Data Analyses – Comprehensibility Experiments

Ute Schmid and Christina Zeller
Cognitive Systems Group, University of Bamberg
{ute.schmid,christina.zeller}@uni-bamberg.de

May 16, 2016

Contents

1 Hypotheses

Table 1: Hypotheses concerning comprehensibility, meaningful names, and predicate invention.

	Hypothesis	Confirmation
H1	Comprehensibility manifests itself in high scores and fast solution times	confirmed (OS)
H2	Comprehesibility means to be able to give a meaningful name to a program	confirmed (OS, BA-only score)
H3	Predicate invention helps comprehensibility if it reduces complexity of the program	partially (OS-score, BA-time)
H4	Predicate invention helps comprehensibility if it reduces complexity and if one can give a meaningful name	not confirmed (OS)
H5	If coming up with a meaningful name needs a long time, it will probably be the false concept	confirmed (OS)

Additional:

- H5 holds only for persons who are not absolute beginners.
- Time for comprehension of a program is composed of reading time (longer for longer programs) + perception/attention + cognitive complexity. The recursive ancestor is as short as grandfather but cognitively more complex, therefore, inspection time should be longer.
- ...

2 Imperial Data

2.1 Files and Variables

- Data file: `data_analysis20160511.csv`

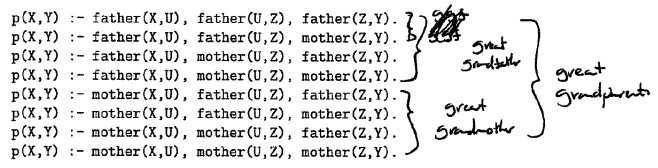
2.2 Participants

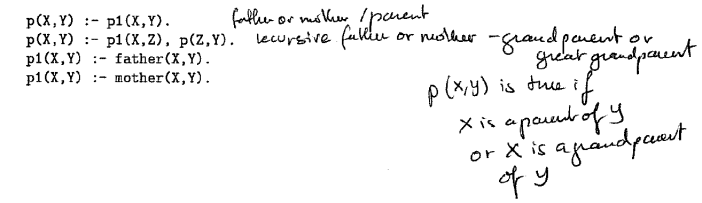
- Data collection was on March 4 2016 in a Prolog class, time: 30 min
- n=21; male = 16, female = 4, others = n.a.
  mean age = 23.69, sd = 2.10 (only 13 gave the age)

2.3 Summary of Results

- From the 21 subjects only 4 answered at least one item wrongly. ↪ ceiling effect, no statistical analysis per factor possible
- If we look at items per problem: from 81 solved problems (4 per subject with 1 problem sheet missing for one subject and 2 problem sheets missing for another subject) for 7 problems were wrongly answered items.
  Strangley, the errors were all made in the "family tree" condition and not in the "abstract tree" condition, 6 times for a problem with predicate invention and one time in a condition without predicate invention. One error was in acestor, 2 for greatgrandparent, 2 for grandparent, 2 for grandfather.
- Naming (last question): worked perfectly for the family tree condition (however one person wrote relative instead of ancestor which is wrong). Naming worked also very well in the abstract condition, e.g. grandfather was described as grandparent with respect to q1.
- Note taking: further qualitative information in the answer sheets, namely: some subjects made notes to support mental interpretation ("calculation") and some wrote meaningful names next to the program, e.g., named predicate p (corresponding to the question to name the predicate), or also named p1 as parent.
  - Astonishingly, about the same percentage of subjects made notes to support mental evaluation in the abstract condition and in the family tree condition (about 25 percent in each condition).
  - One interesting finding is that in the abstract condition with predicate invention only 1 subject made notes whereas without predicate invention 6 problems were solved using notes. Meaning that predicate invention can help to represent the problem in a more compact way and thereby facilitate mental evaluation.
  - Notes with meaningful names: 5 participants gave meaningful names for at least one problem; ID3 for all problems, ID15 for grandfather and ancestor, ID18 for ancestor, ID19 for greatgrandparent and ancestor, ID20 for greatgrandparent.
  - Notes with evaluations: 8 for grandfather, 5 for grandparent, 4 for greatgranparent and ancestor; ID3 for all problems (did also naming)

2.4 Notes with Naming





```
p(X,Y)  :- p1(X,U), p1(U,Z), p1(Z,Y).
p1(X,Y) :- father(X,Y).
p1(X,Y) :- mother(X,Y).
```

2.5 Notes with Evaluations

2.6 Details about Incorrect Answers

```
grandfather (gf): Problem 1
grandparents (gp): Problem 2
great-grandpartend (ggp): Problem 3
ancestor (anc): Problem 4

ID 7, concrete, grandfather, with predicate invention:
wrongly answered gf1,gf5,gf6,gf7

  gf1: What is the result of p(bill,bob)?
  correct: true
  answered: false

  gf5: What is the result of p(john,sam)?
  correct: true
  answered: false

  gf6: What is the complete solution set of p(X,bob)?
  correct: X = bill
  answered: false

  gf7: What is the complete solution set of p(john,X)?
  correct: X = sam; jo
  answered: X = sam

ID 7, concrete, grandparent, with predicate invention:
wrongly answered gp6,gp7

  gp6: What is the complete solution set of p(X,bob)?
  correct: X = bill; alice
  answered: X = bill

  gp7: What is the complete solution set of p(john,X)?
  correct: X = sam; jo
  answered: X = sam

ID 7, concrete, ancestor, with predicate invention:
```

```
wrongly answered anc6

  anc6: What is the complete solution set of p(X,bob)?
  correct: X = jake; matilda; alice; bill; jill; ted
  answered: false


ID 13, concrete, great-grandparent, with predicate invention:
wrongly answered ggp6

  ggp6: What is the complete solution set of p(X,bob)?
  correct: X = jake; matilda
  answered: X = jake


ID 15, concrete, grandfather, with predicate invention:
wrongly answered gf3,gf6

  gf3: What is the result of p(bob,bill)?
  correct: false
  answered: true

  gf6: What is the complete solution set of p(X,bob)?
  correct: X = bill
  answered: two marks set: X = bill and X = bill; alice

ID 15, concrete, grandparend, with predicate invention:
wrongly answered gp3

  gp3: What is the result of p(bob,bill)?
  correct: false
  answered: true

ID 15, concrete, great-grandparent, without predicate invention:
wrongly answered ggp3

  ggp3: What is the result of p(jo,jake)?
  correct: false
  answered: true

ID 15, concrete, ancestor, with predicate invention:
wrongly answered anc3

  anc3: What is the result of p(jo,jake)?
  correct: false
  answered: true


Naming (Can you give p/2 a meaningful name or shortly describe what p/2 does?):

ID 4, abstract, anc, with predicate invention:
  answer: directly related

ID 7, concrete, anc, with predicate invention:
  answer: related

ID 8, abstract, gf, with predicate invention:
  answer: grandparent with parent node of
```

```
ID 10, abstract, gf, without predicate invention:
  answer: p(X,Y) -> X is the grandparent of Y.

ID 14, abstract, gf, without predicate invention:
  answer: 'grandchild/grandparent' relationship (X grandparent of Y)

ID 18, abstract, gf, without predicate invention:
  answer: p(X,Y) is true if X is a parent according to q1 of a parent
       of y accordint to either q1 or q2
```

3 Osnabrueck Data

3.1 Files and Variables

- Data file: `osna_rdata_collapsed2_naming.csv` ('collapsed' means that variables for different conditions were integrated into one new variable and naming is coded as correct/incorrect/don't know)
- The additional variables are created via an R script `osna_preprocessing.r`
- Variable names and coding for the raw data are documented in `osna_variables_listing.pdf`
- Problems are referred to differently in different contexts (due to naming restrictions in soscisurvey):
  - grandfather = Problem 1 = P1 = GF
  - grandparent = Problem 2 = P2 = GP
  - greatgrandparent = Problem 3 = P3 = PP
  - ancestor = Problem 4 = P4 = AC
- Factor names:
  - Name private/public = abstract/concrete = a/c
  - Predicate invention with/without = PIOrder 1/2; **with is coded as '1' for problems 1 and 3 and with is coded as '2' 2 for problems 2 and 4**. This is due to the balanced assignement of the conditions to problems
  - Naming first/last = F/L

  For quick graphs, the default naming is used, so 'a1_F' means private with predicate invention for problem 1 or 3 with naming first, etc.
- Dependent variables:
  - Score is calculated as sum over the 7 closed questions (min: 0, max: 7) (because of a meaningful 0, calculating a log score is not meaningful
  - Time is calculated as sum over seconds per page for closed question including a first page where the problem is presented. This page time is included to control that some participants might study the problem first, others only when they look at the first question. Problem: in the condition naming first, participants might only start to think about the problem when they are on the naming page which is NOT included in the time variable! Analyses are done with log times.
  - Naming question was coded by two persons (Christina, Ute) as correct (1), incorrect (0), dontknow (-1), additionally -2 was coded for don't know answers with more than four symbols (i.e., answers such as -, no, nno are -1; answers such as I have absolutely no idea are -2). Thereby we could discriminate between time effects if wanted.

3.2 Participants

- Data collection was on April 13 and 14 2016 in a cognitive science class.
- Originally: $n = 87$, 299, 323, 344 did not complete (a2L, a1L, a2F), we have no demographical data for these
- Complete: $n = 84$, male = 46, female = 37, other = 1; mean age 24.48, sd = 5.90 (one participant gave age with 71)
- 6 participants were excluded because they answered 'don't know' for more than 50% of the closed items over all problems and did give no answer for problems 3 and 4 (292–a2F, 311–a2L, 312–a2L, 318–a1F, 337–c2F, 342–c1F)
  - $n = 78$
  - male = 43, female = 35, others = 0
  - mean age = 23.55, sd = 2.47 (participant with age 71 no longer in sample)

3.3 Relation Time and Score (H1)

3.3.1 Variables

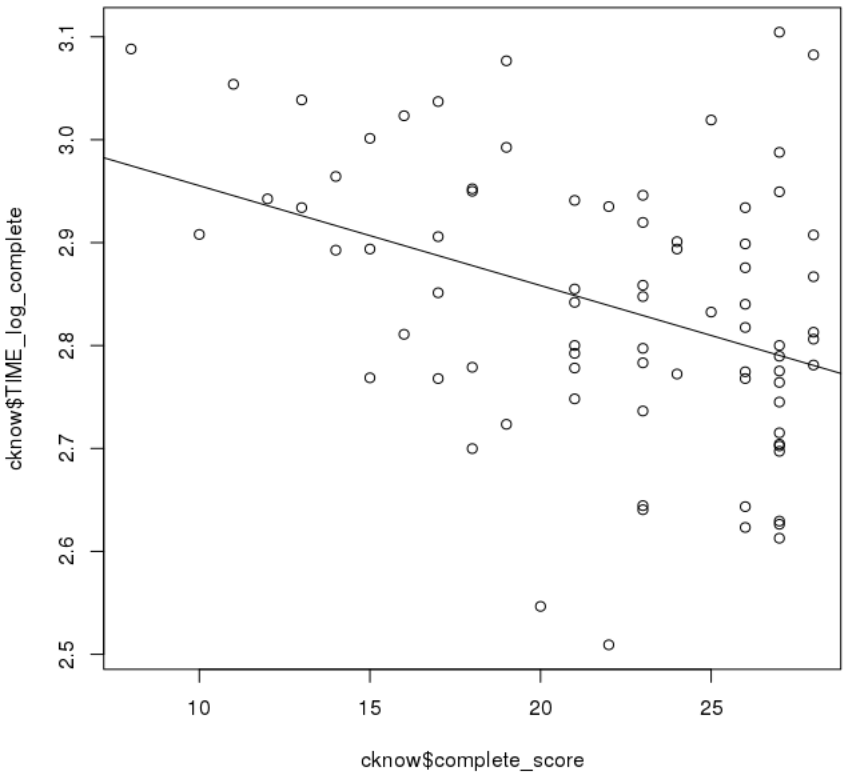
- Variables per problem and aggregated (complete)
- Time is log, score is sum of closed items (0 to 7; wrong and don't know are not discriminated)
- Time_log_complete
- Time_log_P1_noNa (time for naming not included), analogously for P2, P3, P4

3.3.2 Summary of Results

- H1 is confirmed: Participants with higher scores have lower solution times: negative correlation of score and log. time for an aggregation over all 4 problems and for problems 1, 2, 3.

	Grandfather	Grandparent	Greatgrandparent	Ancestor	Complete
r	-0.23	-0.34	-0.36	0.06	-0.38
p	0.04	0.002	0.001	0.59	< 0.001

3.3.3 Graphs



3.3.4 R Output

```
> cor.test(cknow$TIME_log_complete, cknow$complete_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_complete and cknow$complete_score
t = -3.5506, df = 76, p-value = 0.0006635
alternative hypothesis: true correlation is not equal to 0
```

```
95 percent confidence interval:
 -0.5532890 -0.1688423
sample estimates:
       cor
-0.3772007


> cor.test(cknow$TIME_log_P1_noNa, cknow$P1_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_P1_noNa and cknow$P1_score
t = -2.0804, df = 76, p-value = 0.04086
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.43232000 -0.01011181
sample estimates:
       cor
-0.2321203


> cor.test(cknow$TIME_log_P2_noNa, cknow$P2_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_P2_noNa and cknow$P2_score
t = -3.1639, df = 76, p-value = 0.002239
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5239105 -0.1283681
sample estimates:
       cor
-0.3411536


> cor.test(cknow$TIME_log_P3_noNa, cknow$P3_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_P3_noNa and cknow$P3_score
t = -3.4168, df = 76, p-value = 0.001021
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5433205 -0.1549620
sample estimates:
       cor
-0.3649092


> cor.test(cknow$TIME_log_P4_noNa, cknow$P4_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_P4_noNa and cknow$P4_score
t = 0.5435, df = 76, p-value = 0.5884
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1625614  0.2808621
```

```
sample estimates:
       cor
0.06222015
```

3.4 Correct Naming Implies High Performance (H2)

3.4.1 Variables

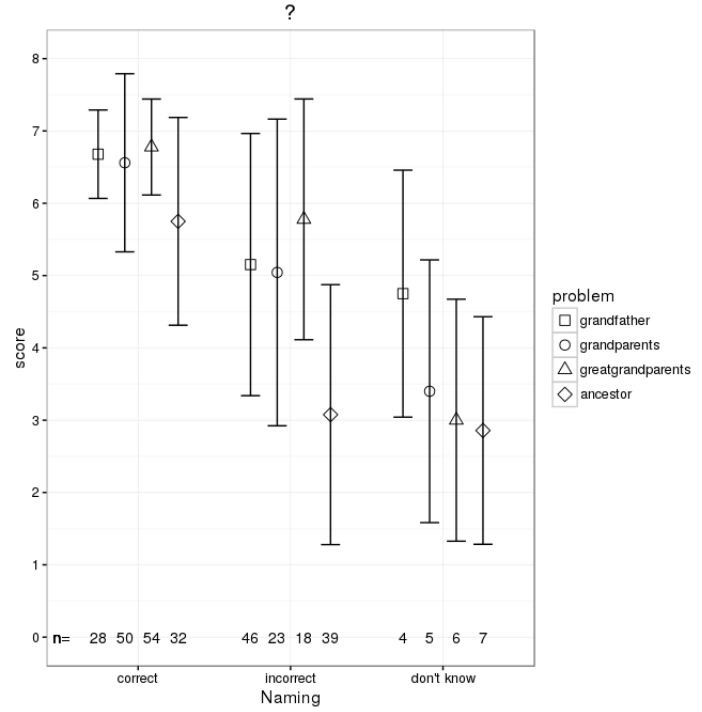
- Levene test shows that variances are homogenous for log time for all problems.
  Levene shows that variances are not homogenous for score.
  ↪ Therefore, we do analyses with GLM (and not Anova) with dummy coding (contrast) for all hypotheses when indicated.
- Naming as independent/predictor variable and score rsp. log time as dependent/criterion variable
- P1 _naming_c3 (1 = correct, 0 = incorrect, -1 = don't know answers) and for the other problems; this variable is dummy coded (contrast)
- To get an impression about the time used per problem, the not log time is given descriptively, statistics is done with log times.

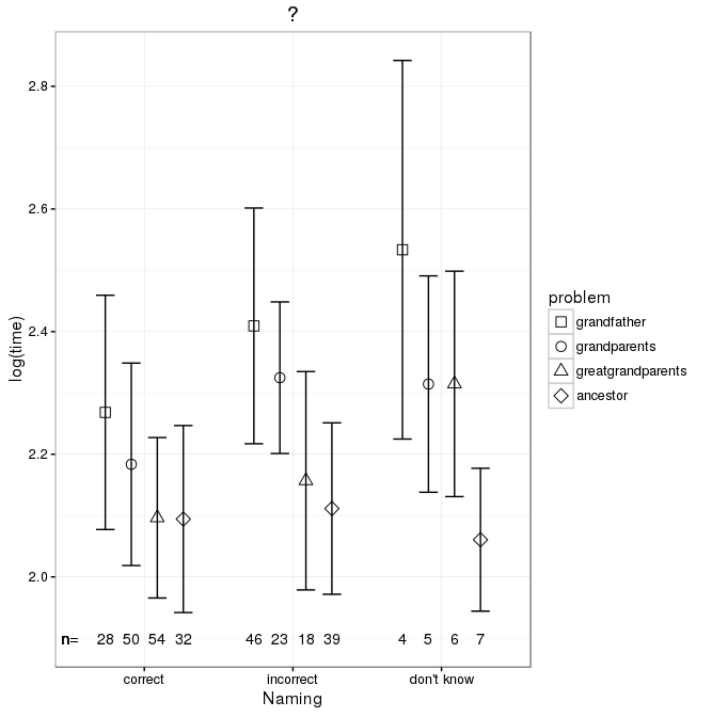
3.4.2 Summary of Results

- H2 confirmed: Participants who name the problem correctly have significantly higher scores and significantly lower answer times (except time ancestor).
- Significances are reported for the contrast between correct/incorrect.

	Correct (1)	Incorrect (0)	DontKnow (-1/-2)	Test
Grandfather	n=28	46	4	
Score	Mean 6.68 (sd = 0.61)	5.15 (1.81)	4.75 (1.71)	-1.53, $p < 0.001$
Time	206.11 (114.67)	281.13 (119.72)	423.25 (353.64)	
LogTime	2.27 (0.19)	2.41 (0.19)	2.53 (0.31)	0.14, p=0.004
Grandparent	50	23	5	
Score	6.56 (1.23)	5.04 (2.12)	3.4 (1.82)	-1.52, $p < 0.001$
Time	164.38 (68.57)	219.35 (60.07)	220.2 (87.71)	
LogTime	2.18 (0.17)	2.32 (0.12)	2.31 (0.18)	0.14, $p < 0.001$
Greatgrandparent	54	18	6	
Score	6.76 (0.66)	5.78 (1.66)	3 (1.67)	-1, $p < 0.001$
Time	130.83 (44.21)	156.61 (78.66)	221.5 (85.48)	
LogTime	2.1 (0.13)	2.16 (0.18)	2.31 (0.18)	0.06, p = 0.13 contr 1/-1, 0.22, p = 0.001
Ancestor	32	39	7	
Score	5.75 (1.44)	3.08 (1.8)	2.86 (1.57)	-2.67, $p < 0.001$
Time	131.59 (44.81)	135.95 (45.14)	118.71 (33.85)	
LogTime	2.09 (0.15)	2.11 (0.14)	2.06 (0.12)	n.s.

