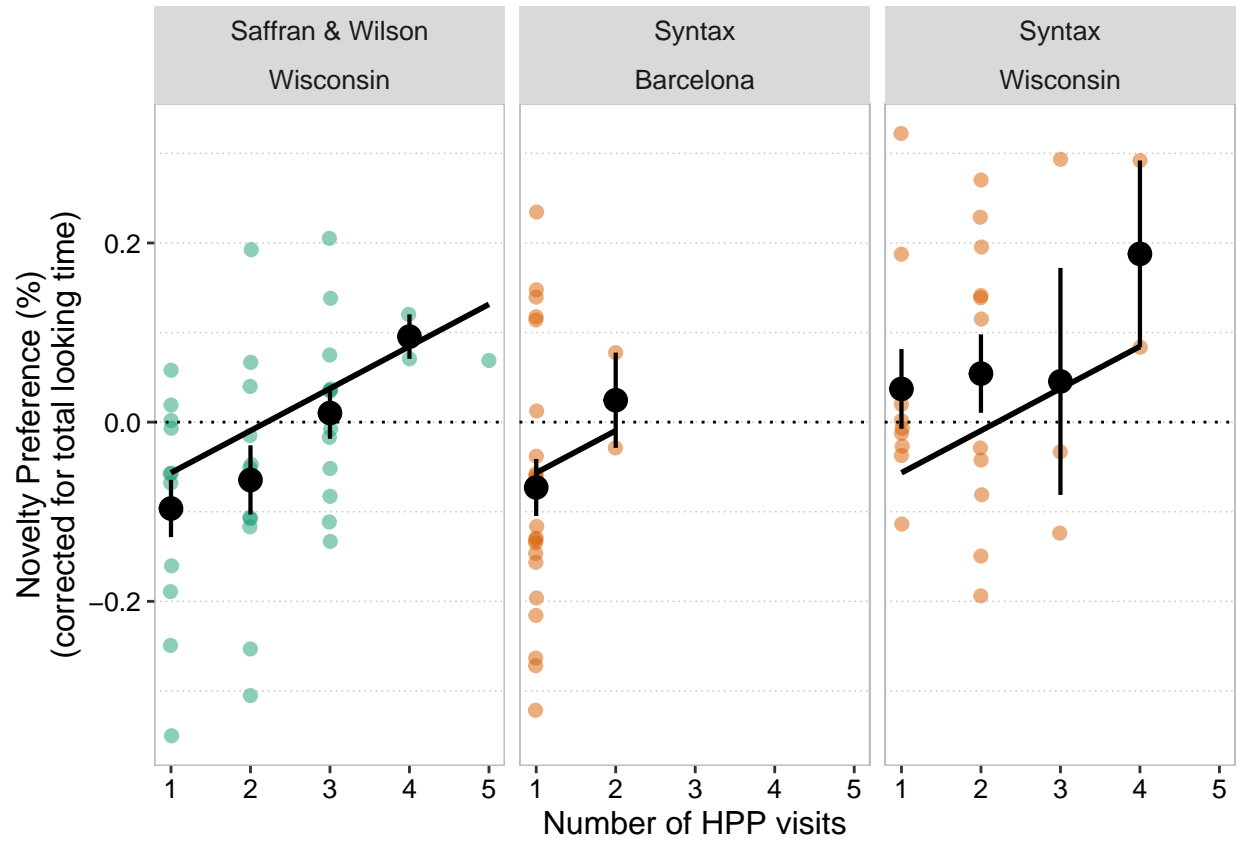


Split the prediction by study/location

Split by study (Santolin vs. Saffran & Wilson)



####Split by study and location (Wisconsin vs. Barcelona)

