Analysis Record - SubDrop Kids project

Melissa <mekline@mit.edu>

2018-01-06

SubDrop Analysis document

Here is (will be) the pipeline for executing all code to get you from subdrop_reconciled.csv to all analyses reported in the paepr

Preliminaries

(Code is suppressed here for the most part; see the RMD file if interested). It loads all the libraries,

makes some convenience functions,

and loads the data

Check here to see that your data was loaded:

```
head(subtable[,1:10], n=3)
```

```
##   Experimenter          Design   Experiment Condition Correct Subject..
## 1        Jenny Between-2trials ParentSecret        SD       1         1
## 2        Jenny Between-2trials ParentSecret        SD       2         2
## 3        Jenny Between-2trials ParentSecret        SD       2         3
##   Gender                 Video.Name Days.Old Age.Years
## 1      F subject_drop_040414_01.mpg     1895         5
## 2      M subject_drop_040414_02.mpg     2436         6
## 3      F subject_drop_040414_03.mpg     2269         6
```

Data cleaning

So boring, so necessary. Relabeling columns and fixing factor/character encoding issues.

Here is a full report on how many rows are dropped from the analyzed dataset and why. It's principally bilingualism, reported developmental delay, error with consent/no consent, age outside the intended sample (but tested anyhow because children's museum) A note on "ExpErrorJ": A series of major implementation flaws (e.g. puppets did not see the events they were supposedly describing) were discovered after several months :(- RA implementation was very inconsistently implemented so a large # of participants must be excluded)

```
table(dropped$Final.Reason, dropped$Experiment)
```

```
##
##                         ParentSecret ParentSecretControl
##   BILINGUAL                       16                   0
##   DEVELOPMENTAL DISORDER           1                   1
##   ExpError                         6                   2
##   ExpErrorJ                       67                   0
##   FUSSOUT                          5                   5
##   KID INTERFERENCE                 6                   1
##   NO CONSENT                       2                   0
```

```
##   PARENT INTERFERENCE                 2                   0
##   WRONG AGE                           5                   8
```

This is followed by yet more data cleaning. One major note: We tried 2-trial (between-subj) and 4-trial (within-subj) versions of the task. With within-subj 4-trial version, we saw big carryover effects during the first part of data collection, and intended to stop collecting the 4-trial version (but did not do so consistently because the extra trials were still in the book.) So, we only ever analyzed just the 1st 2 trials, treating this data as a between-subjects comparison.

Descriptive statistics

First we report some basic descriptive stats for the datasets. We begin by splitting the data into the 'main' and 'control' versions, and limiting the latter to 3- and 4-year-old participants (a few kids of other ages were run in that version thanks to the museum context.)

First, report the n kids in each sub-experiment (this was useful for checking updates on subjects needed per condition.

(A note on naming: conditions for the critical experiment are named OD (object-drop) and SD (subject-drop), indicating the *pragmatically correct* answer in a given condition. (In the control experiment, there was a *factually incorrect* answer to contrast with a (correct) shortened sentence that dropped either object or subject).

```
##     3  4  5  6
## OD 12 16 12 12
## SD 15 16 17  9
```

```
##     3  4
## OD 14 15
## SD 14 11
```

For a quick summary statistic, report how often children chose the object-dropping puppet in the two conditions.

```
##        OD        SD
## 0.5192308 0.3947368
```

Inferential statistics

First we need to melt the dataset for logistic regression…

Test #1

Is the choice of puppet (OD vs. SD) sensitive to condition? (That is, do children choose different puppets depending on the nonlinguistic context.) We did not preregister this experiment, but this is the first analysis we tried for these data.

Note that there is no age factor in this model, and that there is a by-trial Condition slope, but no by-participant slope because Condition is manipulated between subjects.

Here and throughout, we evaluate models by comparison to a model lacking the fixed effect of interest.

```
full_maximal_model <- glmer(choseObjectDrop ~ Condition + (Condition|trial) + (1|Subject), data=main.lon

#compare to model w/o fixed effect
no_fixed <- glmer(choseObjectDrop ~ 1 + (Condition|trial) + (1|Subject), data=main.long, family="binomia
```

```r
anova(full_maximal_model, no_fixed)
```

```
## Data: main.long
## Models:
## no_fixed: choseObjectDrop ~ 1 + (Condition | trial) + (1 | Subject)
## full_maximal_model: choseObjectDrop ~ Condition + (Condition | trial) + (1 | Subject)
##                    Df    AIC    BIC  logLik deviance  Chisq Chi Df
## no_fixed            5 301.59 318.52 -145.80   291.59
## full_maximal_model  6 300.77 321.08 -144.38   288.77 2.8232      1
##                    Pr(>Chisq)
## no_fixed
## full_maximal_model    0.09291 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test #2

Next, is the tendency to choose the (pragmatically) correct choice modulated by age? Here, we switch to analyzing the number of correct choices (rather than n times choosing OD) and collapse across conditions, because question order was not counterbalanced (the order was always OD-SD), so the interaction in 'correctness' between age and (between subject) condition is not interpretable.