3.4.3 Graphs





3.4.4 R Output

```
> leveneTest(cknow$P1_score ~ factor(cknow$P1_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value   Pr(>F)
group  2  11.237 5.39e-05 ***
      75
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> describeBy(cknow$P1_score,factor(cknow$P1_naming_c3))
group: 1
  vars  n mean   sd median trimmed mad min max range  skew kurtosis   se
1    1 28 6.68 0.61      7    6.79   0   5   7     2 -1.61     1.37 0.12
------------------------------------------------------------------------------------
group: -1
  vars n mean   sd median trimmed  mad min max range skew kurtosis se
1    1 4 4.75 1.71    4.5    4.75 1.48   3   7     4 0.28    -1.96 0.85
------------------------------------------------------------------------------------
group: 0
  vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
1    1 46 5.15 1.81      5    5.37 1.48   0   7     7 -0.9     0.09 0.27

> cknow$P1_naming_c3 <- relevel(factor(cknow$P1_naming_c3), "1")

> contrasts(factor(cknow$P1_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$P1_score ~ factor(cknow$P1_naming_c3)))

Call:
lm(formula = cknow$P1_score ~ factor(cknow$P1_naming_c3))

Residuals:
    Min      1Q  Median      3Q     Max
-5.1522 -0.6786  0.3214  0.8478  2.2500

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                        6.6786     0.2818  23.698  < 2e-16 ***
factor(cknow$P1_naming_c3)[T.-1]  -1.9286     0.7971  -2.419 0.018 *
factor(cknow$P1_naming_c3)[T.0]   -1.5264     0.3574  -4.270 5.65e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.491 on 75 degrees of freedom
Multiple R-squared:  0.2095,    Adjusted R-squared:  0.1884
F-statistic: 9.936 on 2 and 75 DF,  p-value: 0.0001486


> leveneTest(cknow$P2_score ~ factor(cknow$P2_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2   6.777 0.00197 **
      75
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> describeBy(cknow$P2_score,factor(cknow$P2_naming_c3))
group: 1
  vars  n mean   sd median trimmed mad min max range  skew kurtosis   se
1    1 50 6.56 1.23      7     6.9   0   1   7     6 -3.19     9.83 0.17
------------------------------------------------------------------------------------
group: -1
  vars n mean   sd median trimmed  mad min max range  skew kurtosis   se
```

```
1    1 5  3.4 1.82      4     3.4 1.48   1   5     4 -0.27    -2.08 0.81
------------------------------------------------------------------------------------------
group: 0
  vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
1    1 23 5.04 2.12      6    5.21 1.48   1   7     6 -0.6    -1.35 0.44

> cknow$P2_naming_c3 <- relevel(factor(cknow$P2_naming_c3), "1")

> contrasts(factor(cknow$P2_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$P2_score ~ factor(cknow$P2_naming_c3)))

Call:
lm(formula = cknow$P2_score ~ factor(cknow$P2_naming_c3))

Residuals:
    Min      1Q  Median      3Q     Max
-5.5600  0.0774  0.4400  0.4400  1.9565

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                        6.5600     0.2230  29.418  < 2e-16 ***
factor(cknow$P2_naming_c3)[T.-1]  -3.1600     0.7396  -4.273 5.6e-05 ***
factor(cknow$P2_naming_c3)[T.0]   -1.5165     0.3973  -3.817 0.000276 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.577 on 75 degrees of freedom
Multiple R-squared:  0.2726,    Adjusted R-squared:  0.2532
F-statistic: 14.06 on 2 and 75 DF,  p-value: 6.543e-06


> leveneTest(cknow$P3_score ~ factor(cknow$P3_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  2  8.7709 0.0003777 ***
      75
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> describeBy(cknow$P3_score,factor(cknow$P3_naming_c3))
group: 1
  vars  n mean   sd median trimmed mad min max range  skew kurtosis   se
1    1 54 6.78 0.66      7    6.93   0   3   7     4 -3.92    17.64 0.09
------------------------------------------------------------------------------------------
group: -1
  vars n mean   sd median trimmed  mad min max range skew kurtosis se
1    1 6    3 1.67    2.5       3 1.48   1   5     4 0.21    -1.94 0.68
------------------------------------------------------------------------------------------
group: 0
  vars  n mean   sd median trimmed mad min max range  skew kurtosis   se
1    1 18 5.78 1.66      7    5.94   0   2   7     5 -0.97    -0.49 0.39

> cknow$P3_naming_c3 <- relevel(factor(cknow$P3_naming_c3), "1")
```

```
> contrasts(factor(cknow$P3_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$P3_score ~ factor(cknow$P3_naming_c3)))

Call:
lm(formula = cknow$P3_score ~ factor(cknow$P3_naming_c3))

Residuals:
    Min      1Q  Median      3Q     Max
-3.7778  0.2222  0.2222  0.2222  2.0000

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       6.7778     0.1444  46.939  < 2e-16 ***
factor(cknow$P3_naming_c3)[T.-1] -3.7778     0.4566  -8.273 3.6e-12 ***
factor(cknow$P3_naming_c3)[T.0]  -1.0000     0.2888  -3.463 0.000887 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.061 on 75 degrees of freedom
Multiple R-squared:  0.494,    Adjusted R-squared:  0.4805
F-statistic:  36.6 on 2 and 75 DF,  p-value: 8.073e-12


> leveneTest(cknow$P4_score ~ factor(cknow$P4_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.2925 0.2806
      75

> describeBy(cknow$P4_score,factor(cknow$P4_naming_c3))
group: 1
  vars  n mean   sd median trimmed mad min max range  skew kurtosis   se
1    1 32 5.75 1.44      6    6.04   0   1   7     6 -1.85     2.88 0.25
------------------------------------------------------------------------------------
group: -1
  vars n mean   sd median trimmed  mad min max range  skew kurtosis   se
1    1 7 2.86 1.57      3    2.86 1.48   1   5     4 -0.02    -1.86 0.59
------------------------------------------------------------------------------------
group: 0
  vars  n mean  sd median trimmed mad min max range skew kurtosis se
1    1 39 3.08 1.8      2    2.97   0   0   7     7 0.79     -0.6 0.29

> cknow$P4_naming_c3 <- relevel(factor(cknow$P4_naming_c3), "1")

> contrasts(factor(cknow$P4_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$P4_score ~ factor(cknow$P4_naming_c3)))
```

```
Call:
lm(formula = cknow$P4_score ~ factor(cknow$P4_naming_c3))

Residuals:
   Min     1Q Median     3Q    Max
-4.750 -1.077  0.250  1.143  3.923

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       5.7500     0.2898  19.838  < 2e-16 ***
factor(cknow$P4_naming_c3)[T.-1]  -2.8929     0.6842  -4.228 6.57e-05 ***
factor(cknow$P4_naming_c3)[T.0]   -2.6731     0.3911  -6.835 1.89e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.64 on 75 degrees of freedom
Multiple R-squared:  0.4072,    Adjusted R-squared:  0.3914
F-statistic: 25.76 on 2 and 75 DF,  p-value: 3.039e-09

> leveneTest(cknow$TIME_log_P1_noNa ~ factor(cknow$P1_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.5631 0.5718
      75

> describeBy(cknow$TIME_log_P1_noNa,factor(cknow$P1_naming_c3))
group: 1
  vars  n mean   sd median trimmed  mad  min  max range skew kurtosis   se
1    1 28 2.27 0.19   2.24    2.26 0.12 1.96 2.82  0.86 0.78 0.64 0.04
------------------------------------------------------------------------------------
group: -1
  vars n mean   sd median trimmed  mad  min  max range skew kurtosis   se
1    1 4 2.53 0.31   2.44    2.53 0.16 2.28 2.98  0.69 0.57    -1.81 0.15
------------------------------------------------------------------------------------
group: 0
  vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
1    1 46 2.41 0.19   2.42    2.41 0.18 1.9 2.8  0.91 -0.32    -0.22 0.03

> cknow$P1_naming_c3 <- relevel(factor(cknow$P1_naming_c3), "1")

> contrasts(factor(cknow$P1_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$TIME_log_P1_noNa ~ factor(cknow$P1_naming_c3)))

Call:
lm(formula = cknow$TIME_log_P1_noNa ~ factor(cknow$P1_naming_c3))

Residuals:
     Min       1Q   Median       3Q      Max
-0.51172 -0.09919 -0.01409  0.12148  0.54794

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      2.26830    0.03738  60.686  < 2e-16 ***
```

```
factor(cknow$P1_naming_c3)[T.-1]  0.26522    0.10572   2.509 0.01428 *
factor(cknow$P1_naming_c3)[T.0]   0.14105    0.04741   2.975 0.00394 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1978 on 75 degrees of freedom
Multiple R-squared:  0.137,    Adjusted R-squared:  0.114
F-statistic: 5.955 on 2 and 75 DF,  p-value: 0.003978


> leveneTest(cknow$TIME_log_P2_noNa ~ factor(cknow$P2_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.1496 0.3223
      75

> describeBy(cknow$TIME_log_P2_noNa,factor(cknow$P2_naming_c3))
group: -1
  vars n mean   sd median trimmed  mad  min  max range skew kurtosis   se
1    1 5 2.31 0.18   2.31    2.31 0.24 2.13 2.52  0.39 0.05    -2.16 0.08
------------------------------------------------------------------------------------------
group: 0
  vars  n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
1    1 23 2.32 0.12   2.32    2.33 0.14 2.08 2.52  0.44 -0.26 -1.03 0.03
------------------------------------------------------------------------------------------
group: 1
  vars  n mean   sd median trimmed  mad min  max range skew kurtosis   se
1    1 50 2.18 0.17   2.18    2.17 0.17 1.9 2.59  0.69 0.42    -0.51 0.02

> cknow$P2_naming_c3 <- relevel(factor(cknow$P2_naming_c3), "1")

> contrasts(factor(cknow$P2_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$TIME_log_P2_noNa ~ factor(cknow$P2_naming_c3)))

Call:
lm(formula = cknow$TIME_log_P2_noNa ~ factor(cknow$P2_naming_c3))

Residuals:
     Min       1Q   Median       3Q      Max
-0.28051 -0.11177 -0.00506  0.10289  0.40746

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      2.18360    0.02189  99.774  < 2e-16 ***
factor(cknow$P2_naming_c3)[T.-1]  0.13084    0.07259   1.803 0.075482 .
factor(cknow$P2_naming_c3)[T.0]   0.14122    0.03899   3.622 0.000529 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1548 on 75 degrees of freedom
Multiple R-squared:  0.1627,    Adjusted R-squared:  0.1404
F-statistic: 7.288 on 2 and 75 DF,  p-value: 0.001282
```

```
> leveneTest(cknow$TIME_log_P3_noNa ~ factor(cknow$P3_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.0855  0.343
      75

> describeBy(cknow$TIME_log_P3_noNa,factor(cknow$P3_naming_c3))
group: -1
  vars n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
1    1 6 2.31 0.18   2.36    2.31 0.16 2.07 2.47  0.41 -0.22 -2.05 0.08
------------------------------------------------------------------------------------
group: 0
  vars  n mean   sd median trimmed  mad min  max range skew kurtosis   se
1    1 18 2.16 0.18   2.15    2.14 0.11 1.9 2.63  0.73 0.61     0.65 0.04
------------------------------------------------------------------------------------
group: 1
  vars  n mean   sd median trimmed  mad  min  max range skew kurtosis   se
1    1 54  2.1 0.13    2.1    2.09 0.12 1.83 2.54  0.71 0.42 0.73 0.02

> cknow$P3_naming_c3 <- relevel(factor(cknow$P3_naming_c3), "1")

> contrasts(factor(cknow$P3_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$TIME_log_P3_noNa ~ factor(cknow$P3_naming_c3)))

Call:
lm(formula = cknow$TIME_log_P3_noNa ~ factor(cknow$P3_naming_c3))

Residuals:
     Min       1Q   Median       3Q      Max
-0.27032 -0.08356 -0.00299  0.08315  0.47465

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                       2.09640    0.01997 104.986 <2e-16 ***
factor(cknow$P3_naming_c3)[T.-1]  0.21837    0.06315   3.458 0.0009 ***
factor(cknow$P3_naming_c3)[T.0]   0.06040    0.03994   1.512 0.1346
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1467 on 75 degrees of freedom
Multiple R-squared:  0.1469,    Adjusted R-squared:  0.1242
F-statistic: 6.458 on 2 and 75 DF,  p-value: 0.002585


> leveneTest(cknow$TIME_log_P4_noNa ~ factor(cknow$P4_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.3772 0.6871
      75

> describeBy(cknow$TIME_log_P4_noNa,factor(cknow$P4_naming_c3))
group: -1
```

```
  vars n mean   sd median trimmed mad  min  max range skew kurtosis   se
1    1 7 2.06 0.12   2.07    2.06 0.1 1.91 2.26  0.36 0.38    -1.16 0.04
-------------------------------------------------------------------------------------------
group: 0
  vars  n mean   sd median trimmed  mad min  max range  skew kurtosis   se
1    1 39 2.11 0.14   2.11    2.11 0.09 1.8 2.46  0.66 -0.02 0.13 0.02
-------------------------------------------------------------------------------------------
group: 1
  vars  n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
1    1 32 2.09 0.15   2.11     2.1 0.15 1.75 2.43  0.68 -0.36 -0.04 0.03

> cknow$P4_naming_c3 <- relevel(factor(cknow$P4_naming_c3), "1")

> contrasts(factor(cknow$P4_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$TIME_log_P4_noNa ~ factor(cknow$P4_naming_c3)))

Call:
lm(formula = cknow$TIME_log_P4_noNa ~ factor(cknow$P4_naming_c3))

Residuals:
     Min       1Q   Median       3Q      Max
-0.34616 -0.06182  0.00254  0.08674  0.34493

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      2.09434    0.02538  82.527 <2e-16 ***
factor(cknow$P4_naming_c3)[T.-1] -0.03372    0.05990  -0.563 0.575
factor(cknow$P4_naming_c3)[T.0]   0.01709    0.03424   0.499 0.619
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1436 on 75 degrees of freedom
Multiple R-squared:  0.01086,    Adjusted R-squared:  -0.01551
F-statistic: 0.4118 on 2 and 75 DF,  p-value: 0.6639

> describeBy(cknow$TIME_P1_noNa,factor(cknow$P1_naming_c3))
group: -1
  vars n   mean     sd median trimmed   mad min max range skew kurtosis     se
1    1 4 423.25 353.64  276.5  423.25 92.66 192 948   756  0.7 -1.72 176.82
-------------------------------------------------------------------------------------------
group: 0
  vars  n   mean     sd median trimmed    mad min max range skew kurtosis    se
1    1 46 281.13 119.72    261  272.13 117.87  79 635   556 0.73 0.22 17.65
-------------------------------------------------------------------------------------------
group: 1
  vars  n   mean     sd median trimmed   mad min max range skew kurtosis    se
1    1 28 206.11 114.67  172.5   189.5 51.89  91 655   564 2.25 5.87 21.67

> describeBy(cknow$TIME_P2_noNa,factor(cknow$P2_naming_c3))
group: -1
  vars n  mean    sd median trimmed   mad min max range skew kurtosis    se
1    1 5 220.2 87.71    202   220.2 99.33 135 328   193 0.16 -2.14 39.23
-------------------------------------------------------------------------------------------
```

```
group: 0
  vars  n   mean    sd median trimmed   mad min max range skew kurtosis    se
1    1 23 219.35 60.07    210  217.74 74.13 121 334   213 0.17 -1.04 12.53
------------------------------------------------------------------------------------
group: 1
  vars  n   mean    sd median trimmed   mad min max range skew kurtosis  se
1    1 50 164.38 68.57  151.5  155.05 60.79  80 390   310 1.26 1.37 9.7

> describeBy(cknow$TIME_P3_noNa,factor(cknow$P3_naming_c3))
group: -1
  vars n  mean    sd median trimmed   mad min max range  skew kurtosis    se
1    1 6 221.5 85.58  238.5   221.5 88.21 117 298   181 -0.12 -2.17 34.94
------------------------------------------------------------------------------------
group: 0
  vars  n   mean    sd median trimmed   mad min max range skew kurtosis    se
1    1 18 156.61 78.66  142.5   144.5 40.77  79 428   349 2.13 5.07 18.54
------------------------------------------------------------------------------------
group: 1
  vars  n   mean    sd median trimmed   mad min max range skew kurtosis   se
1    1 54 130.83 44.21    125   126.7 32.62  67 344   277 2.02 7.66 6.02

> describeBy(cknow$TIME_P4_noNa,factor(cknow$P4_naming_c3))
group: -1
  vars n   mean    sd median trimmed  mad min max range skew kurtosis    se
1    1 7 118.71 33.85    117  118.71 25.2  81 184   103 0.75 -0.74 12.79
------------------------------------------------------------------------------------
group: 0
  vars  n   mean    sd median trimmed   mad min max range skew kurtosis   se
1    1 39 135.95 45.14    129  132.48 23.72  63 286   223 1.06 1.59 7.23
------------------------------------------------------------------------------------
group: 1
  vars  n   mean    sd median trimmed   mad min max range skew kurtosis   se
1    1 32 131.59 44.81  127.5  129.73 42.25  56 267   211 0.69 0.78 7.92
```

3.5 Effect of Predicate Invention on Comprehensibility (H3)

3.5.1 Variables

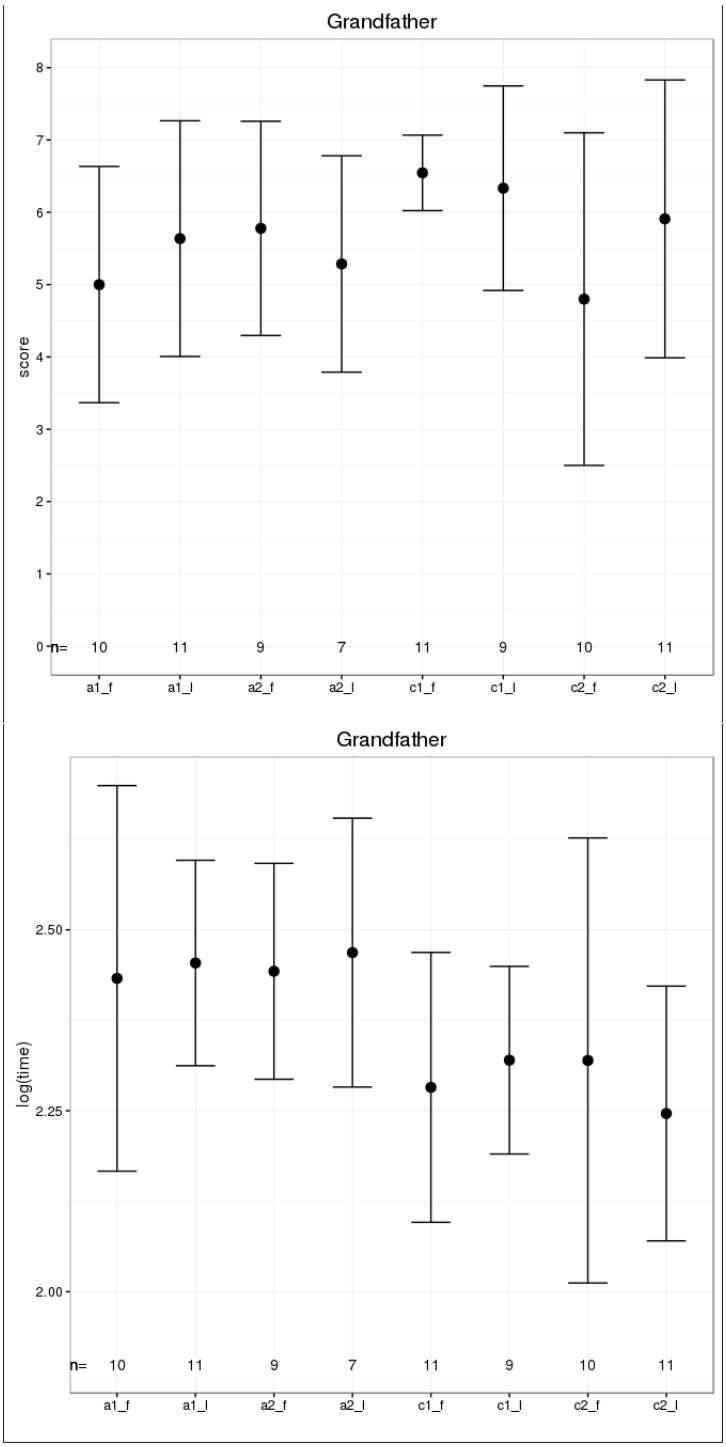
- Factors: Predicate invention with/without, Names: public/private; Naming First/Last (remember, abbreviated with a/c etc., see Section 3.1)
- Dependent variables: score and log time
- We tested whether we get clearer results if we exlude the naming-dontknow participants, but the results remain the same
- Attention: with PI is '1' for problems 1 and 3 and '2' for the other two problems.
  Graphs are currently default labeled: sequence is public the first 4, private the last 4, always with predicate invention before without (but once), always naming first before naming last.

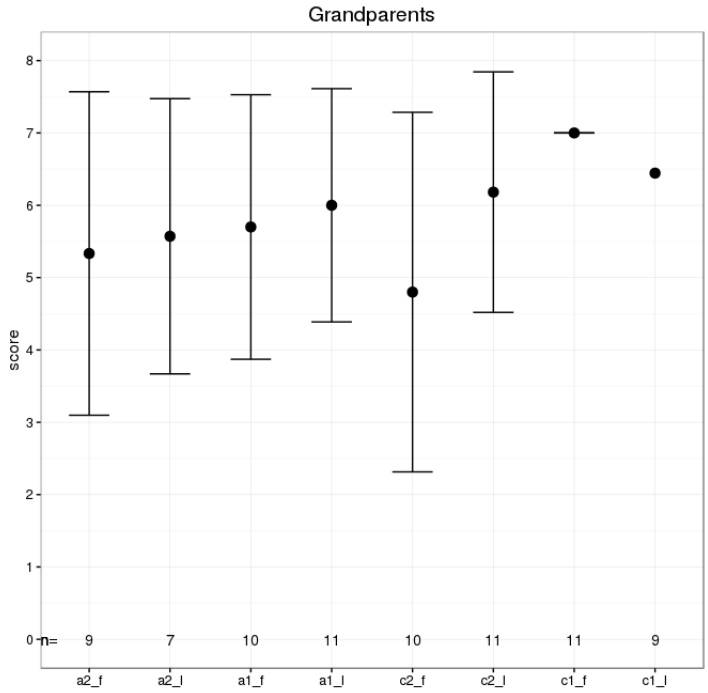
3.5.2 Summary of Results

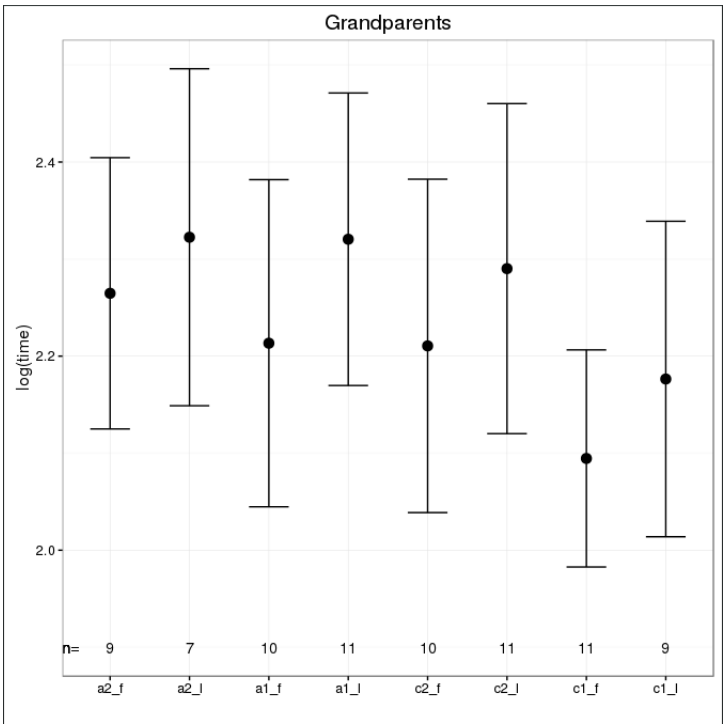
- H3 only concerns the greatgrandparent problem because there the code without predicate invention is significantly longer than the code with predicate invention. The hypothesis is partially confirmed: For the private naming condition with naming last the score is mariginally significantly higher for the condition with predicate invention.
- In addition we tested the effects of all three factors on score and log time:

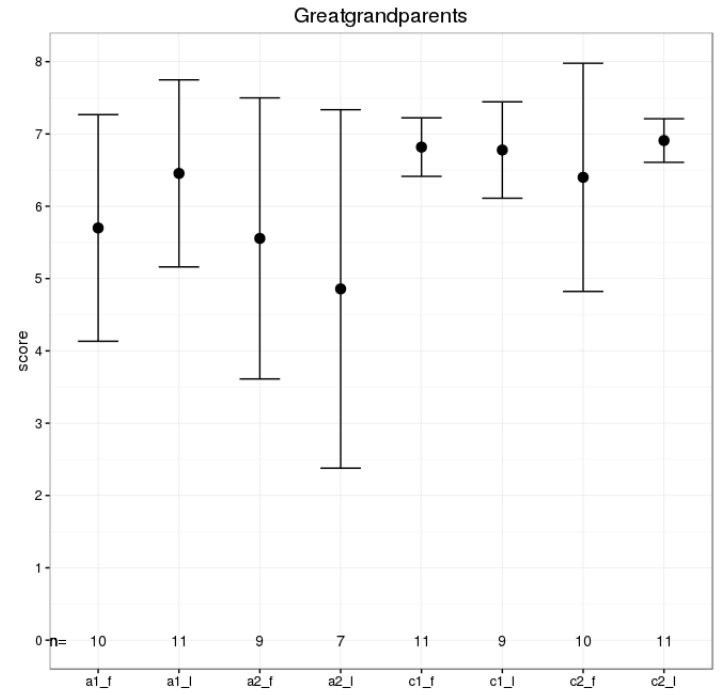
- grandfather and greatgrandparent: marginally significant main effect of time for public vs. private names (with public names is faster)
- grandfather: public/private is a significant main effect on score and interaction of public/private and predicate invention is also significant; public profits from predicate invention, private not?
- greatgrandparent: significant main effect for public/private