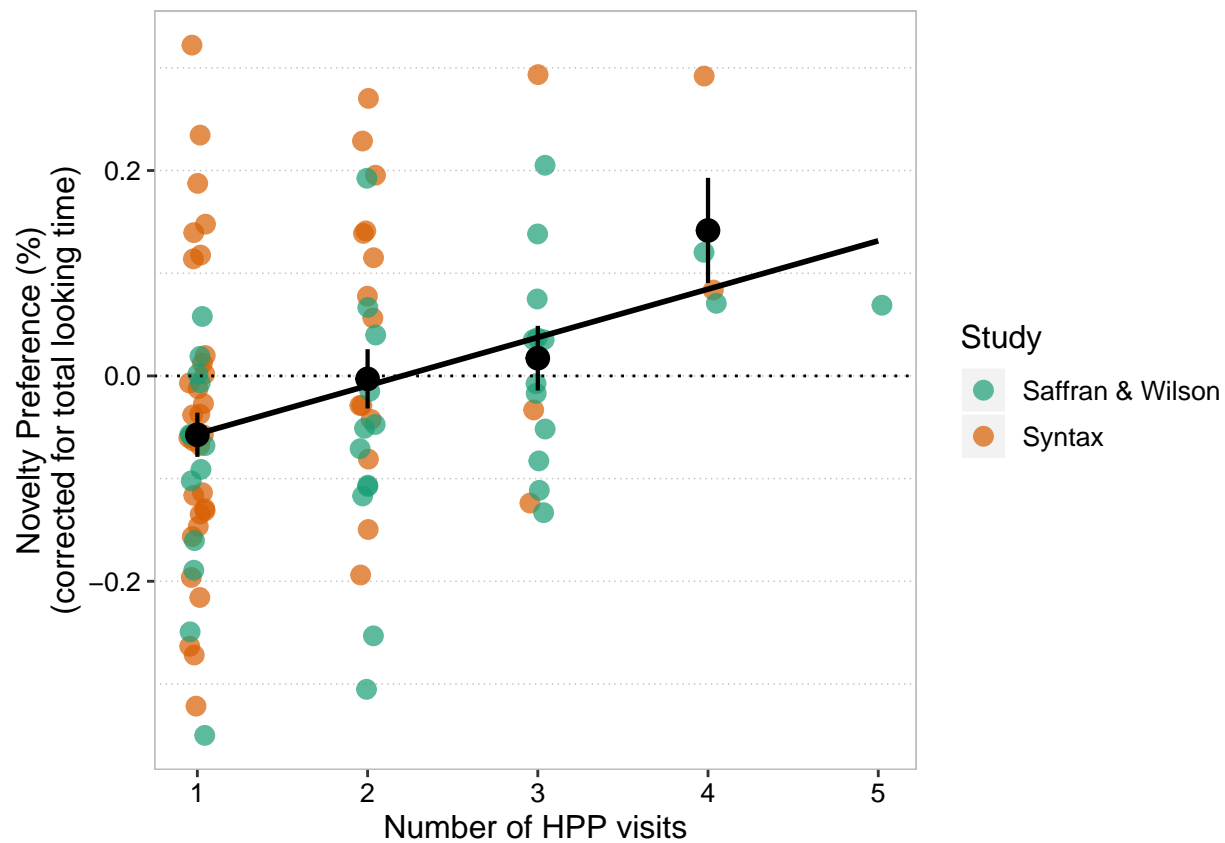


How robust is this analysis to reduced numbers of visits

1 - 5 visits

Fit a linear model predicting novelty preference from number of HPP visits (equivalent to a correlation).