```r
#Scale age-in-days (by z score), to avoid convergence problems
main.long$Scaled.Days.Old <- scale(main.long$Days.Old)

fullmax_age_model <- glmer(pragChoice ~ Scaled.Days.Old + (1|trial) + (1|Subject), data=main.long, famil

#model with same random effects structure as above
no_age <- glmer(pragChoice ~ 1 + (1|trial) + (1|Subject), data=main.long, family="binomial")

anova(fullmax_age_model, no_age)
```

```
## Data: main.long
## Models:
## no_age: pragChoice ~ 1 + (1 | trial) + (1 | Subject)
## fullmax_age_model: pragChoice ~ Scaled.Days.Old + (1 | trial) + (1 | Subject)
##                   Df    AIC    BIC  logLik deviance  Chisq Chi Df
## no_age             3 296.14 306.30 -145.07   290.14
## fullmax_age_model  4 291.80 305.34 -141.90   283.80 6.3465      1
##                   Pr(>Chisq)
## no_age
## fullmax_age_model    0.01176 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test #3 (control experiment)

Parallel to test #2, for the control experiment. Note only 3 and 4yos participated in this version. We don't attempt to interpret differences between the two conditions (same issues as above.)

```r
#Scale age (z score), to avoid convergence problems
cont.long$Scaled.Days.Old <- scale(cont.long$Days.Old)
```

```r
# Logistic Regression model.
full_max_cont_model <- glmer(pragChoice ~ Scaled.Days.Old + (1|trial) + (1|Subject), data=cont.long, fam

no_age_cont_model <- glmer(pragChoice ~ 1 + (1|trial) + (1|Subject), data=cont.long, family="binomial")

anova(full_max_cont_model, no_age_cont_model)
```

```
## Data: cont.long
## Models:
## no_age_cont_model: pragChoice ~ 1 + (1 | trial) + (1 | Subject)
## full_max_cont_model: pragChoice ~ Scaled.Days.Old + (1 | trial) + (1 | Subject)
##                     Df    AIC    BIC  logLik deviance  Chisq Chi Df
## no_age_cont_model    3 134.50 142.54 -64.248   128.50
## full_max_cont_model  4 133.83 144.56 -62.917   125.83 2.6631      1
##                     Pr(>Chisq)
## no_age_cont_model
## full_max_cont_model     0.1027
```

Test 4: Do the 3s and 4s differ on the two different experiments?

Here, we ask whether we should interpret any developmental change *in pragmatic abilities* between the three and 4 year olds, or in other words if we should state definitively that the three-year-olds 'dont' understand' the pragmatic calculation in the main task. We conclude that we can't conclude that!

The full-random-effects models don't initially converge, so we remove the Task/trial slope.

```r
nomax_three_model <- glmer(pragChoice ~ Task*Scaled.Days.Old + (1|trial) + (1|Subject), data=threefour.l
nomaxnoeff_three_model <- glmer(pragChoice ~ Task+Scaled.Days.Old + (1|trial) + (1|Subject), data=three

anova(nomax_three_model,nomaxnoeff_three_model)
```

```
## Data: threefour.long
## Models:
## nomaxnoeff_three_model: pragChoice ~ Task + Scaled.Days.Old + (1 | trial) + (1 | Subject)
## nomax_three_model: pragChoice ~ Task * Scaled.Days.Old + (1 | trial) + (1 | Subject)
##                        Df    AIC    BIC  logLik deviance  Chisq Chi Df
## nomaxnoeff_three_model  5 293.20 310.31 -141.60   283.20
## nomax_three_model       6 294.13 314.65 -141.06   282.13 1.0749      1
##                        Pr(>Chisq)
## nomaxnoeff_three_model
## nomax_three_model          0.2998
```