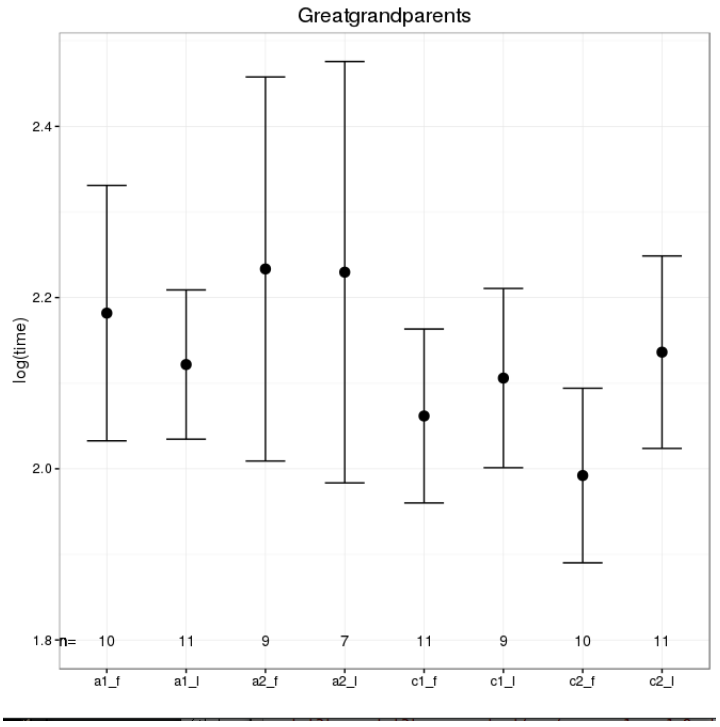
3.5.3 Graphs

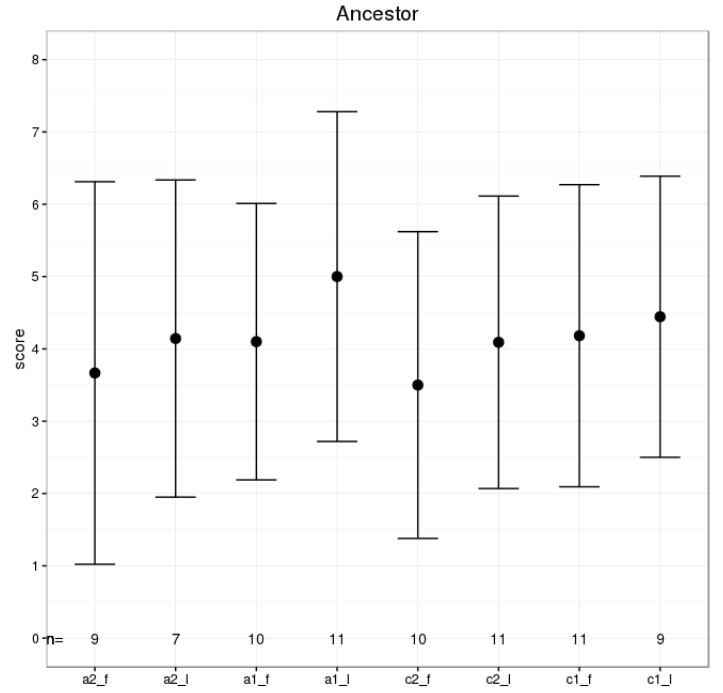


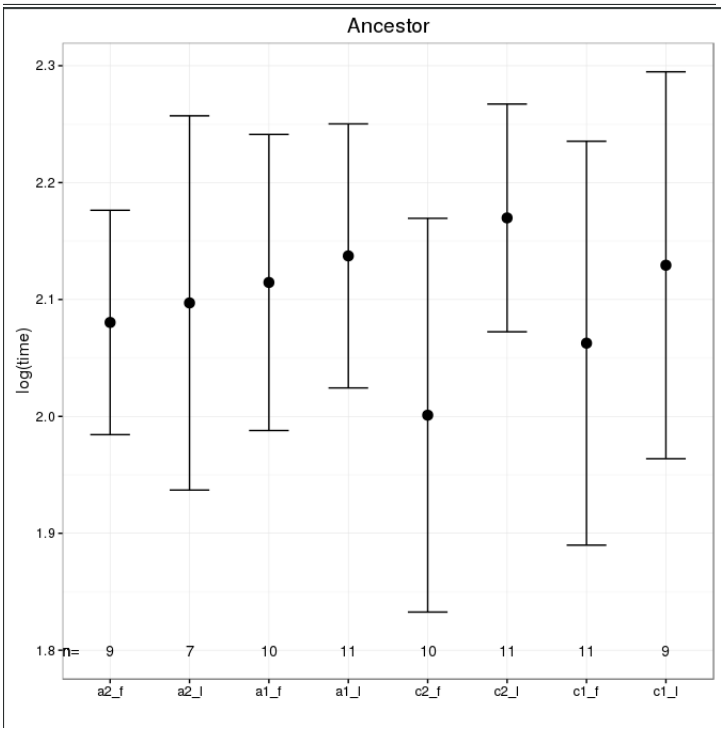

Grandparents
score
8
7
6
5
4
3
2
1
0
n= 9 7 10 11 10 11 11 9
a2_f
a2_l
a1_f
a1_l
c2_f
c2_l
c1_f
c1_l


Grandparents
log(time)
2.4
2.2
2.0
n= 9 7 10 11 10 11 11 9
a2_f
a2_l
a1_f
a1_l
c2_f
c2_l
c1_f
c1_l


Greatgrandparents
score
8
7
6
5
4
3
2
1
0
n=
10
11
9
7
11
9
10
11
a1_f
a1_l
a2_f
a2_l
c1_f
c1_l
c2_f
c2_l


Greatgrandparents
log(time)
2.4
2.2
2.0
1.8
n=
10
11
9
7
11
9
10
11
a1_f
a1_l
a2_f
a2_l
c1_f
c1_l
c2_f
c2_l





3.5.4 R Output

```
# Score for PI with/without Greatgrandparents (private, naming last)

> cknow_al$PIorder <- relevel(factor(cknow_al$PIorder), "1")

> contrasts(cknow_al$PIorder)
  [T.2]
1     0
2     1

> summary(lm(cknow_al$P3_score ~ cknow_al$PIorder))
```

```
Call:
lm(formula = cknow_al$P3_score ~ cknow_al$PIorder)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8571  0.1429  0.5455  0.5455  2.1429

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            6.4545     0.5518  11.698 2.97e-09 ***
cknow_al$PIorder[T.2]  -1.5974     0.8848  -1.805   0.0899 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.83 on 16 degrees of freedom
Multiple R-squared:  0.1692,    Adjusted R-squared:  0.1173
F-statistic: 3.259 on 1 and 16 DF,  p-value: 0.08987

> # PIorder 1: PI Problem 1 and 3; PIorder 2: PI Problem 2 and 4

> cknow$PIorder <- relevel(factor(cknow$PIorder), "1")

> contrasts(cknow$PIorder)
  [T.2]
1     0
2     1

> summary(lm(cknow$TIME_log_P1_noNa ~ cknow$PIorder * cknow$AbsCon * cknow$FiLa))

Call:
lm(formula = cknow$TIME_log_P1_noNa ~ cknow$PIorder * cknow$AbsCon *
    cknow$FiLa)

Residuals:
     Min       1Q   Median       3Q      Max
-0.53528 -0.09815 -0.02544  0.10249  0.65729

Coefficients:
                                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                                             2.432903 0.064004  38.012   <2e-16 ***
cknow$PIorder[T.2]                                      0.009794 0.092995   0.105   0.9164
cknow$AbsCon[T.c]                                      -0.150570 0.088434  -1.703   0.0931 .
cknow$FiLa[T.l]                                         0.021240 0.088434   0.240   0.8109
cknow$PIorder[T.2]:cknow$AbsCon[T.c]                    0.027388 0.128330   0.213   0.8316
cknow$PIorder[T.2]:cknow$FiLa[T.l]                      0.004562 0.134997   0.034   0.9731
cknow$AbsCon[T.c]:cknow$FiLa[T.l]                       0.016343 0.126871   0.129   0.8979
cknow$PIorder[T.2]:cknow$AbsCon[T.c]:cknow$FiLa[T.l] -0.115466 0.185258  -0.623   0.5351
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2024 on 70 degrees of freedom
Multiple R-squared:  0.1566,    Adjusted R-squared:  0.07221
F-statistic: 1.856 on 7 and 70 DF,  p-value: 0.09011


> cknow$PIorder <- relevel(factor(cknow$PIorder), "2")
```

```
> contrasts(cknow$PIorder)
  [T.1]
2     0
1     1

> summary(lm(cknow$TIME_log_P2_noNa ~ cknow$PIorder * cknow$AbsCon * cknow$FiLa))

Call:
lm(formula = cknow$TIME_log_P2_noNa ~ cknow$PIorder * cknow$AbsCon *
    cknow$FiLa)

Residuals:
    Min      1Q  Median      3Q     Max
-0.3139 -0.1354  0.0105  0.1166  0.3132

Coefficients:
                                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                                              2.26472 0.05212  43.451   <2e-16 ***
cknow$PIorder[T.1]                                      -0.05140 0.07184  -0.715    0.477
cknow$AbsCon[T.c]                                       -0.05417 0.07184  -0.754    0.453
cknow$FiLa[T.l]                                          0.05780 0.07880   0.733    0.466
cknow$PIorder[T.1]:cknow$AbsCon[T.c]                    -0.06459 0.09914  -0.651    0.517
cknow$PIorder[T.1]:cknow$FiLa[T.l]                       0.04933 0.10429   0.473    0.638
cknow$AbsCon[T.c]:cknow$FiLa[T.l]                        0.02183 0.10429   0.209    0.835
cknow$PIorder[T.1]:cknow$AbsCon[T.c]:cknow$FiLa[T.l] -0.04710 0.14312  -0.329    0.743
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1564 on 70 degrees of freedom
Multiple R-squared:  0.2022,    Adjusted R-squared:  0.1224
F-statistic: 2.534 on 7 and 70 DF,  p-value: 0.02202


> cknow$PIorder <- relevel(factor(cknow$PIorder), "1")

> contrasts(cknow$PIorder)
  [T.2]
1     0
2     1

> summary(lm(cknow$TIME_log_P3_noNa ~ cknow$PIorder * cknow$AbsCon * cknow$FiLa))

Call:
lm(formula = cknow$TIME_log_P3_noNa ~ cknow$PIorder * cknow$AbsCon *
    cknow$FiLa)

Residuals:
     Min       1Q   Median       3Q      Max
-0.35575 -0.08748  0.02093  0.08693  0.39804

Coefficients:
                                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                                              2.18182 0.04556  47.887   <2e-16 ***
cknow$PIorder[T.2]                                       0.05158 0.06620   0.779   0.4385
cknow$AbsCon[T.c]                                       -0.12016 0.06295  -1.909   0.0604 .
cknow$FiLa[T.l]                                         -0.06007 0.06295  -0.954   0.3433
cknow$PIorder[T.2]:cknow$AbsCon[T.c]                    -0.12113 0.09135  -1.326   0.1892
cknow$PIorder[T.2]:cknow$FiLa[T.l]                       0.05630 0.09610   0.586   0.5598
```

```
cknow$AbsCon[T.c]:cknow$FiLa[T.l]                     0.10433 0.09032   1.155   0.2519
cknow$PIorder[T.2]:cknow$AbsCon[T.c]:cknow$FiLa[T.l]  0.04353 0.13188   0.330   0.7423
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1441 on 70 degrees of freedom
Multiple R-squared:  0.2324,    Adjusted R-squared:  0.1556
F-statistic: 3.027 on 7 and 70 DF,  p-value: 0.007731


> cknow$PIorder <- relevel(factor(cknow$PIorder), "2")

> contrasts(cknow$PIorder)
  [T.1]
2     0
1     1

> summary(lm(cknow$TIME_log_P4_noNa ~ cknow$PIorder * cknow$AbsCon * cknow$FiLa))

Call:
lm(formula = cknow$TIME_log_P4_noNa ~ cknow$PIorder * cknow$AbsCon *
    cknow$FiLa)

Residuals:
     Min       1Q   Median       3Q      Max
-0.33363 -0.07600 -0.00387  0.07803  0.39373

Coefficients:
                                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                                           2.080367 0.046583  44.659   <2e-16 ***
cknow$PIorder[T.1]                                    0.034202 0.064210   0.533    0.596
cknow$AbsCon[T.c]                                    -0.079338 0.064210  -1.236    0.221
cknow$FiLa[T.l]                                       0.016696 0.070427   0.237    0.813
cknow$PIorder[T.1]:cknow$AbsCon[T.c]                  0.027406 0.088608   0.309    0.758
cknow$PIorder[T.1]:cknow$FiLa[T.l]                    0.006032 0.093211   0.065    0.949
cknow$AbsCon[T.c]:cknow$FiLa[T.l]                     0.152057 0.093211   1.631    0.107
cknow$PIorder[T.1]:cknow$AbsCon[T.c]:cknow$FiLa[T.l] -0.108125 0.127915  -0.845    0.401
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1397 on 70 degrees of freedom
Multiple R-squared:  0.1251,    Adjusted R-squared:  0.03766
F-statistic:  1.43 on 7 and 70 DF,  p-value: 0.2071


> cknow$PIorder <- relevel(factor(cknow$PIorder), "1")

> contrasts(cknow$PIorder)
  [T.2]
1     0
2     1

> summary(lm(cknow$P1_score ~ cknow$PIorder * cknow$AbsCon * cknow$FiLa))

Call:
lm(formula = cknow$P1_score ~ cknow$PIorder * cknow$AbsCon *
    cknow$FiLa)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.9091 -0.6364  0.4545  1.0909  2.2000

Coefficients:
                                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                                               5.0000 0.5138   9.732 1.22e-14 ***
cknow$PIorder[T.2]                                        0.7778 0.7465   1.042   0.3010
cknow$AbsCon[T.c]                                         1.5455 0.7099   2.177   0.0328 *
cknow$FiLa[T.l]                                           0.6364 0.7099   0.896   0.3731
cknow$PIorder[T.2]:cknow$AbsCon[T.c]                     -2.5232 1.0301  -2.449   0.0168 *
cknow$PIorder[T.2]:cknow$FiLa[T.l]                       -1.1284 1.0836  -1.041   0.3013
cknow$AbsCon[T.c]:cknow$FiLa[T.l]                        -0.8485 1.0184  -0.833   0.4076
cknow$PIorder[T.2]:cknow$AbsCon[T.c]:cknow$FiLa[T.l]      2.4496 1.4871   1.647   0.1040
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.625 on 70 degrees of freedom
Multiple R-squared:  0.1243,    Adjusted R-squared:  0.03671
F-statistic: 1.419 on 7 and 70 DF,  p-value: 0.2115


> cknow$PIorder <- relevel(factor(cknow$PIorder), "2")

> contrasts(cknow$PIorder)
  [T.1]
2     0
1     1

> summary(lm(cknow$P2_score ~ cknow$PIorder * cknow$AbsCon * cknow$FiLa))

Call:
lm(formula = cknow$P2_score ~ cknow$PIorder * cknow$AbsCon *
    cknow$FiLa)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7000  0.0000  0.5556  1.0000  2.2000

Coefficients:
                                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                                               5.3333 0.5953   8.959 3.15e-13 ***
cknow$PIorder[T.1]                                        0.3667 0.8206   0.447    0.656
cknow$AbsCon[T.c]                                        -0.5333 0.8206  -0.650    0.518
cknow$FiLa[T.l]                                           0.2381 0.9000   0.265    0.792
cknow$PIorder[T.1]:cknow$AbsCon[T.c]                      1.8333 1.1324   1.619    0.110
cknow$PIorder[T.1]:cknow$FiLa[T.l]                        0.0619 1.1912   0.052    0.959
cknow$AbsCon[T.c]:cknow$FiLa[T.l]                         1.1437 1.1912   0.960    0.340
cknow$PIorder[T.1]:cknow$AbsCon[T.c]:cknow$FiLa[T.l]     -1.9993 1.6347  -1.223    0.225
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.786 on 70 degrees of freedom
Multiple R-squared:  0.1291,    Adjusted R-squared:  0.04202
F-statistic: 1.482 on 7 and 70 DF,  p-value: 0.1878


> cknow$PIorder <- relevel(factor(cknow$PIorder), "1")
```

```
> contrasts(cknow$PIorder)
  [T.2]
1     0
2     1

> summary(lm(cknow$P3_score ~ cknow$PIorder * cknow$AbsCon * cknow$FiLa))

Call:
lm(formula = cknow$P3_score ~ cknow$PIorder * cknow$AbsCon *
    cknow$FiLa)

Residuals:
    Min      1Q  Median      3Q     Max
-4.4000  0.0909  0.2222  0.5864  2.1429

Coefficients:
                                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                                           5.7000 0.4382  13.008   <2e-16 ***
cknow$PIorder[T.2]                                   -0.1444 0.6367  -0.227    0.821
cknow$AbsCon[T.c]                                     1.1182 0.6054   1.847    0.069 .
cknow$FiLa[T.l]                                       0.7545 0.6054   1.246    0.217
cknow$PIorder[T.2]:cknow$AbsCon[T.c]                 -0.2737 0.8786  -0.312    0.756
cknow$PIorder[T.2]:cknow$FiLa[T.l]                   -1.4530 0.9242  -1.572    0.120
cknow$AbsCon[T.c]:cknow$FiLa[T.l]                    -0.7949 0.8686  -0.915    0.363
cknow$PIorder[T.2]:cknow$AbsCon[T.c]:cknow$FiLa[T.l]  2.0025 1.2683   1.579    0.119
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.386 on 70 degrees of freedom
Multiple R-squared:  0.1945,    Adjusted R-squared:  0.114
F-statistic: 2.415 on 7 and 70 DF,  p-value: 0.02831


> cknow$PIorder <- relevel(factor(cknow$PIorder), "2")

> contrasts(cknow$PIorder)
  [T.1]
2     0
1     1

> summary(lm(cknow$P4_score ~ cknow$PIorder * cknow$AbsCon * cknow$FiLa))

Call:
lm(formula = cknow$P4_score ~ cknow$PIorder * cknow$AbsCon *
    cknow$FiLa)

Residuals:
   Min     1Q Median     3Q    Max
-4.000 -2.091  0.200  1.900  3.333

Coefficients:
                                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                                           3.6667 0.7183   5.105 2.73e-06 ***
cknow$PIorder[T.1]                                    0.4333 0.9901   0.438    0.663
cknow$AbsCon[T.c]                                    -0.1667 0.9901  -0.168    0.867
cknow$FiLa[T.l]                                       0.4762 1.0859   0.439    0.662
cknow$PIorder[T.1]:cknow$AbsCon[T.c]                  0.2485 1.3663   0.182    0.856
```

```
cknow$PIorder[T.1]:cknow$FiLa[T.l]                      0.4238 1.4372   0.295    0.769
cknow$AbsCon[T.c]:cknow$FiLa[T.l]                       0.1147 1.4372   0.080    0.937
cknow$PIorder[T.1]:cknow$AbsCon[T.c]:cknow$FiLa[T.l]  -0.7521 1.9723  -0.381    0.704
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.155 on 70 degrees of freedom
Multiple R-squared:  0.04448,    Adjusted R-squared:  -0.05108
F-statistic: 0.4655 on 7 and 70 DF,  p-value: 0.8562
```

3.6 Correct Naming Helps to Profit from Predicate Invention (H4)

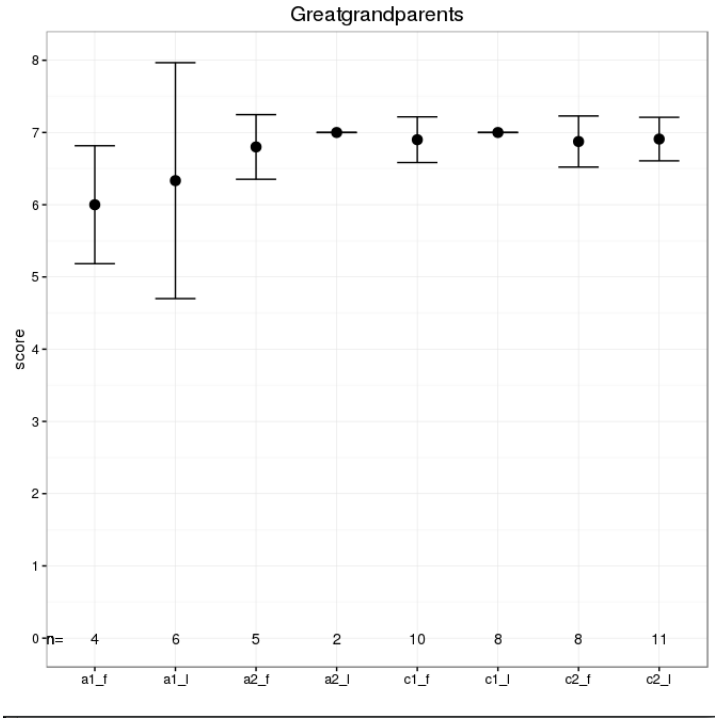
3.6.1 Variables

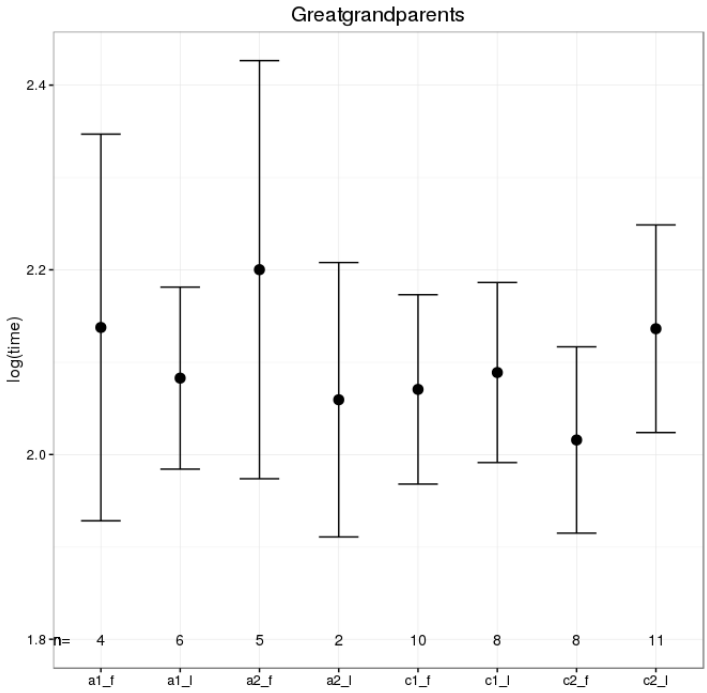
- For greatgrandparent we selected the participants who gave a correct name (n=54).

3.6.2 Summary of Results

- H4: All the participants who had correct naming have high score values.