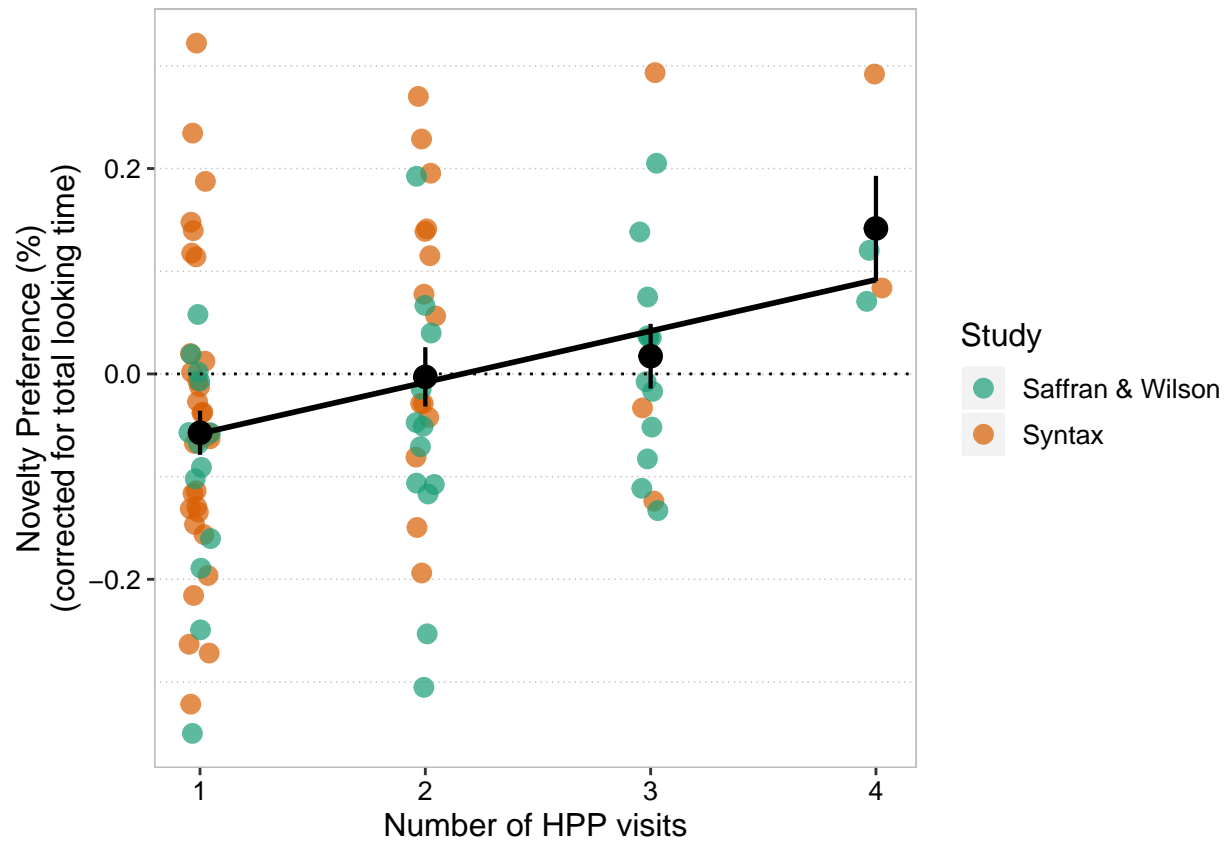
	Coefficient	SEM	F	p-value
(Intercept)	-0.104	0.031	-3.306	0.001
hpp	0.047	0.015	3.051	0.003



#### 1 - 4 visits

Fit a linear model predicting novelty preference