(Note that we also are underpowered here: to *detect* a difference in performance (assuming that 3yos were actually at chance on the pragmatic version and at the observed level on the control task)) we'd need 103 kids!

```
##
##      Chi squared power calculation
##
##               w = 0.3065049
##               N = 102.5565
##              df = 2
##       sig.level = 0.05
##           power = 0.8
##
```

```
## NOTE: N is the number of observations
```

Year-by-year Interpretation

The above describes the main significant effects we see in the dataset, but it's helpful to understand the dataset by binning by year. So, we conduct multinomial tests against chance for each year in both experiments.

```
## [1] "threes"
```

```
##
##  Exact Multinomial Test, distance measure: p
##
##     Events    pObs    p.value
##        406  0.0183     0.5717
```

```
## [1] "fours"
```

```
##
##  Exact Multinomial Test, distance measure: p
##
##     Events    pObs    p.value
##        561   2e-04     0.0066
```

```
## [1] "fives"
```

```
##
##  Exact Multinomial Test, distance measure: p
##
##     Events    pObs    p.value
##        465  0.0011     0.0324
```

```
## [1] "sixes"
```

```
##
##  Exact Multinomial Test, distance measure: p
##
##     Events    pObs    p.value
##        253   4e-04     0.0072
```

```
## [1] "threes, control experiment"
```

```
##
##  Exact Multinomial Test, distance measure: p
##
##     Events    pObs    p.value
##        435   9e-04     0.0258
```

```
## [1] "fours, control experiment"
```

```
##
##  Exact Multinomial Test, distance measure: p
##
##     Events    pObs    p.value
##        378       0          0
```

Similarly, we ask whether the threes and fours (respectively) differ on the two tasks.

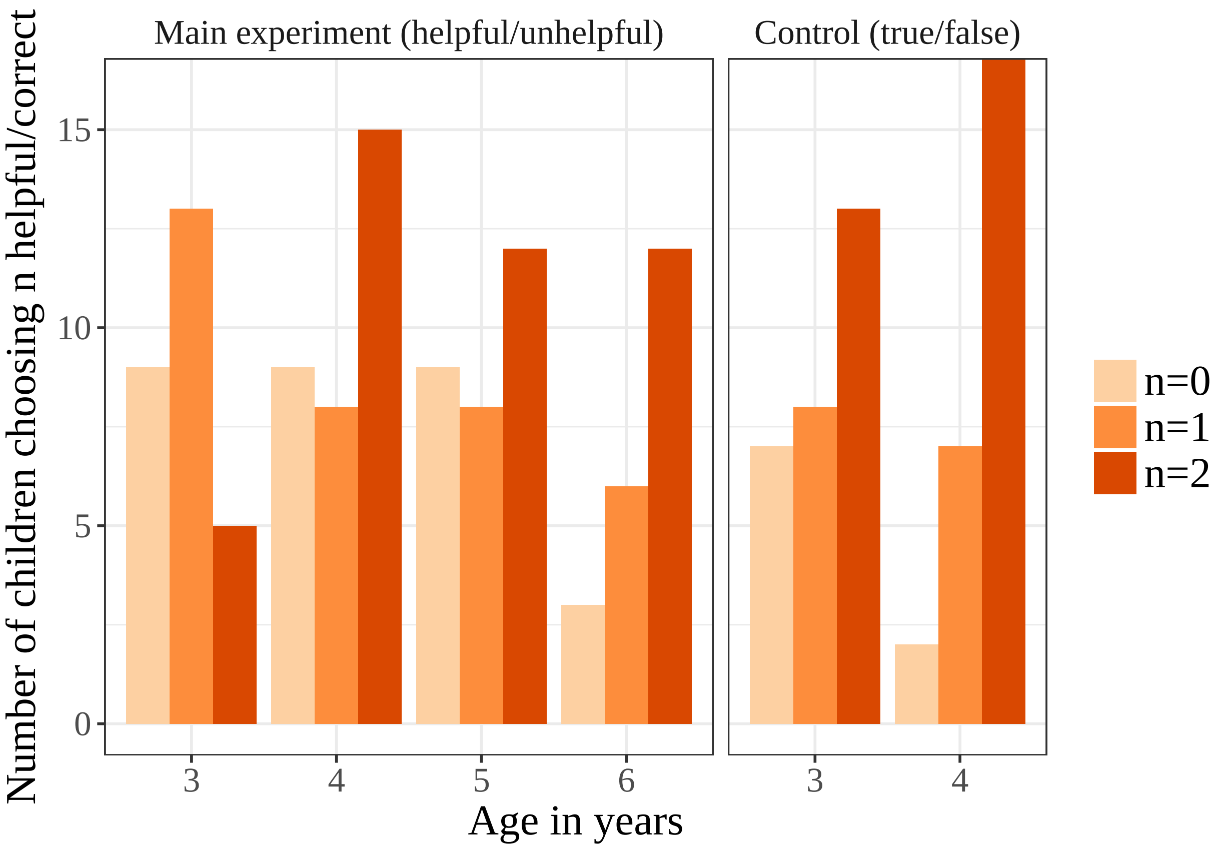
```
##
##  Fisher's Exact Test for Count Data
##
```