3.6.3 Graphs





3.6.4 R Output

```
> cknow_cn3$PIorder <- relevel(factor(cknow_cn3$PIorder), "1")

> contrasts(cknow_cn3$PIorder)
  [T.2]
1     0
2     1

> summary(lm(cknow_cn3$P3_score ~ cknow_cn3$PIorder * cknow_cn3$AbsCon * cknow_cn3$FiLa))

Call:
lm(formula = cknow_cn3$P3_score ~ cknow_cn3$PIorder * cknow_cn3$AbsCon *
    cknow_cn3$FiLa)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3333  0.0000  0.0909  0.1250  1.0000

Coefficients:
                                                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                       6.00000    0.31979  18.762   <2e-16 ***
cknow_cn3$PIorder[T.2]                                            0.80000    0.42905   1.865   0.0686 .
cknow_cn3$AbsCon[T.c]                                             0.90000    0.37839   2.379   0.0216 *
cknow_cn3$FiLa[T.l]                                               0.33333    0.41285   0.807   0.4236
cknow_cn3$PIorder[T.2]:cknow_cn3$AbsCon[T.c]
                                                                 -0.82500    0.52548  -1.570   0.1233
cknow_cn3$PIorder[T.2]:cknow_cn3$FiLa[T.l]
                                                                 -0.13333    0.67587  -0.197   0.8445
cknow_cn3$AbsCon[T.c]:cknow_cn3$FiLa[T.l]
                                                                 -0.23333    0.51234  -0.455   0.6509
cknow_cn3$PIorder[T.2]:cknow_cn3$AbsCon[T.c]:cknow_cn3$FiLa[T.l]
                                                                  0.06742    0.79823   0.084   0.9331
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.6396 on 46 degrees of freedom
Multiple R-squared:  0.1935,    Adjusted R-squared:  0.07082
F-statistic: 1.577 on 7 and 46 DF,  p-value: 0.1663


> cknow_cn3$PIorder <- relevel(factor(cknow$PIorder), "1")

> contrasts(cknow$PIorder)
  [T.1]
2     0
1     1

> summary(lm(cknow_cn3$TIME_log_P3_noNa ~ cknow_cn3$PIorder * cknow_cn3$AbsCon * cknow_cn3$FiLa))

Call:
lm(formula = cknow_cn3$TIME_log_P3_noNa ~ cknow_cn3$PIorder *
    cknow_cn3$AbsCon * cknow_cn3$FiLa)

Residuals:
     Min       1Q   Median       3Q      Max
-0.31151 -0.07456  0.00535  0.08328  0.33639

Coefficients:
                                                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                       2.13758    0.06451  33.133   <2e-16 ***
cknow_cn3$PIorder[T.2]                                            0.06259    0.08655   0.723    0.473
cknow_cn3$AbsCon[T.c]                                            -0.06707    0.07633  -0.879    0.384
cknow_cn3$FiLa[T.l]                                              -0.05486    0.08329  -0.659    0.513
cknow_cn3$PIorder[T.2]:cknow_cn3$AbsCon[T.c]                     -0.11737    0.10601  -1.107    0.274
cknow_cn3$PIorder[T.2]:cknow_cn3$FiLa[T.l]                       -0.08601    0.13635  -0.631    0.531
cknow_cn3$AbsCon[T.c]:cknow_cn3$FiLa[T.l]                         0.07313    0.10336   0.708    0.483
cknow_cn3$PIorder[T.2]:cknow_cn3$AbsCon[T.c]:cknow_cn3$FiLa[T.l]  0.18821    0.16103   1.169    0.249
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.129 on 46 degrees of freedom
Multiple R-squared:  0.1556,    Adjusted R-squared:  0.02714
F-statistic: 1.211 on 7 and 46 DF,  p-value: 0.316
```

3.7 Long Time for Naming Indicates Incorrectness (H5)

3.7.1 Variables

- Needs to be calculated per problem because different participants were correct for different problems.
- Needs to be calculated separately for with/without predicate invention because in the with predicate invention condition, participants have to name two predicates, in the without condition only one.
- Variables for naming: GF16_01, GP16_01, PP16_01, AN18_01; because for dontknow there were short and long answers, we separate this group into -1 (short, e.g., no, -) and -2 (long, e.g., I have absolutely no idea)
- Time for naming is generated as new variable from different variables because of naming first/last

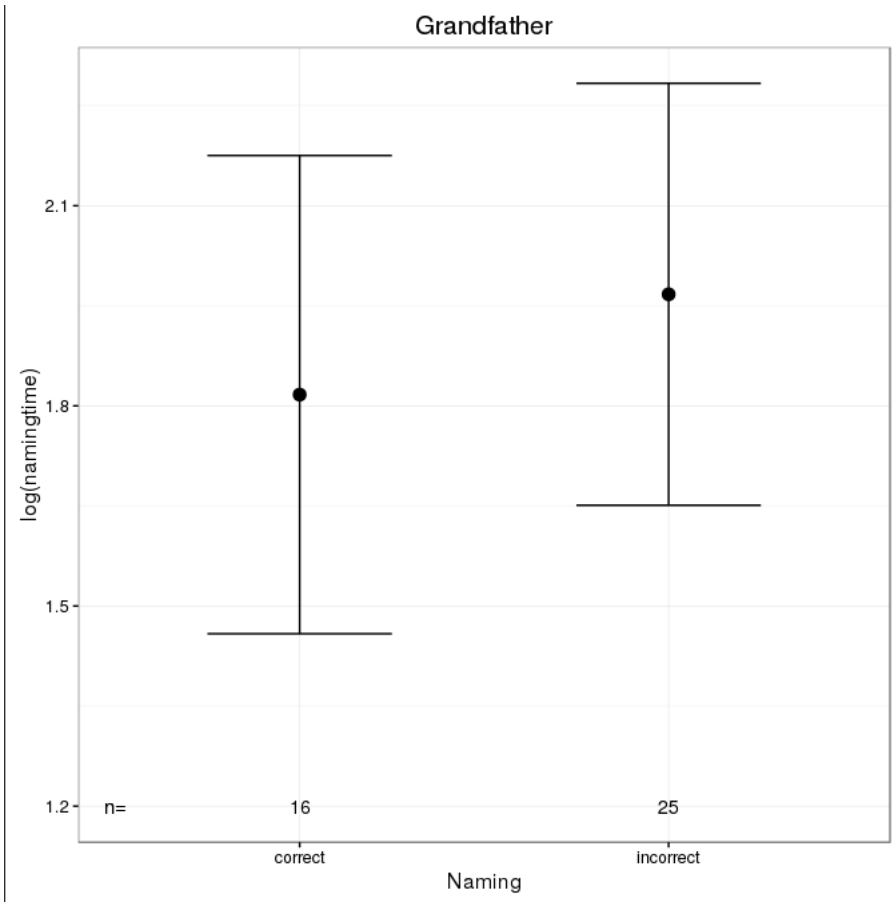
- Analysis: GLM with factor correct naming (4 conditions) and dependent variable log namingtime (do not confuse with answering time for closed questions!)

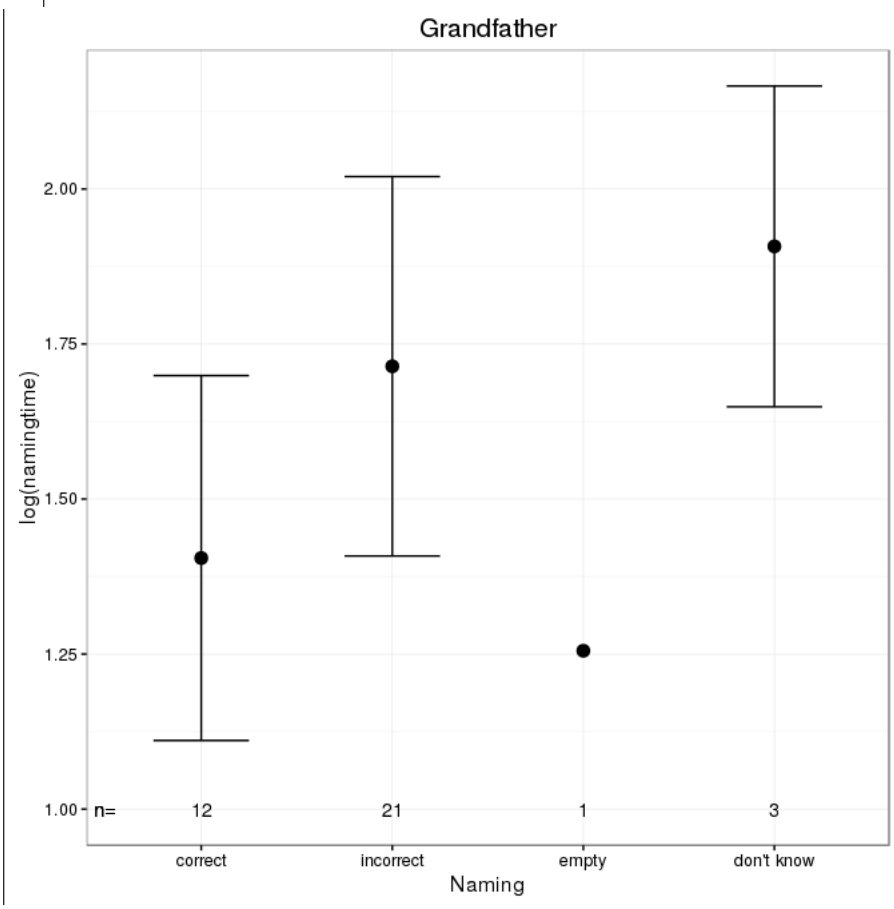
3.7.2 Summary of Results

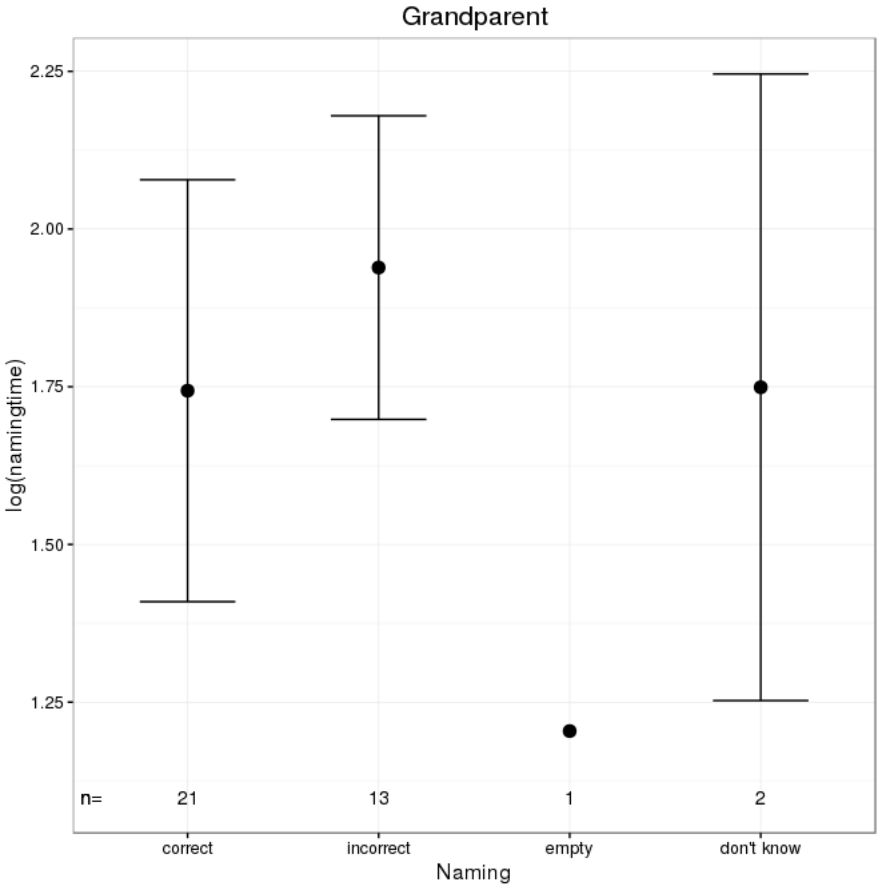
- H5: partially confirmed – correct is always faster than incorrect (with the exception ancestor without predicate invention)
- Significant difference between correct/incorrect for grandfather without predicate invention, marginally significant for grandparent.
- For ancestor with predicate invention there is a significant difference between correct and "dontknow" (slowest)

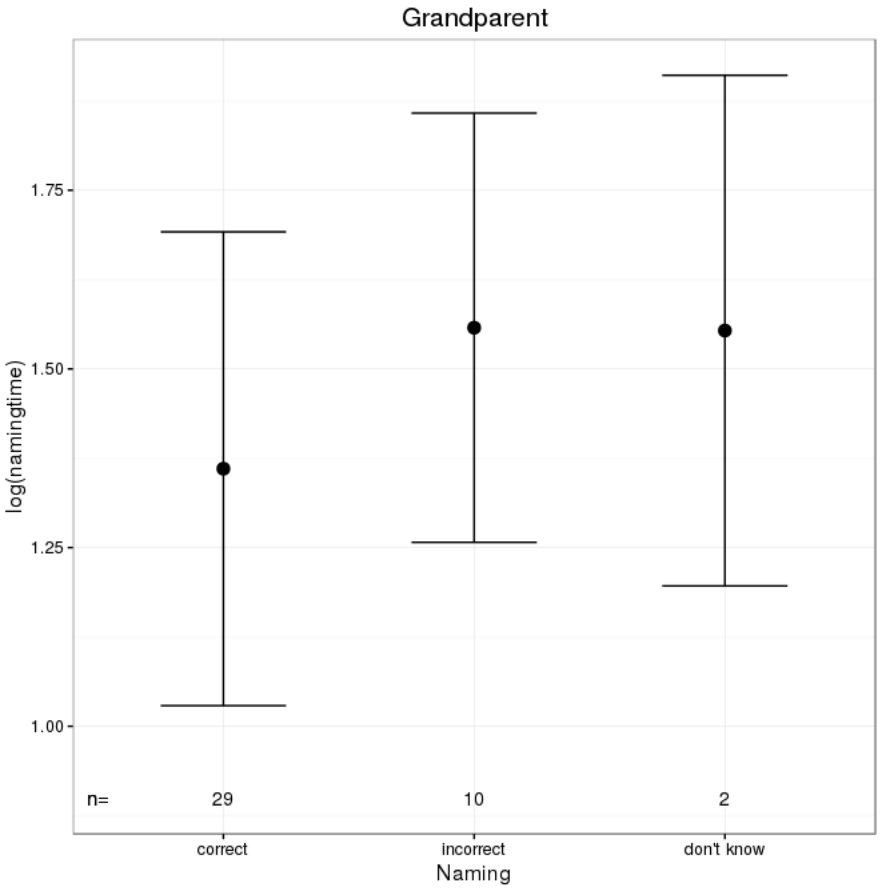
3.7.3 Graphs

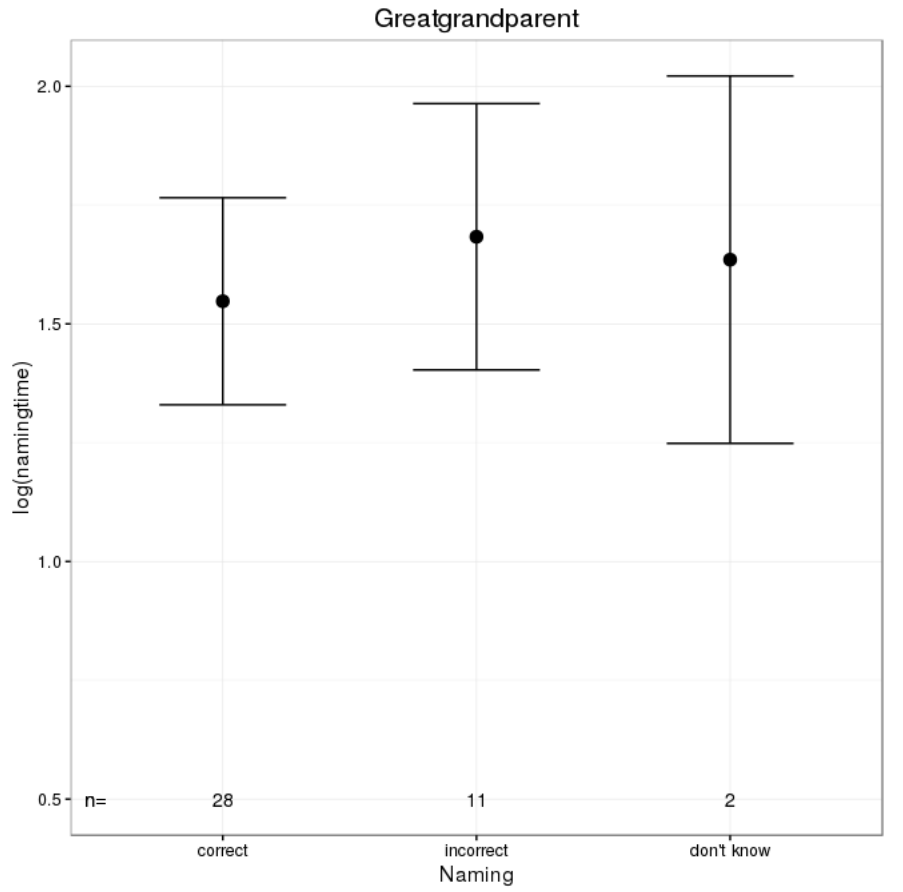
(always PI before noPI)

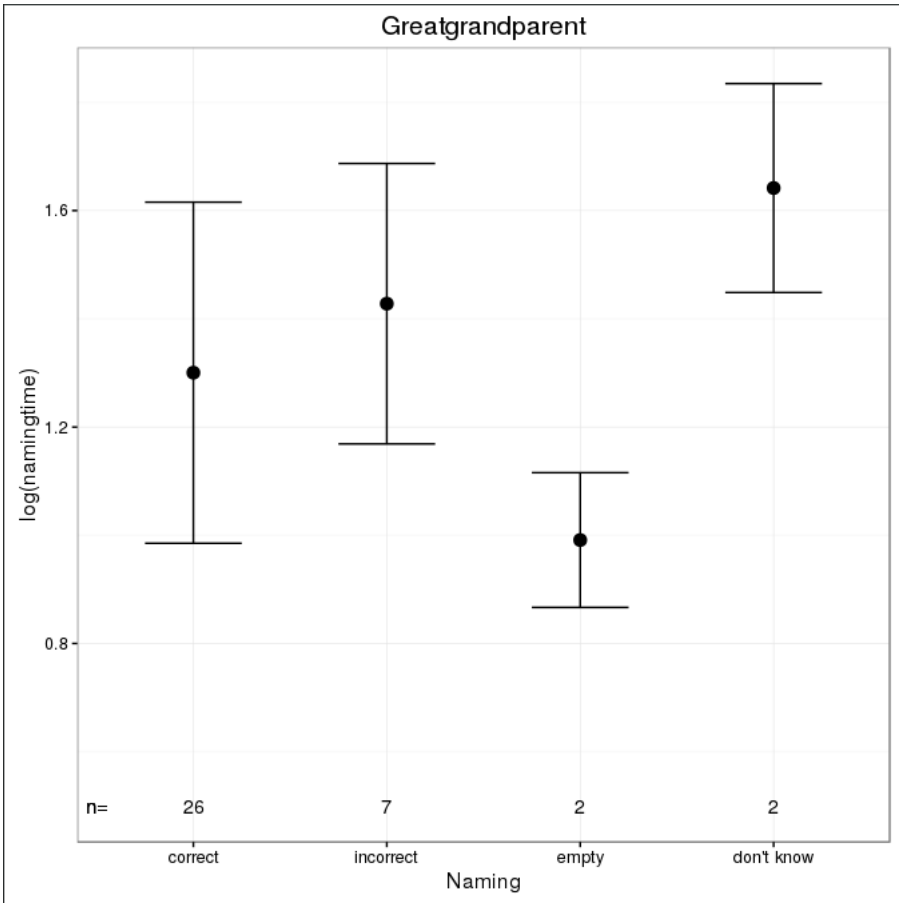

Grandfather
log(namingtime)
2.1
1.8
1.5
1.2
n=
16
25
correct
incorrect
Naming

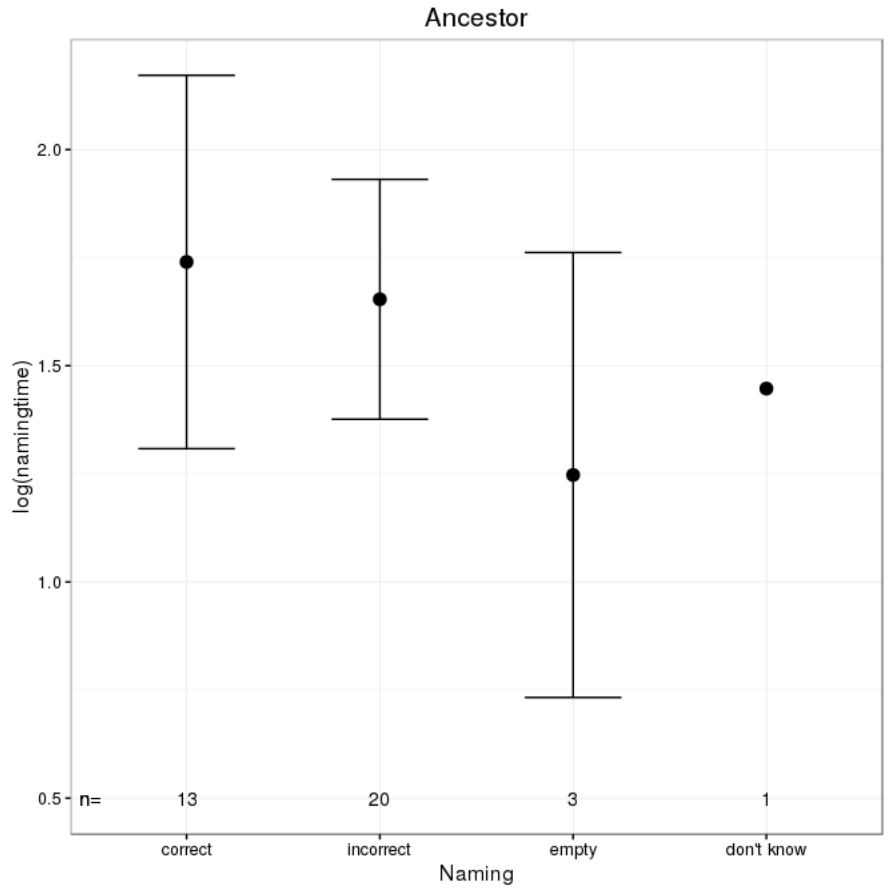

Grandfather
log(namingtime)
2.00
1.75
1.50
1.25
1.00
n=
12
21
1
3
correct
incorrect
empty
don't know
Naming