from number of HPP visits (equivalent to a correlation).

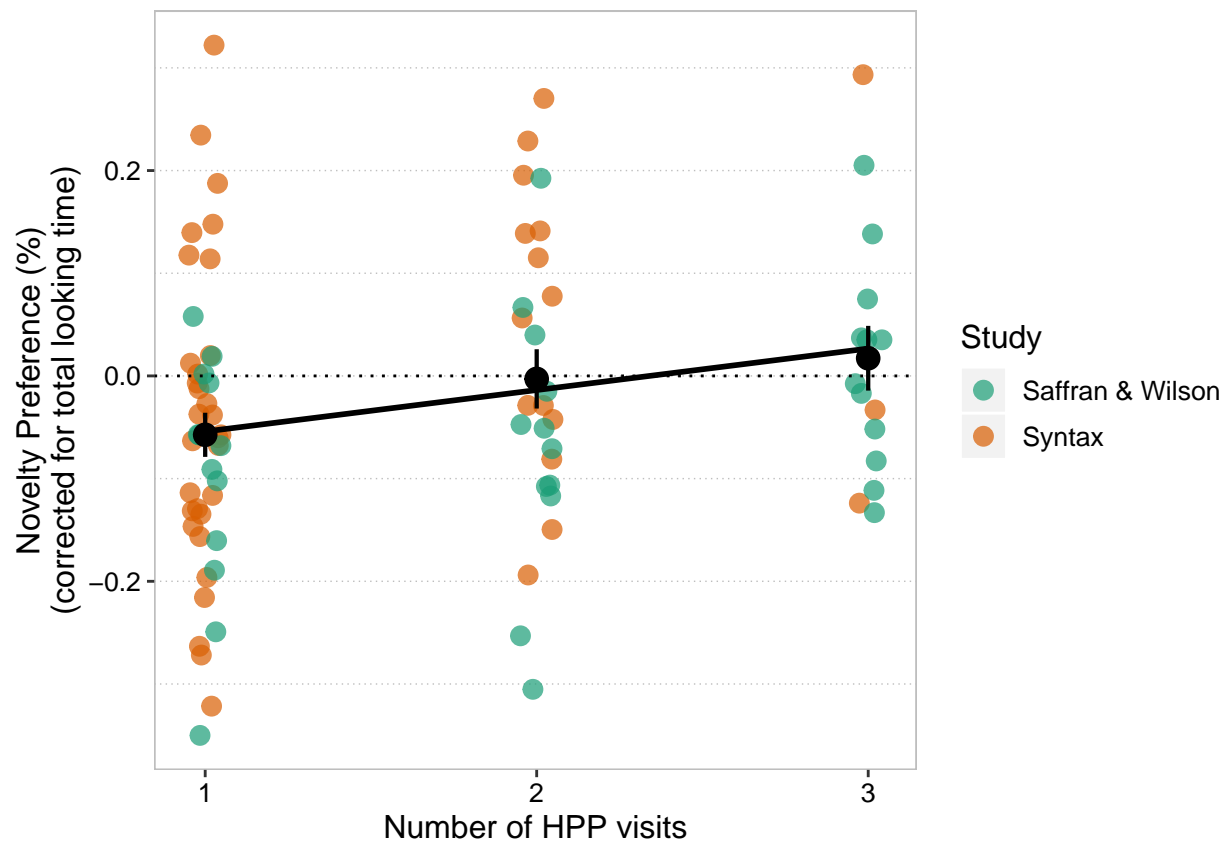
	Coefficient	SEM	F	p-value
(Intercept)	-0.108	0.033	-3.298	0.001
hpp	0.050	0.017	3.011	0.003