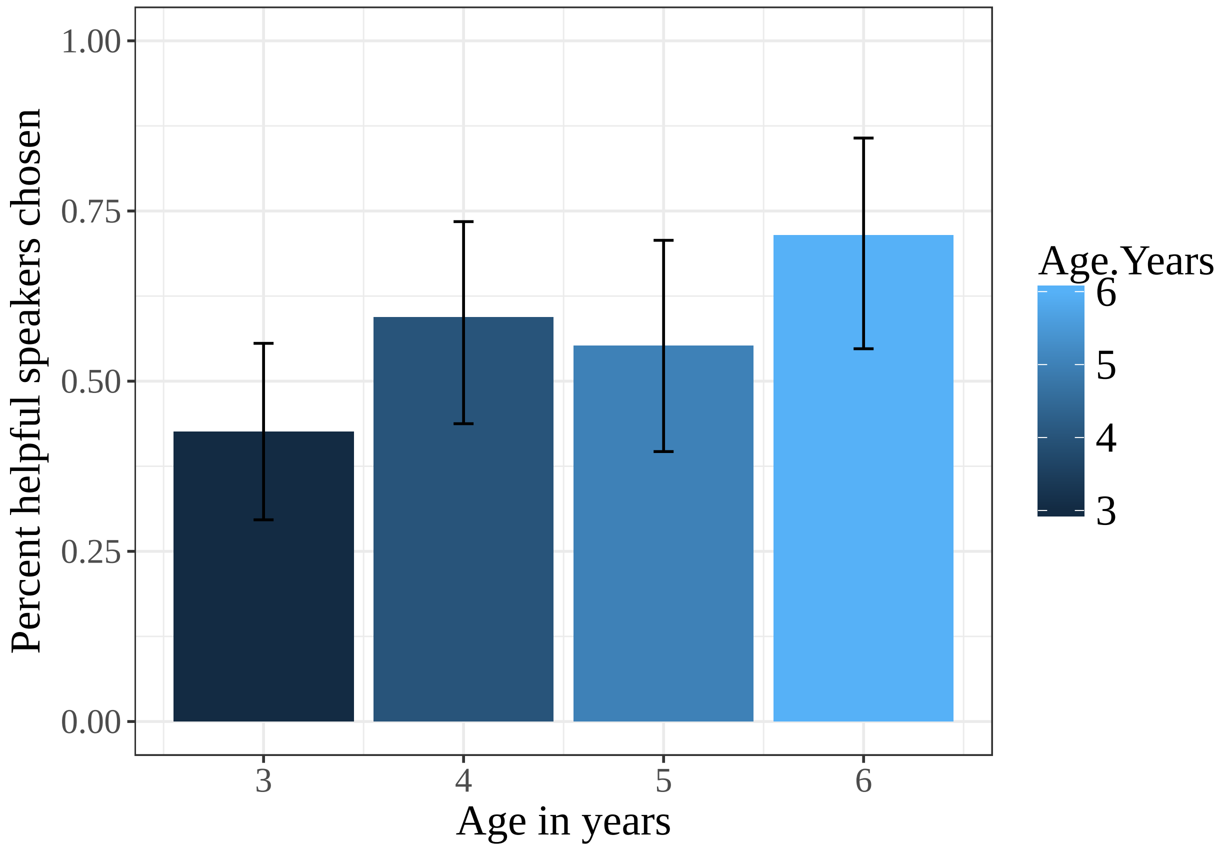
```
## data:  threetab
## p-value = 0.08499
## alternative hypothesis: two.sided
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  fourtab
## p-value = 0.1391
## alternative hypothesis: two.sided
```