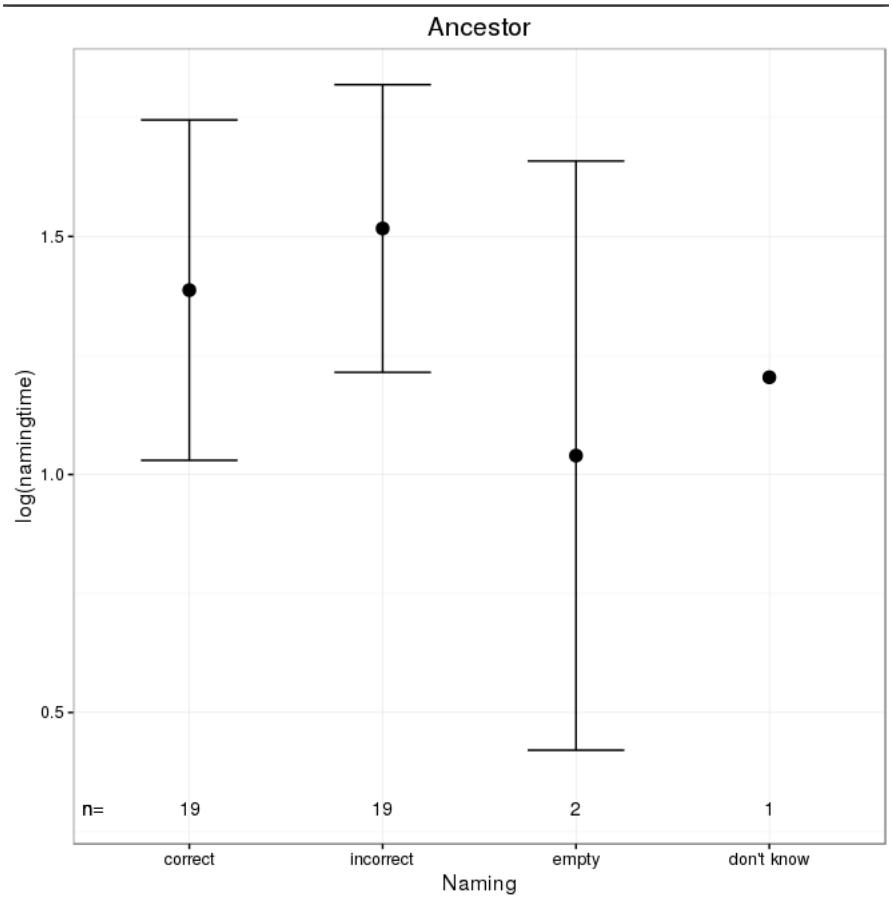

Grandparent
2.25
2.00
1.75
1.50
1.25
log(namingtime)
n=
21
13
1
2
correct
incorrect
empty
don't know
Naming


Grandparent
1.75
1.50
1.25
1.00
log(namingtime)
n=
29
10
2
correct
incorrect
don't know
Naming


Greatgrandparent
log(namingtime)
2.0
1.5
1.0
0.5
n=
28
11
2
correct
incorrect
don't know
Naming


Greatgrandparent
log(namingtime)
1.6
1.2
0.8
n=
26
7
2
2
correct
incorrect
empty
don't know
Naming


Ancestor
log(namingtime)
2.0
1.5
1.0
0.5
n=
13
20
3
1
correct
incorrect
empty
don't know
Naming


Ancestor
log(namingtime)
1.5
1.0
0.5
n=
19
19
2
1
correct
incorrect
empty
don't know
Naming

3.7.4 R Output

```
> piP13 <- cknow[cknow$PIorder == 1,]

> piP24 <- cknow[cknow$PIorder == 2,]

> # Problem 1 predicate invention (in piP13)

> piP13$GF16_01 <- relevel(factor(piP13$GF16_01), "1")

> contrasts(factor(piP13$GF16_01))
  [T.0]
1     0
0     1

> summary(lm(piP13$TIME_log_P1_Na ~ piP13$GF16_01))

Call:
lm(formula = piP13$TIME_log_P1_Na ~ piP13$GF16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.82109 -0.20379  0.02841  0.30462  0.54198

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.81679    0.08328   21.81   <2e-16 ***
piP13$GF16_01[T.0]  0.15043    0.10666    1.41    0.166
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3331 on 39 degrees of freedom
Multiple R-squared:  0.04853,    Adjusted R-squared:  0.02414
F-statistic: 1.989 on 1 and 39 DF,  p-value: 0.1663


> # Problem 1 no predicate invention (in piP24)

> piP24$GF16_01 <- relevel(factor(piP24$GF16_01), "1")

> contrasts(factor(piP24$GF16_01))
   [T.0] [T.-1] [T.-2]
1      0      0      0
0      1      0      0
-1     0      1      0
-2     0      0      1

> summary(lm(piP24$TIME_log_P1_Na ~ piP24$GF16_01))

Call:
lm(formula = piP24$TIME_log_P1_Na ~ piP24$GF16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.63491 -0.25170  0.03409  0.18940  0.57816

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)          1.40484    0.08643  16.254   <2e-16 ***
piP24$GF16_01[T.0]   0.30926    0.10835   2.854   0.0074 **
piP24$GF16_01[T.-1] -0.14957    0.31164  -0.480   0.6344
piP24$GF16_01[T.-2]  0.50267    0.19327   2.601   0.0138 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2994 on 33 degrees of freedom
Multiple R-squared:  0.275,    Adjusted R-squared:  0.2091
F-statistic: 4.173 on 3 and 33 DF,  p-value: 0.01306


> # Problem 2 predicate invention (in piP24)

> piP24$GP16_01 <- relevel(factor(piP24$GP16_01), "1")

> contrasts(factor(piP24$GP16_01))
   [T.0] [T.-1] [T.-2]
1      0      0      0
0      1      0      0
-1     0      1      0
-2     0      0      1

> summary(lm(piP24$TIME_log_P2_Na ~ piP24$GP16_01))

Call:
lm(formula = piP24$TIME_log_P2_Na ~ piP24$GP16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.78948 -0.16079  0.00446  0.25191  0.42063

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.74372    0.06773  25.745   <2e-16 ***
piP24$GP16_01[T.0]   0.19521    0.10953   1.782   0.0839 .
piP24$GP16_01[T.-1] -0.53961    0.31768  -1.699   0.0988 .
piP24$GP16_01[T.-2]  0.00543    0.22969   0.024   0.9813
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3104 on 33 degrees of freedom
Multiple R-squared:  0.1757,    Adjusted R-squared:  0.1007
F-statistic: 2.344 on 3 and 33 DF,  p-value: 0.09088


> # Problem 2 no predicate invention (in piP13)

> piP13$GP16_01 <- relevel(factor(piP13$GP16_01), "1")

> contrasts(factor(piP13$GP16_01))
   [T.0] [T.-2]
1      0      0
0      1      0
-2     0      1

> summary(lm(piP13$TIME_log_P2_Na ~ piP13$GP16_01))
```

```
Call:
lm(formula = piP13$TIME_log_P2_Na ~ piP13$GP16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.51524 -0.23539 -0.01791  0.23479  0.93192

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.36034    0.06035  22.540   <2e-16 ***
piP13$GP16_01[T.0]   0.19727    0.11919   1.655    0.106
piP13$GP16_01[T.-2]  0.19327    0.23761   0.813    0.421
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.325 on 38 degrees of freedom
Multiple R-squared:  0.07556,    Adjusted R-squared:  0.02691
F-statistic: 1.553 on 2 and 38 DF,  p-value: 0.2247

> # Problem 3 predicate invention (in piP13)

> piP13$PP16_01 <- relevel(factor(piP13$PP16_01), "1")

> contrasts(factor(piP13$PP16_01))
   [T.0] [T.-2]
1      0      0
0      1      0
-2     0      1

> summary(lm(piP13$TIME_log_P3_Na ~ piP13$PP16_01))

Call:
lm(formula = piP13$TIME_log_P3_Na ~ piP13$PP16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.47922 -0.16483 -0.00358  0.14255  0.53514

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.54765    0.04564  33.907   <2e-16 ***
piP13$PP16_01[T.0]   0.13570    0.08594   1.579    0.123
piP13$PP16_01[T.-2]  0.08746    0.17678   0.495    0.624
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2415 on 38 degrees of freedom
Multiple R-squared:  0.06339,    Adjusted R-squared:  0.01409
F-statistic: 1.286 on 2 and 38 DF,  p-value: 0.2881


> # Problem 3 no predicate invention (in piP24)

> piP24$PP16_01 <- relevel(factor(piP24$PP16_01), "1")

> contrasts(factor(piP24$PP16_01))
   [T.0] [T.-1] [T.-2]
1      0      0      0
```

```
0      1      0      0
-1     0      1      0
-2     0      0      1

> summary(lm(piP24$TIME_log_P3_Na ~ piP24$PP16_01))

Call:
lm(formula = piP24$TIME_log_P3_Na ~ piP24$PP16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.60141 -0.18644 -0.02163  0.10355  0.59724

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.30038    0.05853  22.217   <2e-16 ***
piP24$PP16_01[T.0]   0.12754    0.12709   1.004    0.323
piP24$PP16_01[T.-1] -0.30925    0.21900  -1.412    0.167
piP24$PP16_01[T.-2]  0.34127    0.21900   1.558    0.129
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2985 on 33 degrees of freedom
Multiple R-squared:  0.1487,    Adjusted R-squared:  0.07128
F-statistic: 1.921 on 3 and 33 DF,  p-value: 0.1454


> # Problem 4 predicate invention (in piP24)

> piP24$AN18_01 <- relevel(factor(piP24$AN18_01), "1")

> contrasts(factor(piP24$AN18_01))
   [T.0] [T.-1] [T.-2]
1      0      0      0
0      1      0      0
-1     0      1      0
-2     0      0      1

> summary(lm(piP24$TIME_log_P4_Na ~ piP24$AN18_01))

Call:
lm(formula = piP24$TIME_log_P4_Na ~ piP24$AN18_01)

Residuals:
    Min      1Q  Median      3Q     Max
-0.9619 -0.1624 -0.0627  0.1913  0.6102

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.74007    0.09928  17.527   <2e-16 ***
piP24$AN18_01[T.0]  -0.08631    0.12752  -0.677    0.503
piP24$AN18_01[T.-1] -0.49276    0.22927  -2.149    0.039 *
piP24$AN18_01[T.-2] -0.29292    0.37146  -0.789    0.436
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.358 on 33 degrees of freedom
Multiple R-squared:  0.1301,    Adjusted R-squared:  0.05107
```

```
F-statistic: 1.646 on 3 and 33 DF,  p-value: 0.1977

> # Problem 4 no predicate invention (in piP13)

> piP13$AN18_01 <- relevel(factor(piP13$AN18_01), "1")

> contrasts(factor(piP13$AN18_01))
   [T.0] [T.-1] [T.-2]
1      0      0      0
0      1      0      0
-1     0      1      0
-2     0      0      1

> summary(lm(piP13$TIME_log_P4_Na ~ piP13$AN18_01))

Call:
lm(formula = piP13$TIME_log_P4_Na ~ piP13$AN18_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.61364 -0.24112  0.01475  0.30281  0.50484

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.38725    0.07846  17.682   <2e-16 ***
piP13$AN18_01[T.0]   0.12948    0.11096   1.167    0.251
piP13$AN18_01[T.-1] -0.34766    0.25423  -1.367    0.180
piP13$AN18_01[T.-2] -0.18313    0.35087  -0.522    0.605
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.342 on 37 degrees of freedom
Multiple R-squared:  0.1096,    Adjusted R-squared:  0.03743
F-statistic: 1.519 on 3 and 37 DF,  p-value: 0.2258
```

4 Bamberg Data

4.1 Files and Variables

- Data file: `ba_rdata_collapsed2_naming.csv`
- Raw data preprocessing: `ba_preprocessing.r`
- Variable names and coding: `ba_variables_listing.pdf`

4.2 Naming Answers

```
Grandfather
Correct answers:
Incorrect answers:

Grandparent
Correct answers:
Incorrect answers:

Greatgrandparent
Correct answers:
Incorrect answers:

Ancestors
Correct answers:
Incorrect answers:
```

4.3 Participants

- Data collection was on April 20 2016 in an applied computer science class (after a Prolog introduction in the previous week).
- Since this sample is absolute beginners, we only have the public naming condition. That is, we have only two factors (Predicate invention and naming position).
- Originally, 25 students participated, however, three students needed too long (because class did end), time limit was 22 minutes.
- No participants had to be excluded due to many don't know answers.
- $n = 23$, male=13, female=9, mean age = 24.73, sd=2.31

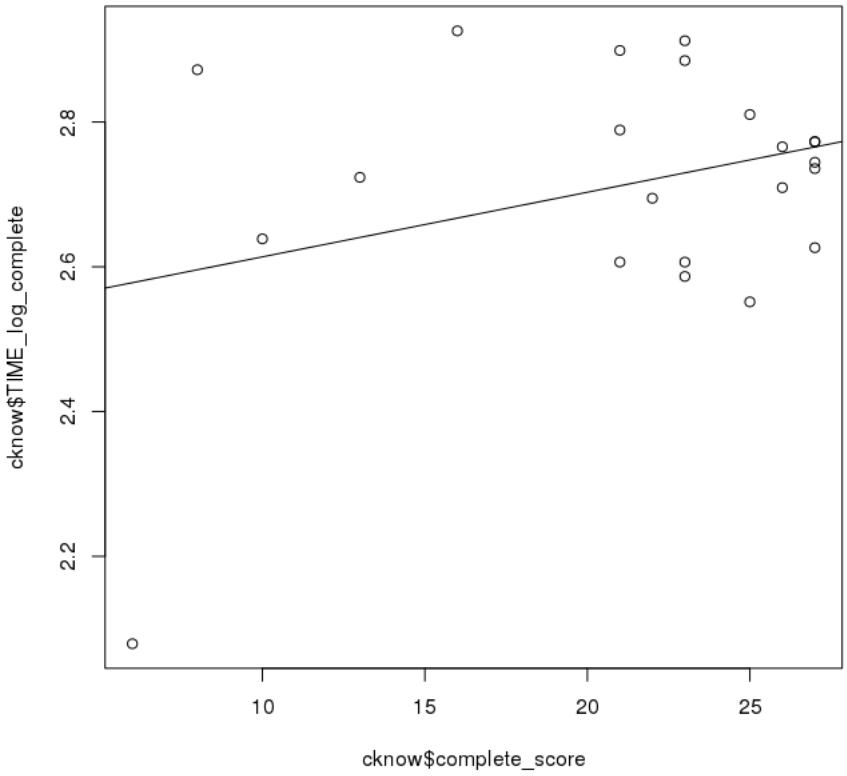
4.4 Relation Time Score (H1)

4.4.1 Summary of Results

- H1 is not confirmed: Participants with higher scores have higher solution times, positive correlation of score and log. time.
- **Absolute beginners need to invest more time to reach good comprehension!**

	Grandfather	Grandparent	Greatgrandparent	Ancestor	Complete
r	0.45	-0.09	0.40	0.28	0.32
p	0.04	n.s.	0.06	n.s. 0.2	n.s. 0.145

4.4.2 Graphs



4.4.3 R Output

```
> cor.test(cknow$TIME_log_complete, cknow$complete_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_complete and cknow$complete_score
t = 1.5166, df = 20, p-value = 0.145
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1161792  0.6541890
sample estimates:
      cor
0.3211604


> cor.test(cknow$TIME_log_P1_noNa, cknow$P1_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_P1_noNa and cknow$P1_score
t = 2.2386, df = 20, p-value = 0.03672
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.0320533 0.7312262
sample estimates:
      cor
```

```
0.4476127


> cor.test(cknow$TIME_log_P2_noNa, cknow$P2_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_P2_noNa and cknow$P2_score
t = -0.3908, df = 20, p-value = 0.7001
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4906484  0.3473079
sample estimates:
        cor
-0.08704655


> cor.test(cknow$TIME_log_P3_noNa, cknow$P3_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_P3_noNa and cknow$P3_score
t = 1.9737, df = 20, p-value = 0.0624
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.02152007  0.70530027
sample estimates:
      cor
0.4037516


> cor.test(cknow$TIME_log_P4_noNa, cknow$P4_score)

    Pearson's product-moment correlation

data:  cknow$TIME_log_P4_noNa and cknow$P4_score
t = 1.3293, df = 20, p-value = 0.1987
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1553426  0.6307632
sample estimates:
      cor
0.2849268
```

4.5 Correct Naming Implies High Performance (H2)

4.5.1 Variables

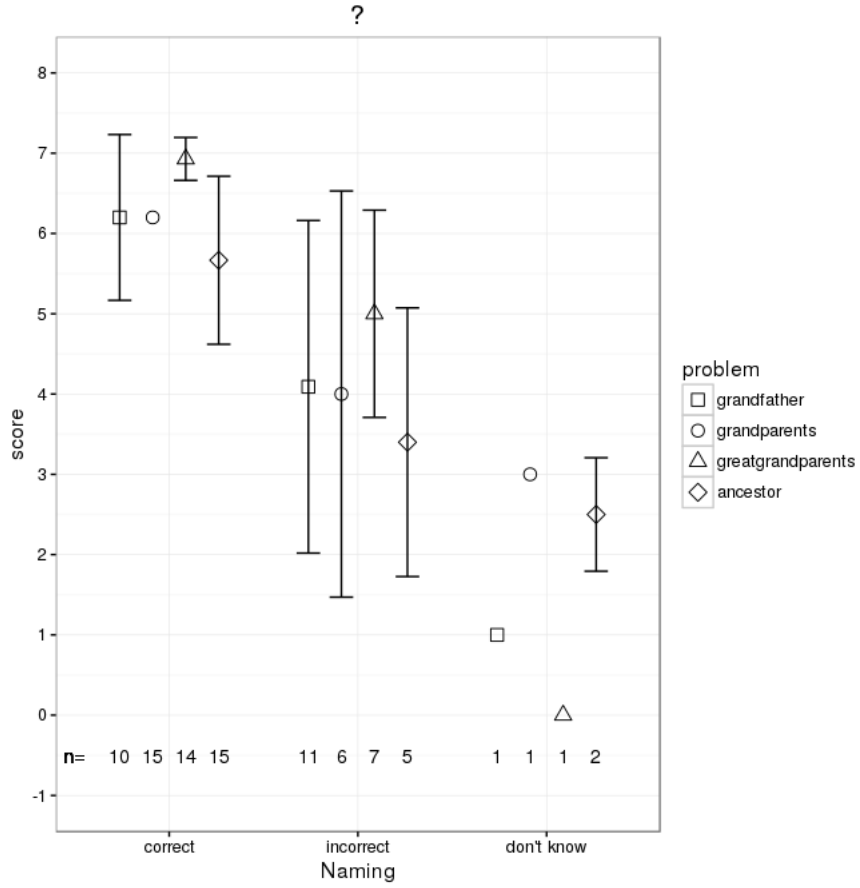
- Levene shows that variances are homogenous for log time for all problems.
  Levene shows that variances are not homogenous for score.
  Therefore, and to be consistent with the Osnabrueck data analyses, we do analyses with GLM (and not Anova) with dummy coding (contrast) for naming as independent/predictor variable and score rsp. log time as dependent/criterion variable.
- P1 _naming_c3 (1 = correct, 0 = incorrect, -1 = don't know answers) and for the other problems; this variable is dummy coded (contrast)
- Not log time is given descriptively, statistics is done with log times.

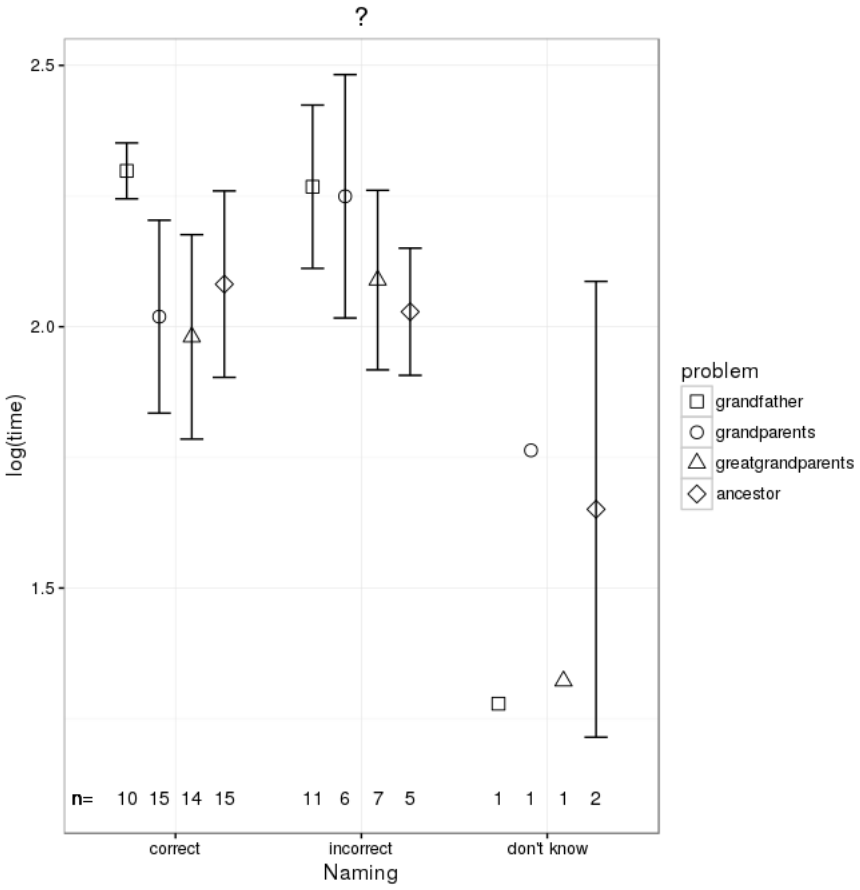
4.5.2 Summary of Results

- H2 is confirmed for score but not for time: Participants who name the problem correctly have significantly higher scores.
- Significances are only reported for contrast correct/incorrect

	Correct	Incorrect	DontKnow	Test
Grandfather	n=10	11	1	
Score	mean 6.2 (sd=1.03)	4.09 (2.07)	1	-2.11, $p = 0.01$
Time	200.1 (26.16)	196.36 (70.38)	19	
LogTime	2.3 (0.05)	2.27 0.16	1.28	0.14, n.s.
Grandparent	15	6	1	
Score	6.2 (1.9)	4.0 (2.53)	3	-2.20, $p = 0.04$
Time	113.13 (44.21)	198.17 (96.43)	58	
LogTime	2.02 (0.18)	2.25 (0.23)	1.76	0.23, $p = 0.03$
Greatgrandparent	14	7	1	
Score	6.93 (0.27)	5.0 (1.29)	0	-1.93, $p < 0.001$
Time	104.5 (43.57)	132.14 (59.73)	21	
LogTime	1.98 (0.2)	2.09 (0.17)	1.32	n.s.
Ancestor	15	5	2	
Score	5.67 (1.05)	3.4 (1.67)	2.5 (0.71)	-2.27, $p = 0.002$
Time	130.6 (56.2)	110.2 (30.65)	56.5 (48.79)	
LogTime	2.08 (0.18)	2.03 (0.12)	1.65 (0.44)	n.s.