##### 1 - 3 visits

Fit a linear model predicting novelty preference from number of HPP visits (equivalent to a correlation).

	Coefficient	SEM	F	p-value
(Intercept)	-0.095	0.036	-2.591	0.011
hpp	0.040	0.020	2.020	0.047

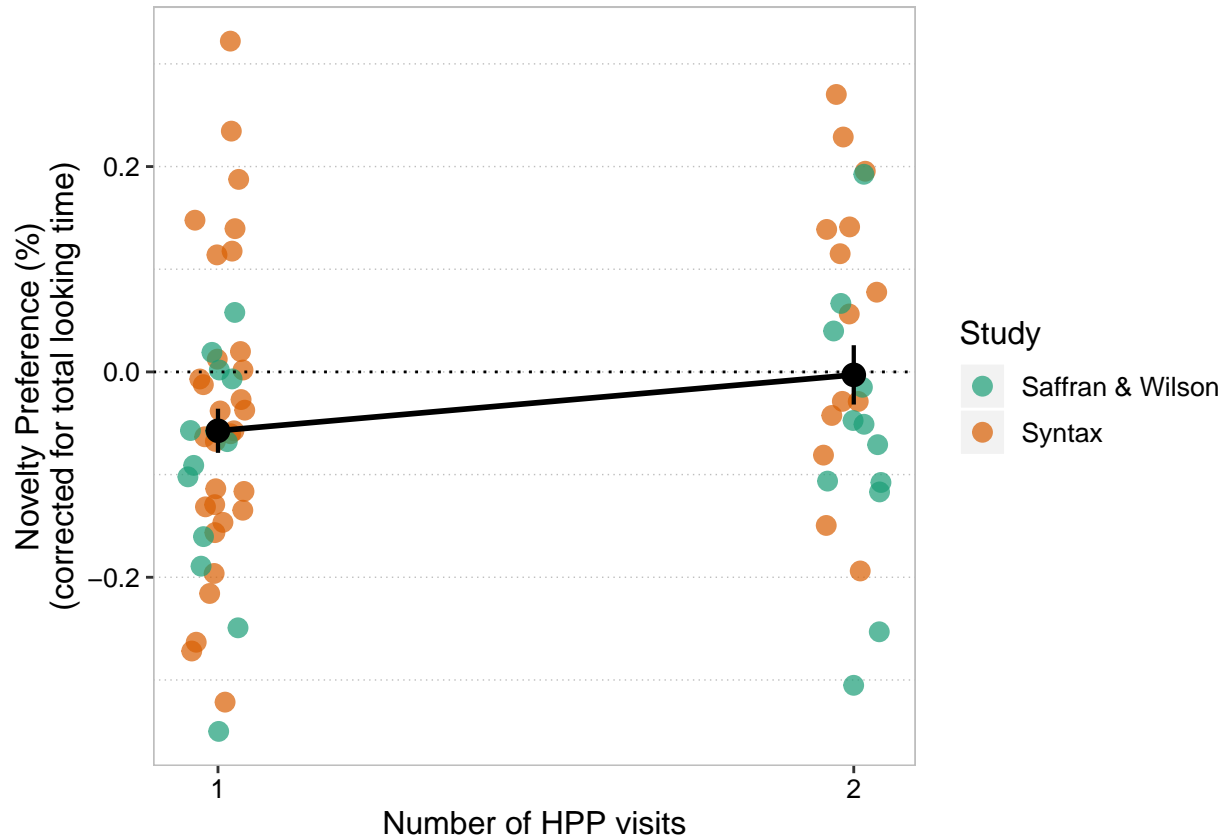


### 1 - 2 visits

Fit a linear model predicting novelty preference from number of HPP visits (equivalent to a correlation).

	Coefficient	SEM	F	p-value
(Intercept)	-0.112	0.052	-2.154	0.035
hpp	0.054	0.036	1.526	0.132





## Linear mixed-effects analyses

### Reproducing analysis with summarized difference scores

Here we specify a model that tries to predict the corrected novelty preference score (CNPS) as a function of the number of HPP, including a random intercept the variable *Study*. This essentially reproduces the model reported in the shared analysis script from 10 July 2019, with a few slight technical differences. For instance, I will simply fit the model with the maximal random effects structure here as the “default” and use Wald chi-squared tests to get p-values (not my preferred method, which is Kenward-Rogers approximation, but this has some odd issues in this instance - see below).

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: novel_pref_corr ~ hpp.c + (1 + hpp.c | Study)
## Data: data
##
## REML criterion at convergence: -93.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.1050 -0.6421 -0.0060  0.6537  2.7109
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## Study (Intercept) 0.002630 0.05128
## hpp.c 0.000111 0.01054 1.00
```

```
## Residual          0.017867 0.13367
## Number of obs: 90, groups: Study, 2
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept) -0.02052    0.03891  -0.527
## hpp.c        0.05971    0.01745   3.422
##
## Correlation of Fixed Effects:
##      (Intr)
## hpp.c 0.393
## convergence code: 0
## boundary (singular) fit: see ?isSingular

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: novel_pref_corr
##           Chisq Df Pr(>Chisq)
## (Intercept)  0.278  1  0.5979974
## hpp.c       11.707  1  0.0006225 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lmer(novel_pref_corr ~ hpp.c + (1 + hpp.c | Study), data = data)
```

$$CNPS_{ps} = \beta_0 + Study_{0s} + (\beta_1 + Study_{1s})HPP_i + e_i$$

### Approach using long-format data & lme4

The best (most powerful) analysis would involve fitting linear mixed-effects models to the trial-by-trial looking time data (*LT*). This model would look something like this, fit on the trial-level looking data:

```
lmer(looking.time ~ looking.type * hpp + (1 + looking.type | baby) + (1 | item) + (1 + looking.type * hpp | Study))
```

$$LT_{ips} = \beta_0 + Item_{0i} + Participant_{0p} + Study_{0s} + (\beta_1 + Participant_{1p})HPP_i + (\beta_2 + Study_{1s})Condition_i + (\beta_3 + Participant_{3p} + Study_{3s})$$

For now, let's consider a mixed-effects model fit to the average looking time data.

I think the model that we want to test would involve predicting an interaction between item type (familiar vs. novel) and number of hpp visits on looking time - the effect of familiar vs. novel items on looking behavior **depends on** the number of headturn visits an infant has had.

```
lmer(looking.time ~ looking.type * hpp + (1 | baby) + (1 + hpp | Study))
```

$$LT_{is} = \beta_0 + Participant_{0p} + Study_{0s} + (\beta_1 + Participant_{1p} + Study_{1s})HPP_i + Condition_c + e_{is}$$

We include a by-participant random intercept, since we have multiple observations per baby. Since item type has only two levels, we fit the model without a by-participant random slope (since this overdetermines the model; my understanding is that it's OK to still force the model to accept the random slope, but it shouldn't matter for the estimates).