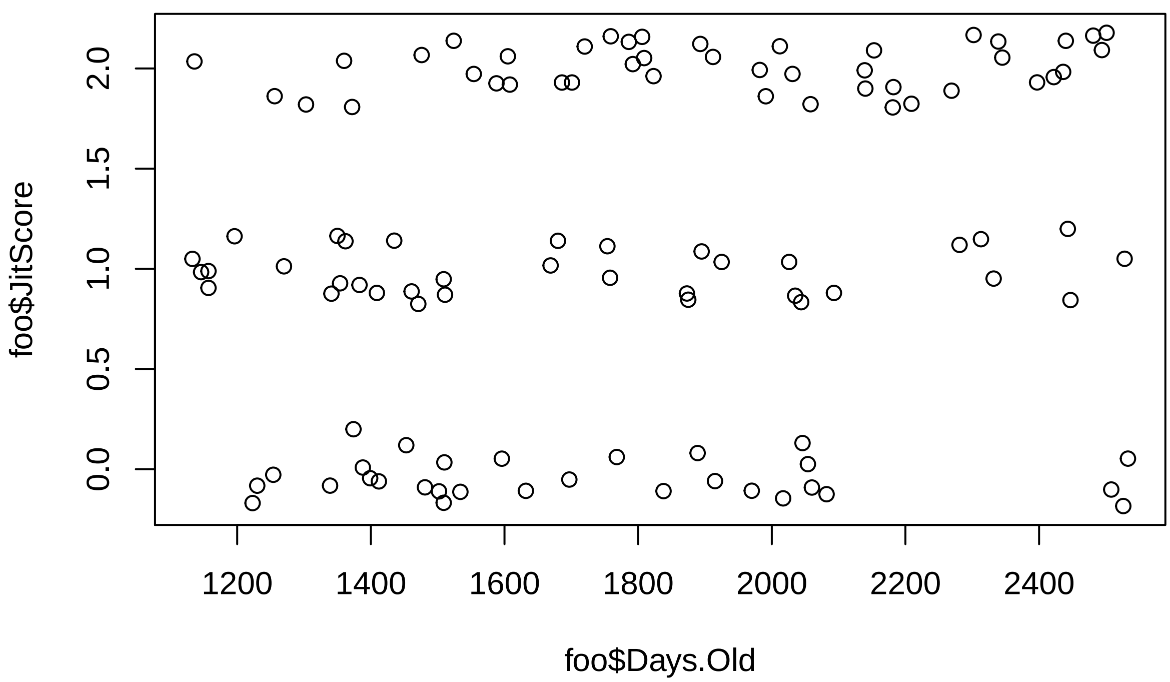
Graphs

This is the graph reported in the paper, and some simpler ones used in talks.





And a fairly uninformative dot plot.



Session information

The R session information for compiling this document is shown below.

```
sessionInfo()
```

```
## R version 3.4.1 (2017-06-30)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.1
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] bindrcpp_0.2       pwr_1.2-1          RColorBrewer_1.1-2
##  [4] tufte_0.2          liftr_0.7          bootstrap_2017.2
##  [7] ggplot2_2.2.1      EMT_1.1            lsr_0.5
## [10] dplyr_0.7.2        binom_1.1-1        multcomp_1.4-8
## [13] TH.data_1.0-8      MASS_7.3-47        survival_2.41-3
## [16] mvtnorm_1.0-6      lme4_1.1-13        Matrix_1.2-10
## [19] stringr_1.2.0      irr_0.84           lpSolve_5.6.13
##
## loaded via a namespace (and not attached):
##  [1] zoo_1.8-0        reshape2_1.4.2   splines_3.4.1    lattice_0.20-35
##  [5] colorspace_1.3-2 htmltools_0.3.6  yaml_2.1.14      rlang_0.1.1
##  [9] nloptr_1.0.4     glue_1.1.1       bindr_0.1        plyr_1.8.4
## [13] munsell_0.4.3    gtable_0.2.0     codetools_0.2-15 evaluate_0.10.1
## [17] labeling_0.3     knitr_1.16       Rcpp_0.12.12     backports_1.1.0
## [21] scales_0.4.1     digest_0.6.12    stringi_1.1.5    grid_3.4.1
## [25] rprojroot_1.2    tools_3.4.1      sandwich_2.4-0   magrittr_1.5
## [29] lazyeval_0.2.0   tibble_1.3.3     pkgconfig_2.0.1  assertthat_0.2.0
## [33] minqa_1.2.4      rmarkdown_1.6    rstudioapi_0.6   R6_2.2.2
## [37] nlme_3.1-131     compiler_3.4.1
```