4.5.3 Graphs


?
score
8
7
6
5
4
3
2
1
0
-1
n= 10 15 14 15 11 6 7 5 1 1 1 2
correct
incorrect
don't know
Naming
problem
grandfather
grandparents
greatgrandparents
ancestor


?
log(time)
2.5
2.0
1.5
n= 10 15 14 15 11 6 7 5 1 1 1 2
correct
incorrect
don't know
Naming
problem
grandfather
grandparents
greatgrandparents
ancestor

4.5.4 R Output

```
> leveneTest(cknow$P1_score ~ factor(cknow$P1_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2  4.0455 0.03438 *
      19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> describeBy(cknow$P1_score,factor(cknow$P1_naming_c3))
group: 1
  vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
1    1 10  6.2 1.03    6.5    6.38 0.74   4   7     3 -0.89    -0.58 0.33
------------------------------------------------------------------------------------------
group: -1
  vars n mean sd median trimmed mad min max range skew kurtosis se
1    1 1    1 NA      1       1   0   1   1     0   NA       NA NA
------------------------------------------------------------------------------------------
group: 0
  vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
1    1 11 4.09 2.07      4    4.11 2.97   1   7     6 0.07    -1.57 0.62

> cknow$P1_naming_c3 <- relevel(factor(cknow$P1_naming_c3), "1")

> contrasts(factor(cknow$P1_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$P1_score ~ factor(cknow$P1_naming_c3)))

Call:
lm(formula = cknow$P1_score ~ factor(cknow$P1_naming_c3))

Residuals:
     Min       1Q   Median       3Q      Max
-3.09091 -1.09091 -0.04545  0.80000  2.90909

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                        6.2000     0.5257  11.794 3.47e-10 ***
factor(cknow$P1_naming_c3)[T.-1]  -5.2000     1.7436  -2.982 0.00765 **
factor(cknow$P1_naming_c3)[T.0]   -2.1091     0.7264  -2.904 0.00911 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.662 on 19 degrees of freedom
Multiple R-squared:  0.4281,    Adjusted R-squared:  0.3679
F-statistic: 7.112 on 2 and 19 DF,  p-value: 0.004948


> leveneTest(cknow$P2_score ~ factor(cknow$P2_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.1765 0.3298
      19
```

```
> describeBy(cknow$P2_score,factor(cknow$P2_naming_c3))
group: 1
  vars  n mean  sd median trimmed mad min max range  skew kurtosis se
1    1 15  6.2 1.9      7    6.62   0   0   7     7 -2.38     4.74 0.49
------------------------------------------------------------------------------------
group: -1
  vars n mean sd median trimmed mad min max range skew kurtosis se
1    1 1    3 NA      3       3   0   3   3     0   NA       NA NA
------------------------------------------------------------------------------------
group: 0
  vars n mean   sd median trimmed  mad min max range  skew kurtosis   se
1    1 6    4 2.53    4.5       4 2.97   1   7     6 -0.19    -1.94 1.03

> cknow$P2_naming_c3 <- relevel(factor(cknow$P2_naming_c3), "1")

> contrasts(factor(cknow$P2_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$P2_score ~ factor(cknow$P2_naming_c3)))

Call:
lm(formula = cknow$P2_score ~ factor(cknow$P2_naming_c3))

Residuals:
   Min     1Q Median     3Q    Max
 -6.20  -0.15   0.80   0.80   3.00

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       6.2000     0.5377  11.531 5.07e-10 ***
factor(cknow$P2_naming_c3)[T.-1]  -3.2000     2.1508  -1.488 0.1532
factor(cknow$P2_naming_c3)[T.0]   -2.2000     1.0059  -2.187 0.0415 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.083 on 19 degrees of freedom
Multiple R-squared:  0.2472,    Adjusted R-squared:  0.1679
F-statistic: 3.119 on 2 and 19 DF,  p-value: 0.06739


> leveneTest(cknow$P3_score ~ factor(cknow$P3_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2  3.6182 0.04662 *
      19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> describeBy(cknow$P3_score,factor(cknow$P3_naming_c3))
group: 1
  vars  n mean   sd median trimmed mad min max range  skew kurtosis   se
1    1 14 6.93 0.27      7       7   0   6   7     1 -2.98     7.41 0.07
------------------------------------------------------------------------------------
group: -1
```

```
  vars n mean sd median trimmed mad min max range skew kurtosis se
1    1 1    0 NA      0       0   0   0   0     0   NA       NA NA
------------------------------------------------------------------------------------------
group: 0
  vars n mean   sd median trimmed mad min max range skew kurtosis se
1    1 7    5 1.29      6       5   0   3   6     3 -0.4    -1.87 0.49

> cknow$P3_naming_c3 <- relevel(factor(cknow$P3_naming_c3), "1")

> contrasts(factor(cknow$P3_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$P3_score ~ factor(cknow$P3_naming_c3)))

Call:
lm(formula = cknow$P3_score ~ factor(cknow$P3_naming_c3))

Residuals:
     Min       1Q   Median       3Q      Max
-2.00000  0.07143  0.07143  0.07143  1.00000

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       6.9286     0.2027  34.182  < 2e-16 ***
factor(cknow$P3_naming_c3)[T.-1] -6.9286     0.7850  -8.826 3.78e-08 ***
factor(cknow$P3_naming_c3)[T.0]  -1.9286     0.3511  -5.493 2.67e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7584 on 19 degrees of freedom
Multiple R-squared:  0.8344,    Adjusted R-squared:  0.817
F-statistic: 47.87 on 2 and 19 DF,  p-value: 3.808e-08


> leveneTest(cknow$P4_score ~ factor(cknow$P4_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.3312 0.2877
      19

> describeBy(cknow$P4_score,factor(cknow$P4_naming_c3))
group: 1
  vars  n mean   sd median trimmed mad min max range  skew kurtosis   se
1    1 15 5.67 1.05      6    5.92   0   2   6     4 -2.86     7.07 0.27
------------------------------------------------------------------------------------------
group: -1
  vars n mean   sd median trimmed  mad min max range skew kurtosis se
1    1 2  2.5 0.71    2.5     2.5 0.74   2   3     1    0    -2.75 0.5
------------------------------------------------------------------------------------------
group: 0
  vars n mean   sd median trimmed  mad min max range skew kurtosis se
1    1 5  3.4 1.67      3     3.4 1.48   2   6     4 0.52    -1.63 0.75

> cknow$P4_naming_c3 <- relevel(factor(cknow$P4_naming_c3), "1")
```

```
> contrasts(factor(cknow$P4_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$P4_score ~ factor(cknow$P4_naming_c3)))

Call:
lm(formula = cknow$P4_score ~ factor(cknow$P4_naming_c3))

Residuals:
    Min      1Q  Median      3Q     Max
-3.6667 -0.2167  0.3333  0.3333  2.6000

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                        5.6667     0.3080  18.399 1.44e-13 ***
factor(cknow$P4_naming_c3)[T.-1]  -3.1667     0.8979  -3.527 0.00225 **
factor(cknow$P4_naming_c3)[T.0]   -2.2667     0.6160  -3.680 0.00159 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.193 on 19 degrees of freedom
Multiple R-squared:  0.5386,    Adjusted R-squared:   0.49
F-statistic: 11.09 on 2 and 19 DF,  p-value: 0.0006437


> leveneTest(cknow$TIME_log_P1_noNa ~ factor(cknow$P1_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2  4.3059 0.02869 *
      19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> describeBy(cknow$TIME_log_P1_noNa,factor(cknow$P1_naming_c3))
group: -1
  vars n mean sd median trimmed mad  min  max range skew kurtosis se
1    1 1 1.28 NA   1.28    1.28   0 1.28 1.28     0   NA       NA NA
------------------------------------------------------------------------------------
group: 0
  vars  n mean   sd median trimmed  mad min  max range  skew kurtosis   se
1    1 11 2.27 0.16   2.27    2.27 0.15   2 2.51  0.51 -0.03 -1.12 0.05
------------------------------------------------------------------------------------
group: 1
  vars  n mean   sd median trimmed  mad  min  max range skew kurtosis   se
1    1 10  2.3 0.05   2.29    2.29 0.04 2.23 2.42  0.19 0.86 -0.18 0.02

> cknow$P1_naming_c3 <- relevel(factor(cknow$P1_naming_c3), "1")

> contrasts(factor(cknow$P1_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$TIME_log_P1_noNa ~ factor(cknow$P1_naming_c3)))
```

```
Call:
lm(formula = cknow$TIME_log_P1_noNa ~ factor(cknow$P1_naming_c3))

Residuals:
      Min        1Q    Median        3Q       Max
-0.272170 -0.049070 -0.005501  0.047124  0.237345

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      2.29817    0.03768  60.986  < 2e-16 ***
factor(cknow$P1_naming_c3)[T.-1] -1.01942    0.12498  -8.157 1.26e-07 ***
factor(cknow$P1_naming_c3)[T.0]  -0.03037    0.05207  -0.583 0.567
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1192 on 19 degrees of freedom
Multiple R-squared:  0.7817,    Adjusted R-squared:  0.7587
F-statistic: 34.02 on 2 and 19 DF,  p-value: 5.262e-07


> leveneTest(cknow$TIME_log_P2_noNa ~ factor(cknow$P2_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.7097 0.5044
      19

> describeBy(cknow$TIME_log_P2_noNa,factor(cknow$P2_naming_c3))
group: -1
  vars n mean sd median trimmed mad  min  max range skew kurtosis se
1    1 1 1.76 NA   1.76    1.76   0 1.76 1.76     0   NA       NA NA
------------------------------------------------------------------------------------------
group: 0
  vars n mean   sd median trimmed  mad  min  max range skew kurtosis  se
1    1 6 2.25 0.23   2.28    2.25 0.15 1.86 2.56   0.7 -0.4    -1.14 0.1
------------------------------------------------------------------------------------------
group: 1
  vars  n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
1    1 15 2.02 0.18   2.05    2.03 0.21 1.68 2.27  0.59 -0.32 -1.29 0.05

> cknow$P2_naming_c3 <- relevel(factor(cknow$P2_naming_c3), "1")

> contrasts(factor(cknow$P2_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$TIME_log_P2_noNa ~ factor(cknow$P2_naming_c3)))

Call:
lm(formula = cknow$TIME_log_P2_noNa ~ factor(cknow$P2_naming_c3))

Residuals:
     Min       1Q   Median       3Q      Max
-0.39222 -0.08383  0.02219  0.11006  0.30795

Coefficients:
```

```
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      2.01921    0.05123  39.418 <2e-16 ***
factor(cknow$P2_naming_c3)[T.-1] -0.25579    0.20490  -1.248 0.2271
factor(cknow$P2_naming_c3)[T.0]   0.23034    0.09584   2.404 0.0266 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1984 on 19 degrees of freedom
Multiple R-squared:  0.3036,    Adjusted R-squared:  0.2303
F-statistic: 4.143 on 2 and 19 DF,  p-value: 0.03213


> leveneTest(cknow$TIME_log_P3_noNa ~ factor(cknow$P3_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.7551 0.4836
      19

> describeBy(cknow$TIME_log_P3_noNa,factor(cknow$P3_naming_c3))
group: -1
  vars n mean sd median trimmed mad  min  max range skew kurtosis se
1    1 1 1.32 NA   1.32    1.32   0 1.32 1.32     0   NA       NA NA
------------------------------------------------------------------------------------------
group: 0
  vars n mean   sd median trimmed  mad  min max range skew kurtosis   se
1    1 7 2.09 0.17      2    2.09 0.12 1.92 2.4  0.48 0.66    -1.23 0.06
------------------------------------------------------------------------------------------
group: 1
  vars  n mean  sd median trimmed  mad  min  max range  skew kurtosis   se
1    1 14 1.98 0.2   2.03    1.99 0.23 1.66 2.24  0.58 -0.26 -1.35 0.05

> cknow$P3_naming_c3 <- relevel(factor(cknow$P3_naming_c3), "1")

> contrasts(factor(cknow$P3_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$TIME_log_P3_noNa ~ factor(cknow$P3_naming_c3)))

Call:
lm(formula = cknow$TIME_log_P3_noNa ~ factor(cknow$P3_naming_c3))

Residuals:
     Min       1Q   Median       3Q      Max
-0.31780 -0.11583  0.01824  0.11326  0.31039

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      1.98056    0.05034  39.342  < 2e-16 ***
factor(cknow$P3_naming_c3)[T.-1] -0.65834    0.19498  -3.377 0.00317 **
factor(cknow$P3_naming_c3)[T.0]   0.10872    0.08720   1.247 0.22759
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1884 on 19 degrees of freedom
Multiple R-squared:  0.4334,    Adjusted R-squared:  0.3738
```

```
F-statistic: 7.267 on 2 and 19 DF,  p-value: 0.004529


> leveneTest(cknow$TIME_log_P4_noNa ~ factor(cknow$P4_naming_c3))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2  3.2294 0.06204 .
      19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> describeBy(cknow$TIME_log_P4_noNa,factor(cknow$P4_naming_c3))
group: -1
  vars n mean   sd median trimmed  mad  min  max range skew kurtosis   se
1    1 2 1.65 0.44   1.65    1.65 0.46 1.34 1.96  0.62    0    -2.75 0.31
------------------------------------------------------------------------------------------
group: 0
  vars n mean   sd median trimmed  mad  min  max range skew kurtosis   se
1    1 5 2.03 0.12   2.06    2.03 0.18 1.89 2.18  0.29 0.02       -2 0.05
------------------------------------------------------------------------------------------
group: 1
  vars  n mean   sd median trimmed  mad  min  max range skew kurtosis   se
1    1 15 2.08 0.18   2.09    2.08 0.17 1.77 2.44  0.67  0.1 -0.67 0.05

> cknow$P4_naming_c3 <- relevel(factor(cknow$P4_naming_c3), "1")

> contrasts(factor(cknow$P4_naming_c3))
   [T.-1] [T.0]
1       0     0
-1      1     0
0       0     1

> summary(lm(cknow$TIME_log_P4_noNa ~ factor(cknow$P4_naming_c3)))

Call:
lm(formula = cknow$TIME_log_P4_noNa ~ factor(cknow$P4_naming_c3))

Residuals:
     Min       1Q   Median       3Q      Max
-0.31055 -0.10923  0.02061  0.10412  0.36108

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                       2.08140    0.04937  42.161  < 2e-16 ***
factor(cknow$P4_naming_c3)[T.-1] -0.43067    0.14393  -2.992 0.00749 **
factor(cknow$P4_naming_c3)[T.0]  -0.05278    0.09874  -0.535 0.59916
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1912 on 19 degrees of freedom
Multiple R-squared:  0.3203,    Adjusted R-squared:  0.2488
F-statistic: 4.477 on 2 and 19 DF,  p-value: 0.02553


> describeBy(cknow$TIME_P1_noNa,factor(cknow$P1_naming_c3))
group: 1
  vars  n  mean    sd median trimmed   mad min max range skew kurtosis   se
1    1 10 200.1 26.16  193.5  196.25 18.53 170 261    91 1.08 0.23 8.27
```

```
------------------------------------------------------------------------------------
group: -1
  vars n mean sd median trimmed mad min max range skew kurtosis se
1    1 1   19 NA     19      19   0  19  19     0   NA       NA NA
------------------------------------------------------------------------------------
group: 0
  vars  n   mean    sd median trimmed  mad min max range skew kurtosis    se
1    1 11 196.36 70.38    185  193.44 59.3  99 320   221 0.48 -1.06 21.22

> describeBy(cknow$TIME_P2_noNa,factor(cknow$P2_naming_c3))
group: 1
  vars  n   mean    sd median trimmed   mad min max range skew kurtosis    se
1    1 15 113.13 44.21    113  112.54 65.23  48 186   138 0.13 -1.4 11.42
------------------------------------------------------------------------------------
group: -1
  vars n mean sd median trimmed mad min max range skew kurtosis se
1    1 1   58 NA     58      58   0  58  58     0   NA       NA NA
------------------------------------------------------------------------------------
group: 0
  vars n   mean    sd median trimmed   mad min max range skew kurtosis    se
1    1 6 198.17 96.43  189.5  198.17 63.01  72 361   289 0.41 -1.14 39.37

> describeBy(cknow$TIME_P3_noNa,factor(cknow$P3_naming_c3))
group: 1
  vars  n  mean    sd median trimmed   mad min max range skew kurtosis    se
1    1 14 104.5 43.57    106  103.67 56.34  46 173   127  0.2 -1.45 11.65
------------------------------------------------------------------------------------
group: -1
  vars n mean sd median trimmed mad min max range skew kurtosis se
1    1 1   21 NA     21      21   0  21  21     0   NA       NA NA
------------------------------------------------------------------------------------
group: 0
  vars n   mean    sd median trimmed  mad min max range skew kurtosis    se
1    1 7 132.14 59.73    100  132.14 25.2  83 251   168 0.97 -0.63 22.58

> describeBy(cknow$TIME_P4_noNa,factor(cknow$P4_naming_c3))
group: 1
  vars  n  mean   sd median trimmed   mad min max range skew kurtosis    se
1    1 15 130.6 56.2    124  124.85 44.48  59 277   218 1.03 0.65 14.51
------------------------------------------------------------------------------------
group: -1
  vars n mean    sd median trimmed   mad min max range skew kurtosis   se
1    1 2 56.5 48.79   56.5    56.5 51.15  22  91    69    0    -2.75 34.5
------------------------------------------------------------------------------------
group: 0
  vars n  mean    sd median trimmed   mad min max range skew kurtosis    se
1    1 5 110.2 30.65    116   110.2 48.93  78 153    75 0.18 -1.87 13.71
```

4.6 Effect of Predicate Invention on Comprehensibility (H3)

4.6.1 Variables

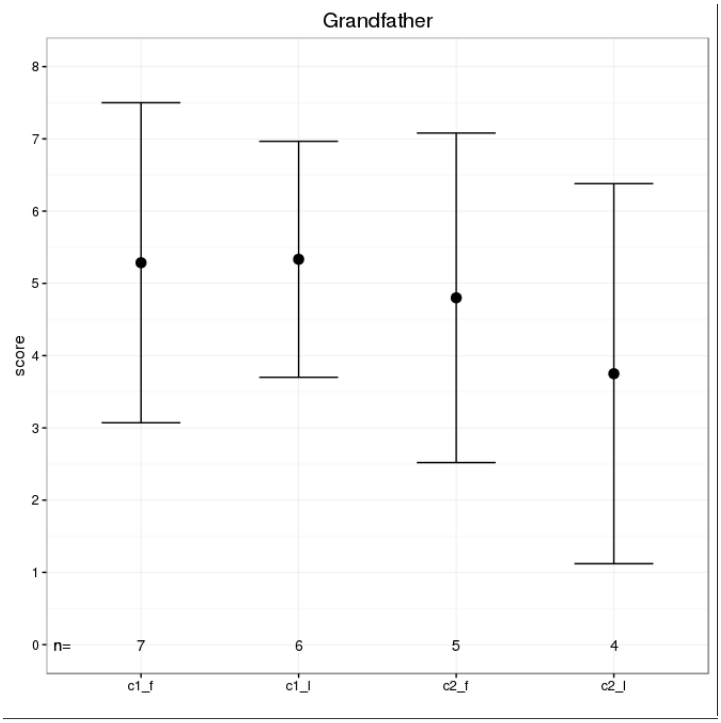
- Factors: Predicate invention with/without; Naming First/Last (for Bamberg there is no public/private variation, only public)
- Dependent variables: score and log time
- Attention: with PI is 1 for problem 1 and 3 and 2 for the other two problems

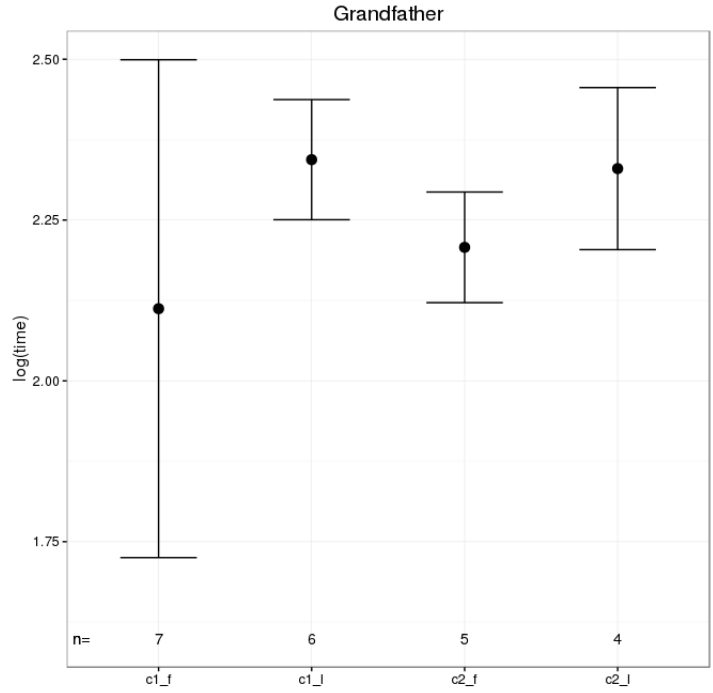
4.6.2 Summary of Results

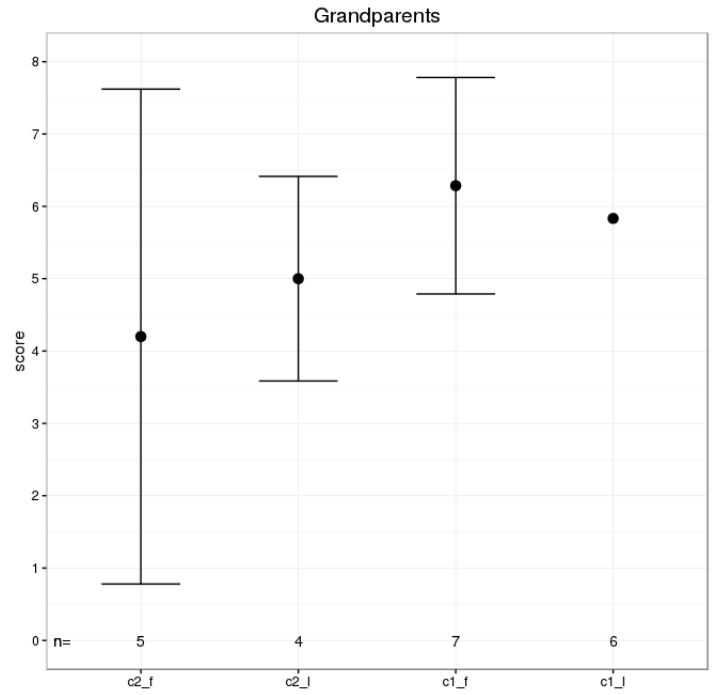
- H3 only concerns the greatgrandparent problem because there the code without predicate invention is significantly longer than the code with predicate invention. The hypothesis is confirmed for time but not for score: with predicate invention is sigfnificantly faster, in addition predicate invention is fastest if naming is first (significant interaction for time); scores are higher for with predicate invention but not significantly so.
- In addtion, we tested the effects of both factors on score and time. There are no significant differences for score, but for time:

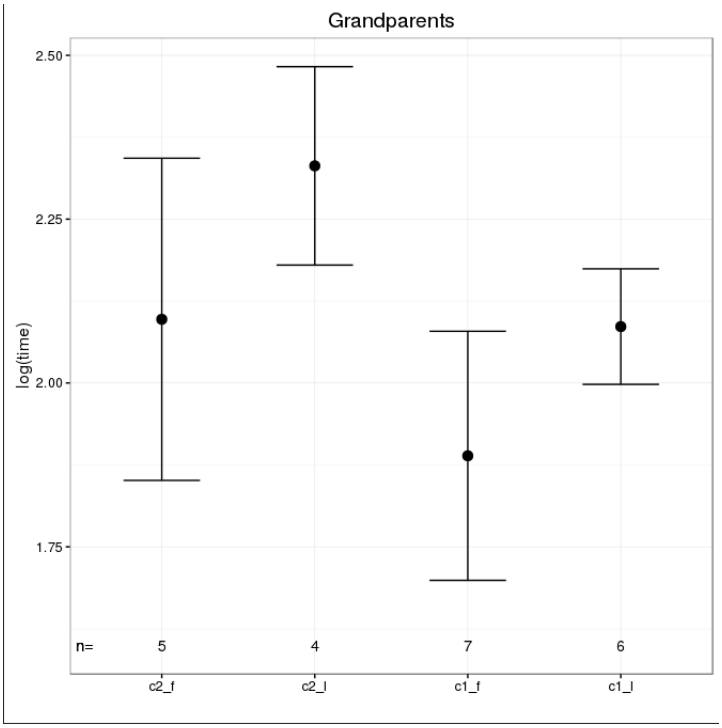
  - grandfather: naming first is significantly faster
  - grandparent: with predicate invention is slower, naming first is faster (no interaction)
  - ancestor: predicate invention is significantly slower

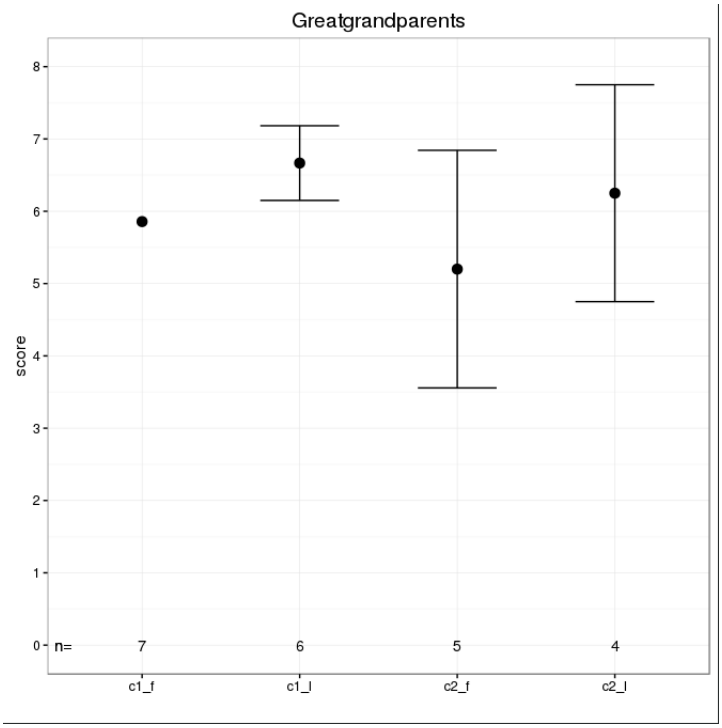
4.6.3 Graphs

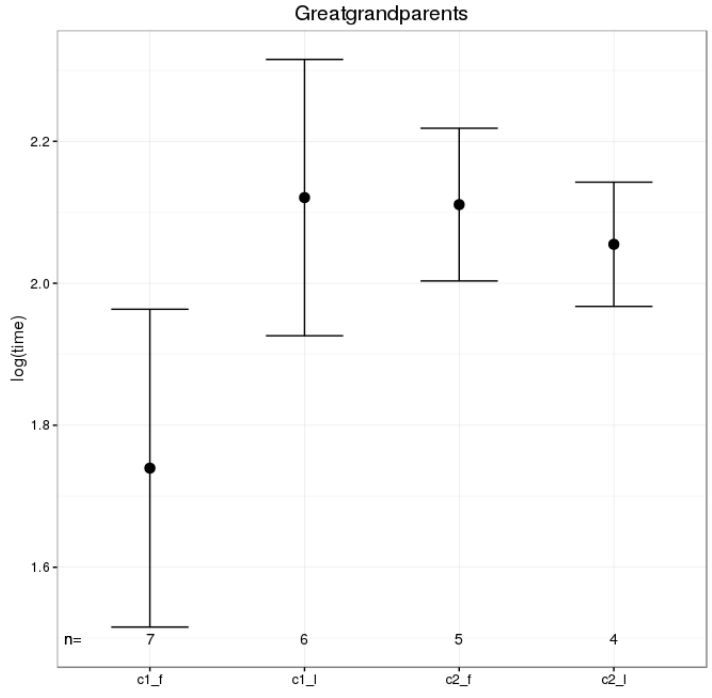

Grandfather
score
8
7
6
5
4
3
2
1
0
n=
7
6
5
4
c1_f
c1_l
c2_f
c2_l

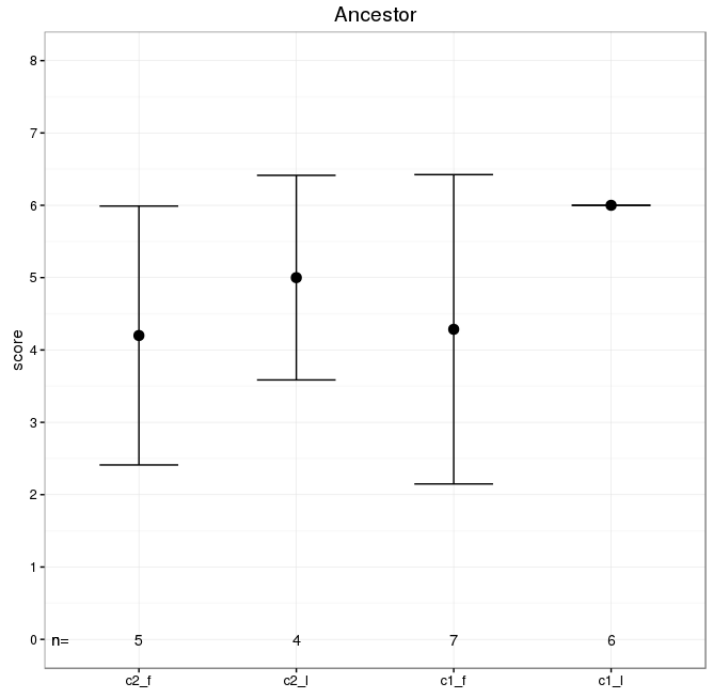

Grandfather
log(time)
2.50
2.25
2.00
1.75
n=
7
6
5
4
c1_f
c1_l
c2_f
c2_l

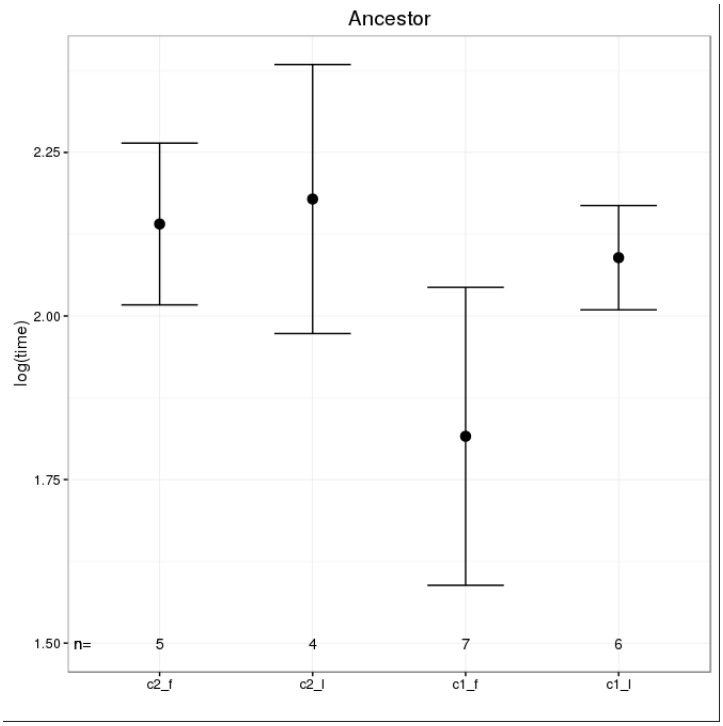

Grandparents
score
8
7
6
5
4
3
2
1
0
n=
5
4
7
6
c2_f
c2_l
c1_f
c1_l


Grandparents
log(time)
2.50
2.25
2.00
1.75
n=
5
4
7
6
c2_f
c2_l
c1_f
c1_l


Greatgrandparents
score
8
7
6
5
4
3
2
1
0
n=
7
6
5
4
c1_f
c1_l
c2_f
c2_l


Greatgrandparents
log(time)
2.2
2.0
1.8
1.6
n=
7
6
5
4
c1_f
c1_l
c2_f
c2_l





4.6.4 R Output

```
> # PIorder 1: PI Problem 1 and 3; PIorder 2: PI Problem 2 and 4

> cknow$PIorder <- relevel(factor(cknow$PIorder), "1")

> contrasts(cknow$PIorder)
  [T.2]
1     0
2     1

> summary(lm(cknow$TIME_log_P1_noNa ~ cknow$PIorder * cknow$FiLa))
```

```
Call:
lm(formula = cknow$TIME_log_P1_noNa ~ cknow$PIorder * cknow$FiLa)

Residuals:
     Min       1Q   Median       3Q      Max
-0.83346 -0.07204  0.00248  0.13914  0.25515

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                        2.11221    0.08995  23.482 5.93e-15 ***
cknow$PIorder[T.2]                 0.09536    0.13935   0.684 0.5025
cknow$FiLa[T.l]                    0.23164    0.13240   1.750 0.0972 .
cknow$PIorder[T.2]:cknow$FiLa[T.l] -0.10924    0.20741  -0.527 0.6048
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.238 on 18 degrees of freedom
Multiple R-squared:  0.1751,    Adjusted R-squared:  0.03763
F-statistic: 1.274 on 3 and 18 DF,  p-value: 0.3134


> cknow$PIorder <- relevel(factor(cknow$PIorder), "2")

> contrasts(cknow$PIorder)
  [T.1]
2     0
1     1

> summary(lm(cknow$TIME_log_P2_noNa ~ cknow$PIorder * cknow$FiLa))

Call:
lm(formula = cknow$TIME_log_P2_noNa ~ cknow$PIorder * cknow$FiLa)

Residuals:
     Min       1Q   Median       3Q      Max
-0.32629 -0.09231 -0.04736  0.10199  0.33386

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                        2.09714    0.07930  26.445 7.4e-16 ***
cknow$PIorder[T.1]                -0.20828    0.10383  -2.006 0.0601 .
cknow$FiLa[T.l]                    0.23414    0.11895   1.968 0.0646 .
cknow$PIorder[T.1]:cknow$FiLa[T.l] -0.03698    0.15454  -0.239 0.8136
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1773 on 18 degrees of freedom
Multiple R-squared:  0.473,    Adjusted R-squared:  0.3852
F-statistic: 5.385 on 3 and 18 DF,  p-value: 0.008005


> cknow$PIorder <- relevel(factor(cknow$PIorder), "1")

> contrasts(cknow$PIorder)
  [T.2]
1     0
2     1
```

```
> summary(lm(cknow$TIME_log_P3_noNa ~ cknow$PIorder * cknow$FiLa))

Call:
lm(formula = cknow$TIME_log_P3_noNa ~ cknow$PIorder * cknow$FiLa)

Residuals:
     Min       1Q   Median       3Q      Max
-0.41739 -0.09372  0.00529  0.11253  0.27898

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                         1.73961    0.06662  26.114 9.23e-16 ***
cknow$PIorder[T.2]                  0.37120    0.10320   3.597 0.00206 **
cknow$FiLa[T.l]                     0.38108    0.09806   3.886 0.00108 **
cknow$PIorder[T.2]:cknow$FiLa[T.l] -0.43698    0.15360  -2.845 0.01075 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1762 on 18 degrees of freedom
Multiple R-squared:  0.5301,    Adjusted R-squared:  0.4517
F-statistic: 6.768 on 3 and 18 DF,  p-value: 0.00299


> cknow$PIorder <- relevel(factor(cknow$PIorder), "2")

> contrasts(cknow$PIorder)
  [T.1]
2     0
1     1

> summary(lm(cknow$TIME_log_P4_noNa ~ cknow$PIorder * cknow$FiLa))

Call:
lm(formula = cknow$TIME_log_P4_noNa ~ cknow$PIorder * cknow$FiLa)

Residuals:
     Min       1Q   Median       3Q      Max
-0.47366 -0.05927  0.01572  0.08016  0.26372

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                         2.14062    0.07681  27.870 2.94e-16 ***
cknow$PIorder[T.1]                 -0.32453    0.10056  -3.227 0.00468 **
cknow$FiLa[T.l]                     0.03815    0.11521   0.331 0.74439
cknow$PIorder[T.1]:cknow$FiLa[T.l]  0.23491    0.14968   1.569 0.13396
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1717 on 18 degrees of freedom
Multiple R-squared:  0.4805,    Adjusted R-squared:  0.3939
F-statistic: 5.549 on 3 and 18 DF,  p-value: 0.007085


> cknow$PIorder <- relevel(factor(cknow$PIorder), "1")

> contrasts(cknow$PIorder)
  [T.2]
```

```
1     0
2     1

> summary(lm(cknow$P1_score ~ cknow$PIorder  * cknow$FiLa))

Call:
lm(formula = cknow$P1_score ~ cknow$PIorder * cknow$FiLa)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2857 -1.6458  0.4333  1.7143  2.2500

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                        5.28571    0.81801   6.462 4.44e-06 ***
cknow$PIorder[T.2]                -0.48571    1.26726  -0.383 0.706
cknow$FiLa[T.l]                    0.04762    1.20408   0.040 0.969
cknow$PIorder[T.2]:cknow$FiLa[T.l] -1.09762    1.88616  -0.582 0.568
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.164 on 18 degrees of freedom
Multiple R-squared:  0.08175,    Adjusted R-squared:  -0.07129
F-statistic: 0.5342 on 3 and 18 DF,  p-value: 0.6647


> cknow$PIorder <- relevel(factor(cknow$PIorder), "2")

> contrasts(cknow$PIorder)
  [T.1]
2     0
1     1

> summary(lm(cknow$P2_score ~ cknow$PIorder * cknow$FiLa))

Call:
lm(formula = cknow$P2_score ~ cknow$PIorder * cknow$FiLa)

Residuals:
    Min      1Q  Median      3Q     Max
-4.8333 -0.8214  0.7143  1.1667  2.8000

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                          4.200      1.028   4.087 0.000692 ***
cknow$PIorder[T.1]                   2.086      1.346   1.550 0.138545
cknow$FiLa[T.l]                      0.800      1.542   0.519 0.610125
cknow$PIorder[T.1]:cknow$FiLa[T.l]  -1.252      2.003  -0.625 0.539612
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.298 on 18 degrees of freedom
Multiple R-squared:  0.1315,    Adjusted R-squared:  -0.01326
F-statistic: 0.9084 on 3 and 18 DF,  p-value: 0.4565


> cknow$PIorder <- relevel(factor(cknow$PIorder), "1")
```

```
> contrasts(cknow$PIorder)
  [T.2]
1     0
2     1

> summary(lm(cknow$P3_score ~ cknow$PIorder  * cknow$FiLa))

Call:
lm(formula = cknow$P3_score ~ cknow$PIorder * cknow$FiLa)

Residuals:
    Min      1Q  Median      3Q     Max
-5.8571 -0.4643  0.5417  1.0571  1.8000

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                         5.8571     0.6886   8.506 1.01e-07 ***
cknow$PIorder[T.2]                 -0.6571     1.0667  -0.616 0.546
cknow$FiLa[T.l]                     0.8095     1.0135   0.799 0.435
cknow$PIorder[T.2]:cknow$FiLa[T.l]  0.2405     1.5877   0.151 0.881
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.822 on 18 degrees of freedom
Multiple R-squared:  0.09484,    Adjusted R-squared:  -0.05602
F-statistic: 0.6287 on 3 and 18 DF,  p-value: 0.6059


> cknow$PIorder <- relevel(factor(cknow$PIorder), "2")

> contrasts(cknow$PIorder)
  [T.1]
2     0
1     1

> summary(lm(cknow$P4_score ~ cknow$PIorder * cknow$FiLa))

Call:
lm(formula = cknow$P4_score ~ cknow$PIorder * cknow$FiLa)

Residuals:
   Min     1Q Median     3Q    Max
-2.286 -0.950  0.000  1.536  1.800

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                        4.20000    0.71669   5.860 1.5e-05 ***
cknow$PIorder[T.1]                 0.08571    0.93837   0.091 0.928
cknow$FiLa[T.l]                    0.80000    1.07504   0.744 0.466
cknow$PIorder[T.1]:cknow$FiLa[T.l] 0.91429    1.39666   0.655 0.521
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.603 on 18 degrees of freedom
Multiple R-squared:  0.211,    Adjusted R-squared:  0.07949
F-statistic: 1.605 on 3 and 18 DF,  p-value: 0.2233
```

4.7 Correct Naming Helps to Profit from Predicate Invention (H4)

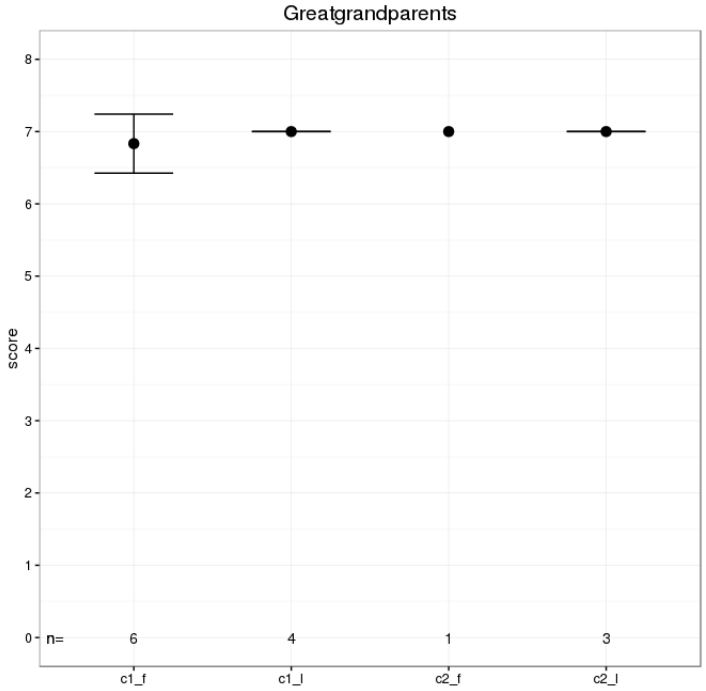
4.7.1 Variables

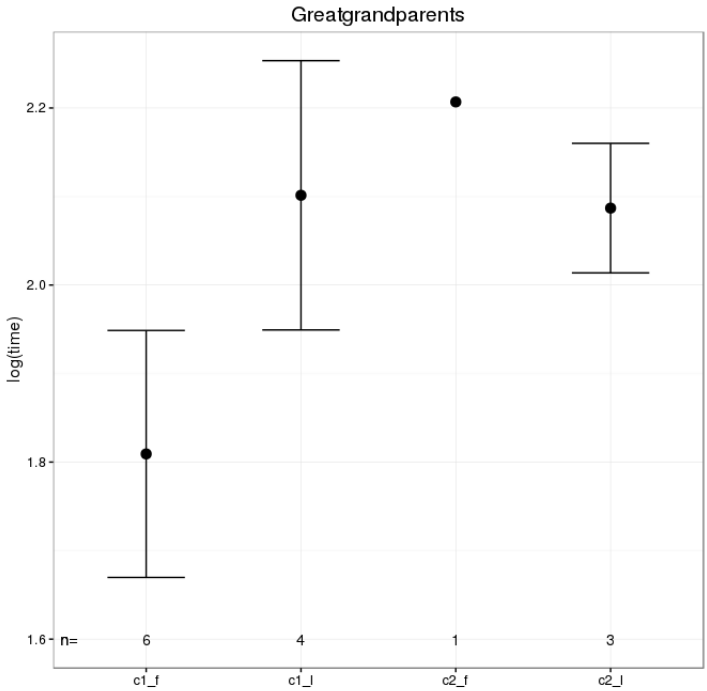
- For greatgrandparent we selected the participants who gave the correct name.