```

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## looking.time ~ looking.type.c * hpp.c + (1 | Baby) + (1 + hpp.c |
##   Study)
##   Data: data_long
##
## REML criterion at convergence: 3209.9
##
## Scaled residuals:
##   Min       1Q   Median       3Q      Max
## -1.7735 -0.5089 -0.0404  0.3505  2.4932
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   Baby      (Intercept)    3620285   1902.70
##   Study     (Intercept)    422113    649.70
##           hpp.c              1581     39.77  -1.00
##   Residual                2056028   1433.89
## Number of obs: 180, groups:  Baby, 90; Study, 2
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      7070.8      513.3  13.775
## looking.type.c    -318.4      213.8   -1.490
## hpp.c             -307.6      252.7   -1.217
## looking.type.c:hpp.c  673.7      226.2    2.979
##
## Correlation of Fixed Effects:
##              (Intr) lkng.. hpp.c
## lkng.typ.c   0.000
## hpp.c        -0.115  0.000
## lkng.typ.:.  0.000  0.000  0.000
## convergence code: 0
## boundary (singular) fit: see ?isSingular

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: looking.time
##               Chisq Df Pr(>Chisq)
## (Intercept)    189.7520  1 < 2.2e-16 ***
## looking.type.c    2.2189  1  0.136328
## hpp.c            1.4816  1  0.223523
## looking.type.c:hpp.c  8.8731  1  0.002894 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The significant interaction term suggests that the number of visits changes the relationship between item type and looking time, i.e. the nature of the familiarity/novelty preference.

The way to get appropriate p-values from lmer models is of course a matter of debate. Here, we used a Wald chi-squared test. Typically, the method I prefer is the Kenward-Rogers approximation which estimates the “correct” degrees of freedom. However, there’s something a bit odd going on when adding the hpp by-study random slope with this method, which is probably because study only has two levels. The chi-squared method should be fine, but this is also something we could do a deeper dive on down the road (it might also be just another argument against including the hpp by-study random slope).

## Is the effect robust when we remove babies with high numbers of visits (4 or 5)?

Checking here that the effect still holds when we remove high leverage points - the handful of babies who had 4 or 5 visits.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## looking.time ~ looking.type.c * hpp.c + (1 | Baby) + (1 + hpp.c |
##   Study)
##   Data: filter(data_long, hpp < 4)
##
## REML criterion at convergence: 3034.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.72994 -0.52285 -0.04872  0.37493  2.43591
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   Baby      (Intercept) 3748272  1936.0
##   Study     (Intercept) 344495  586.9
##           hpp.c          9158   95.7  -1.00
## Residual                2158931 1469.3
## Number of obs: 170, groups:  Baby, 85; Study, 2
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      7098.2     481.2  14.750
## looking.type.c    -326.8     229.2  -1.426
## hpp.c             -216.1     335.4  -0.644
## looking.type.c:hpp.c  665.9     296.5   2.246
##
## Correlation of Fixed Effects:
##              (Intr) lkng.. hpp.c
## lkng.typ.c   0.000
## hpp.c        -0.103  0.000
## lkng.typ.:.  0.000  0.183  0.000

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: looking.time
##              Chisq Df Pr(>Chisq)
## (Intercept)    217.5666  1 < 2e-16 ***
## looking.type.c    2.0322  1  0.15399
## hpp.c            0.4151  1  0.51942
## looking.type.c:hpp.c  5.0434  1  0.02472 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## How to interpret the distribution of absolute values of novelty preference scores?

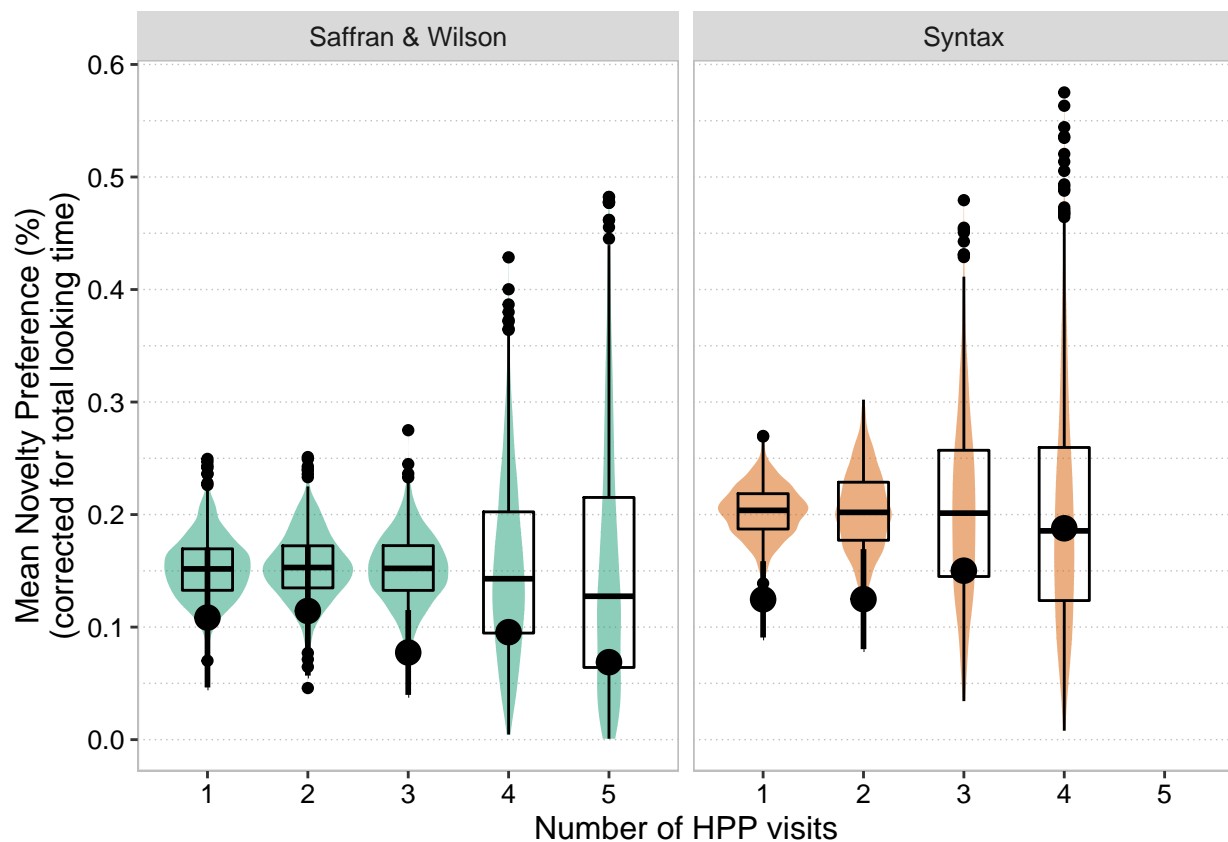
To address this question, I ran two separate small-scale simulations based on the distribution of looking times in our existing data set. The general question is - what might the distribution of absolute novelty

preferences scores look like, in general, even in the absence of underlying systematic individual differences in preference direction.

### Shuffled IDs

The first simulation I ran was simply re-shuffling infant ids across the observed looking times. On each iteration, I re-shuffled subject ids within the set of novel and familiar looking times for each study, while maintaining the proportion of infants with each number of HPP visits. This breaks the association between a specific looking time for familiar stimuli and a specific looking time for novel stimuli.

Below is a plot of the distribution of **average** absolute NPS values after 1000 simulations. The observed absolute NPS values in the current study are plotted in black. As we can see, it seems like the observed absolute NPS values would not be surprising in the absence of systematic relationships between familiar and novel looking time stimuli, given the distribution of those values in the population in general.

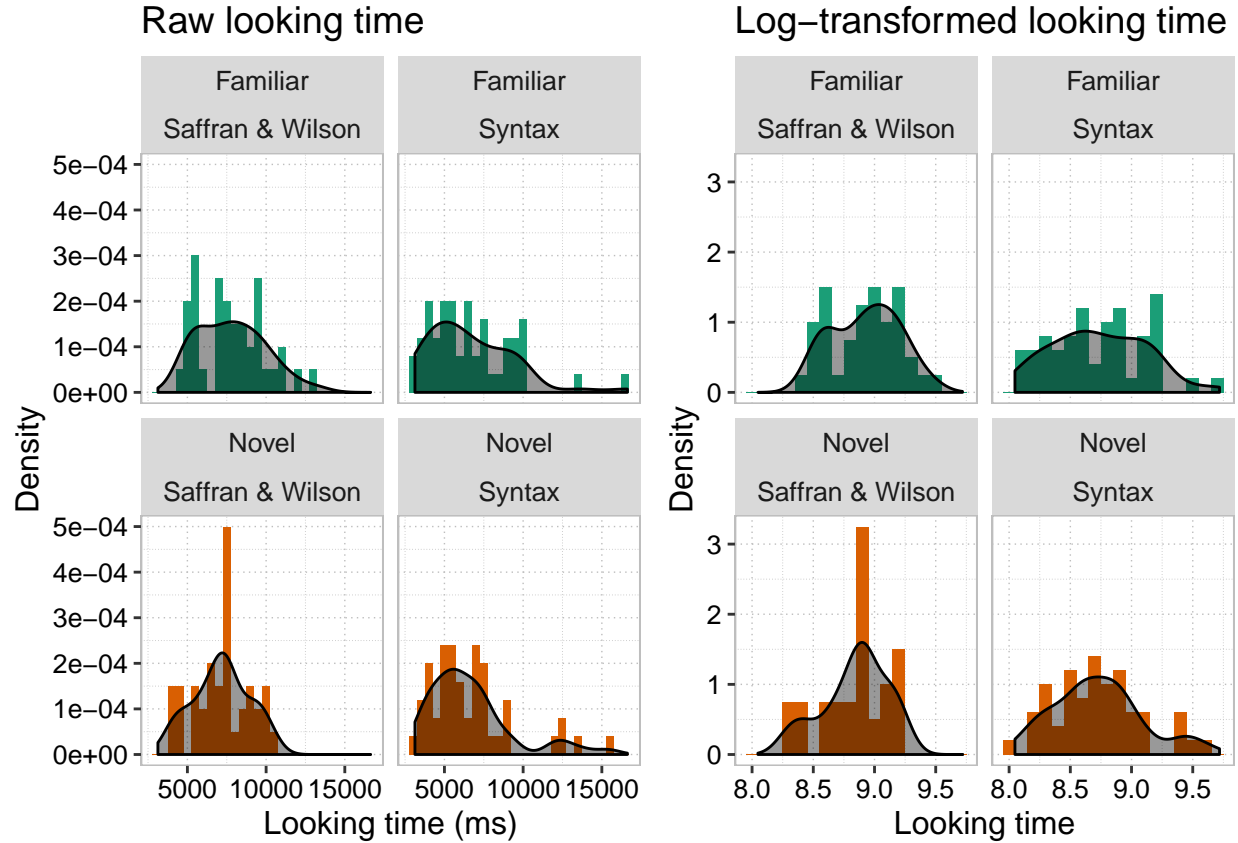


### Simulated data assuming no individual-level effect

Although I find the simulation above informative and fun, it's not entirely obvious to me that it's the strongest test of whether the absolute NPS values are meaningful, since it's not explicitly investigating the distribution of absolute values, **assuming absolutely no systematic preference**.

To investigate this question, I simulated new observations based on the variation in values in the current data set, assuming no significant preference for familiar and novel stimuli for each simulated individual and that the log looking time values are normally distributed.

Empirically, it appears from the plot below that the looking times themselves are not normally distributed, but the log-transformed values are somewhat consistent with a normal distribution.



In this simulation, we regenerate a new dataset repeatedly from scratch, assuming no systematic familiarity/novelty preference within each individual and that the looking time values are normally distributed with the same SD observed in the current study.

One tweak in this simulation is that I account for systematic relationships between familiar and novel looking times within a given subject by - for each subject - sampling from a distribution of average overall looking times, and then shifting the distribution of the looking time values based on that mean. Note that the familiar and novel looking times are sampled from the same distribution for each infant, i.e. each simulated participant has no preference for familiar vs. novel stimuli.

The plot below shows the distribution of average absolute NPS values for each study after 1000 simulations. As you can see, the level of average absolute NPS values observed in our dataset (plotted in black) would not be surprising in a situation where infants have **no preference** between novel and familiar stimuli.