4.7.2 Summary of Results

- All participants with correct naming have maximum scores.
- Predicate invention together with naming first is fastest.
- no statistical analyses due to small n's

4.7.3 Graphs



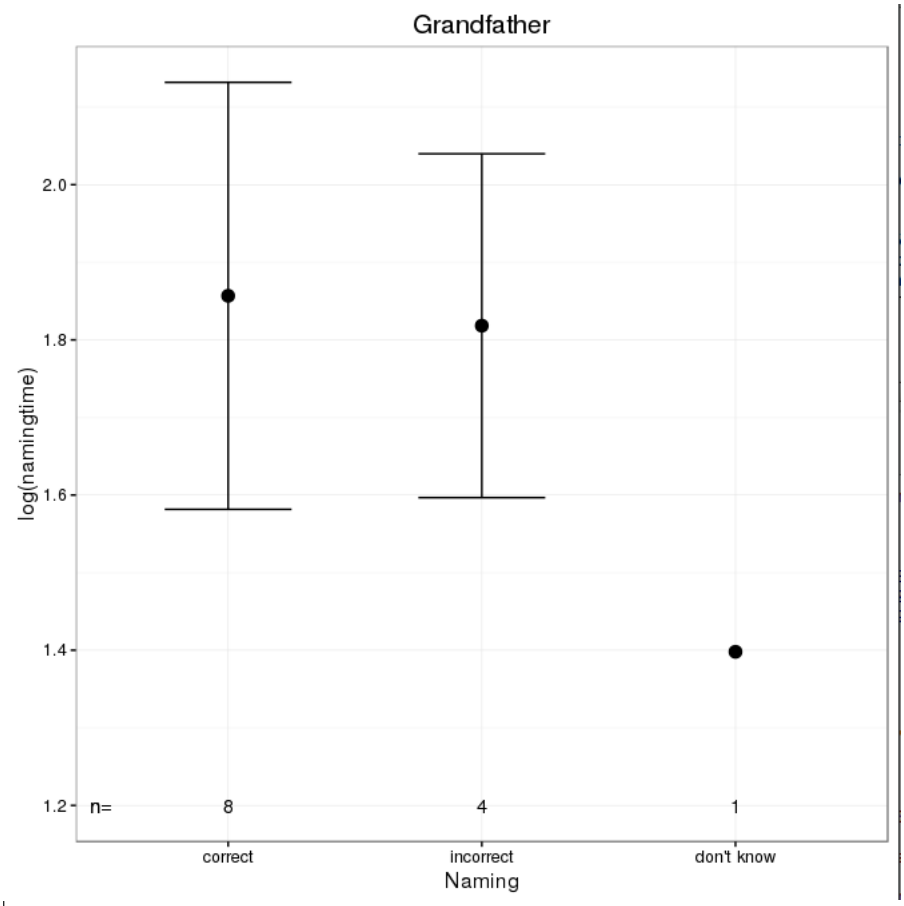


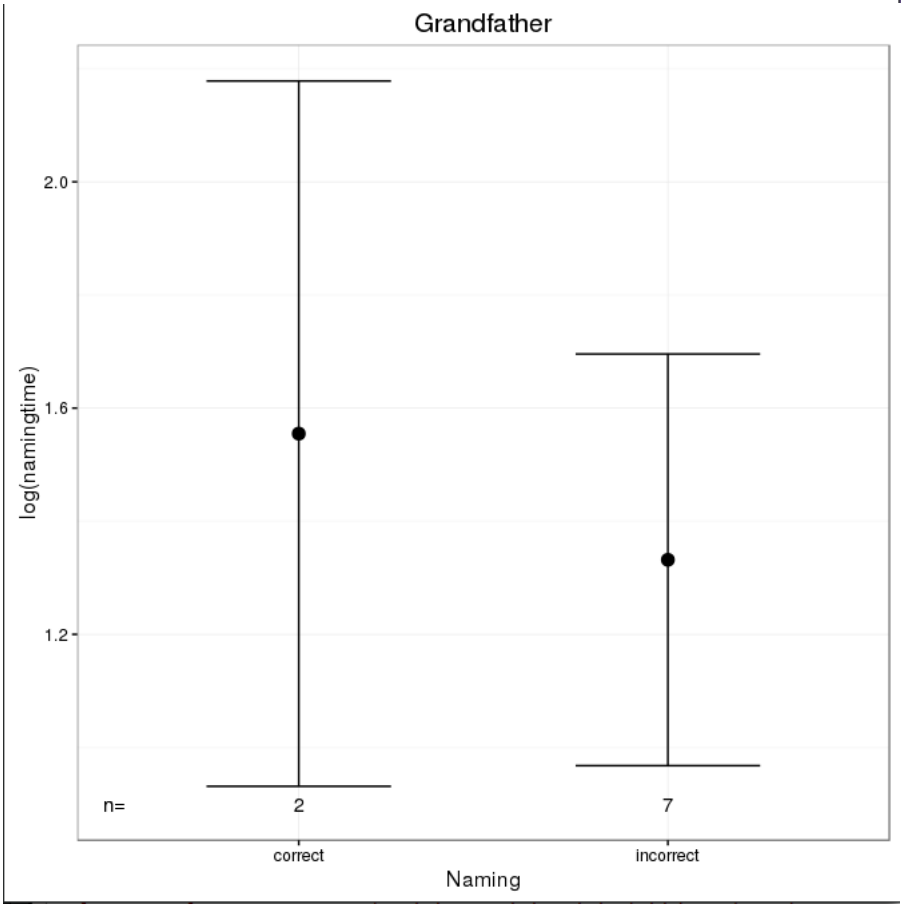
4.8 Long Time for Naming Indicates Incorrectness (H5)

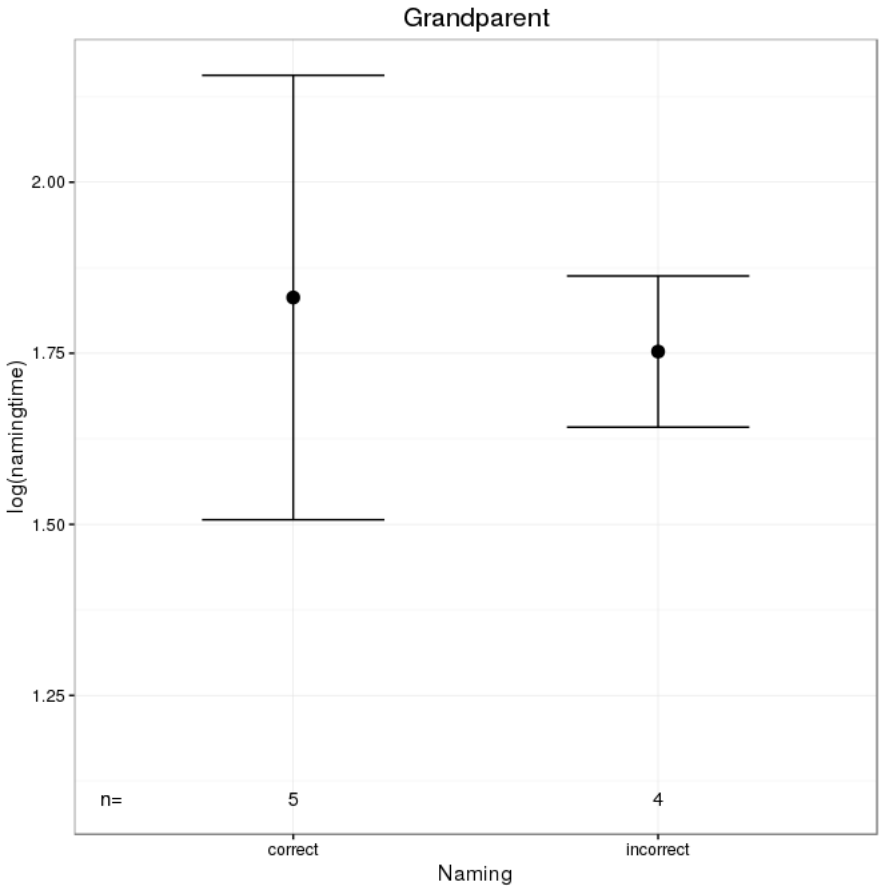
4.8.1 Summary of Results

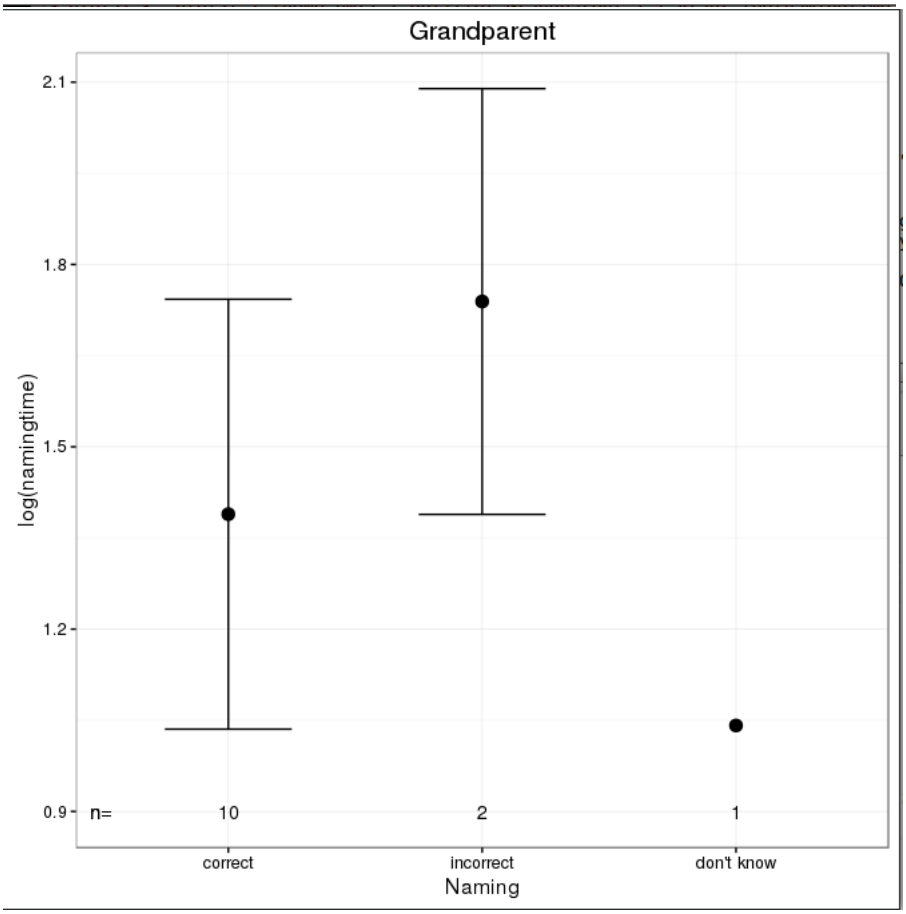
- H5: is partially confirmed for grandparent without predicate invention the trend is in the right direction, for greatgrandparent without predicate invention the difference of time between correct and incorrect is significant, for ancestor without predicate invention marginally significant.
- For greatgrandfather with predicate invention the dontknows are fastest (gave up on understanding).

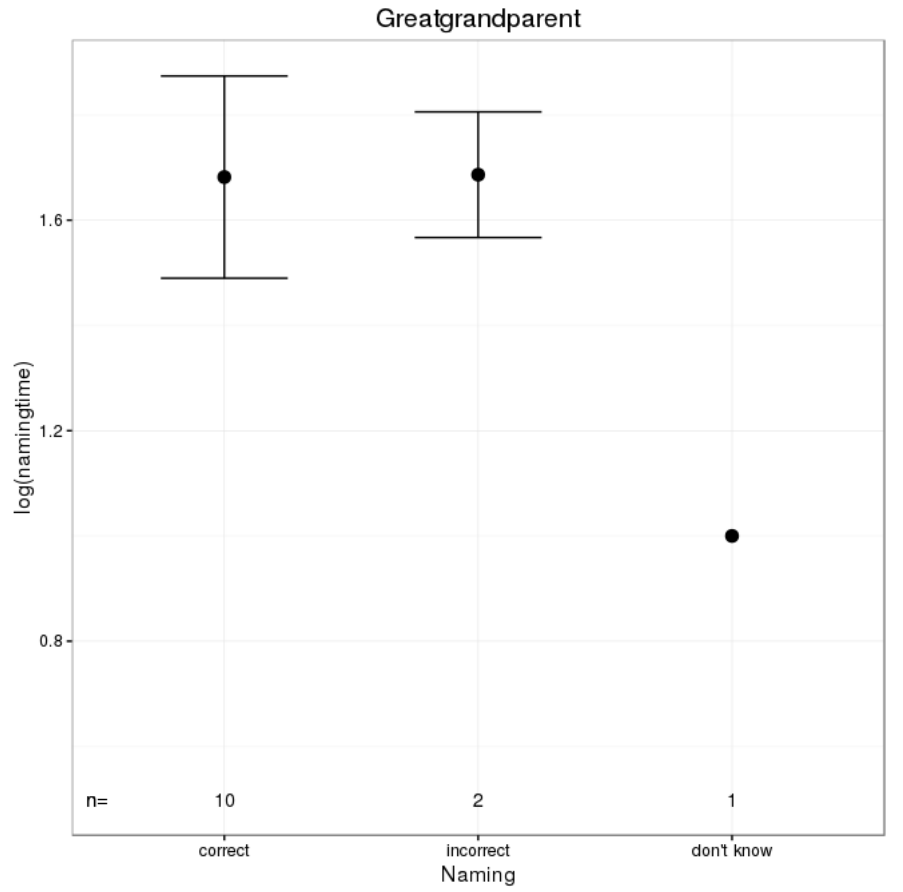
4.8.2 Graphs

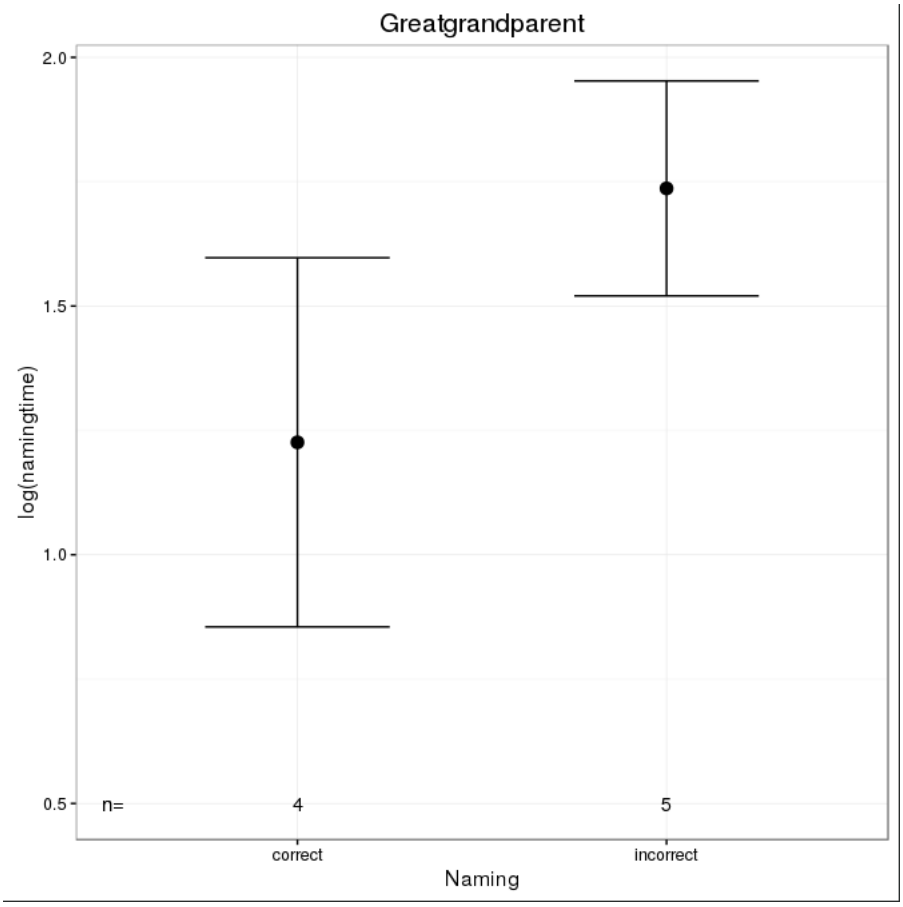

Grandfather
log(namingtime)
2.0
1.8
1.6
1.4
1.2
n=
8
4
1
correct
incorrect
don't know
Naming

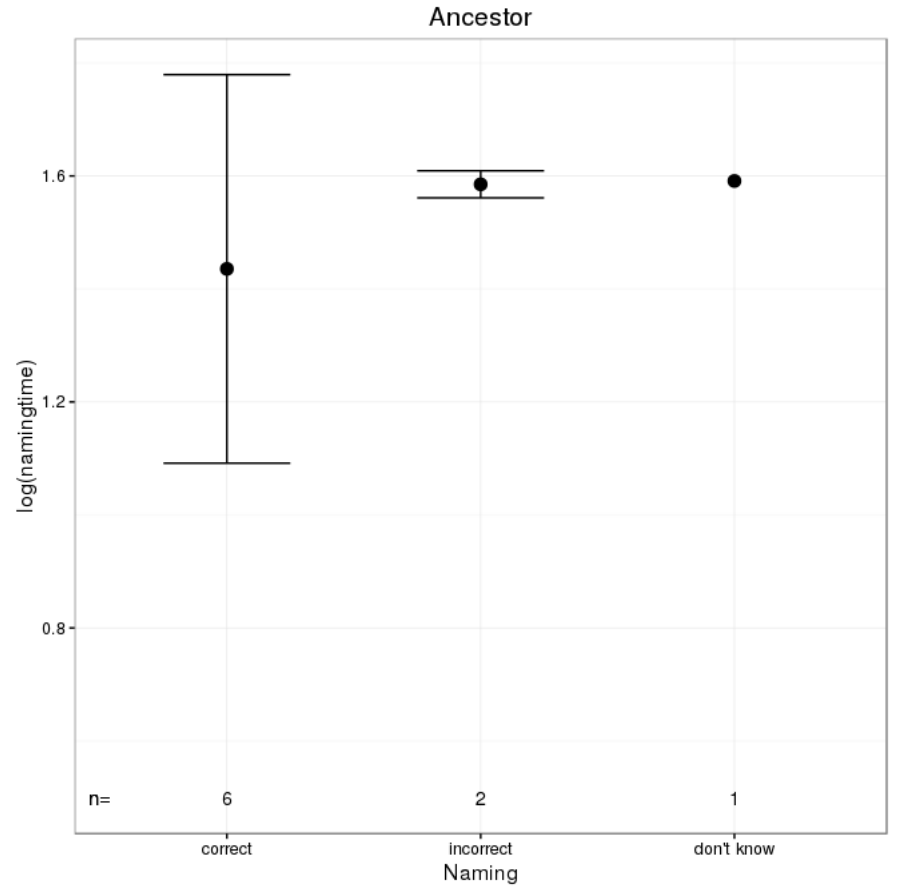

Grandfather
log(namingtime)
2.0
1.6
1.2
n=
2
7
correct
incorrect
Naming

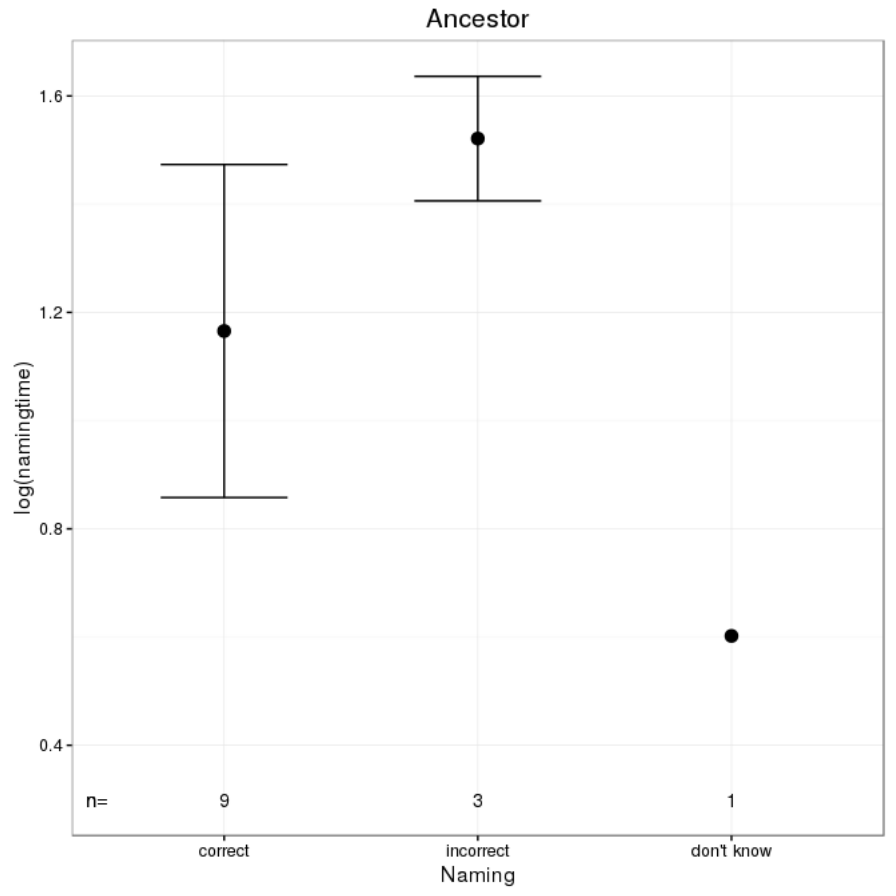

Grandparent
log(namingtime)
2.00
1.75
1.50
1.25
n=
5
4
correct
incorrect
Naming


Grandparent
log(namingtime)
2.1
1.8
1.5
1.2
0.9
n=
10
2
1
correct
incorrect
don't know
Naming


Greatgrandparent
log(namingtime)
1.6
1.2
0.8
n=
10
2
1
correct
incorrect
don't know
Naming


Greatgrandparent
log(namingtime)
2.0
1.5
1.0
0.5
n=
4
5
correct
incorrect
Naming


Ancestor
1.6
1.2
0.8
log(namingtime)
n=
6
2
1
correct
incorrect
don't know
Naming


Ancestor
1.6
1.2
0.8
0.4
log(namingtime)
n=
9
3
1
correct
incorrect
don't know
Naming

4.8.3 R Output

```
> # Problem 1 predicate invention (in piP13)

> piP13$GF16_01 <- relevel(factor(piP13$GF16_01), "1")

> contrasts(factor(piP13$GF16_01))
   [T.0] [T.-2]
1      0      0
0      1      0
-2     0      1

> summary(lm(piP13$TIME_log_P1_Na ~ piP13$GF16_01))

Call:
lm(formula = piP13$TIME_log_P1_Na ~ piP13$GF16_01)

Residuals:
    Min      1Q  Median      3Q     Max
-0.4765 -0.1407  0.0330  0.1603  0.3016

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.85675    0.09202  20.178 1.97e-09 ***
piP13$GF16_01[T.0]  -0.03850    0.15938  -0.242    0.814
piP13$GF16_01[T.-2] -0.45881    0.27605  -1.662    0.127
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2603 on 10 degrees of freedom
Multiple R-squared:  0.2168,    Adjusted R-squared:  0.06021
F-statistic: 1.384 on 2 and 10 DF,  p-value: 0.2946

> # Problem 1 no predicate invention (in piP24)

> piP24$GF16_01 <- relevel(factor(piP24$GF16_01), "1")

> contrasts(factor(piP24$GF16_01))
  [T.0]
1     0
0     1

> summary(lm(piP24$TIME_log_P1_Na ~ piP24$GF16_01))

Call:
lm(formula = piP24$TIME_log_P1_Na ~ piP24$GF16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.48666 -0.44085  0.08322  0.33100  0.44085

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.5548     0.2907   5.348  0.00107 **
piP24$GF16_01[T.0]  -0.2230     0.3296  -0.677  0.52040
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.4111 on 7 degrees of freedom
Multiple R-squared:  0.06138,    Adjusted R-squared:  -0.07271
F-statistic: 0.4578 on 1 and 7 DF,  p-value: 0.5204

> # Problem 2 predicate invention (in piP24)

> piP24$GP16_01 <- relevel(factor(piP24$GP16_01), "1")

> contrasts(factor(piP24$GP16_01))
  [T.0]
1     0
0     1

> summary(lm(piP24$TIME_log_P2_Na ~ piP24$GP16_01))

Call:
lm(formula = piP24$TIME_log_P2_Na ~ piP24$GP16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36906 -0.10904  0.00338  0.15059  0.32388

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.83145    0.11445   16.00 9.04e-07 ***
piP24$GP16_01[T.0] -0.07896    0.17167   -0.46     0.66
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2559 on 7 degrees of freedom
Multiple R-squared:  0.02933,    Adjusted R-squared:  -0.1093
F-statistic: 0.2115 on 1 and 7 DF,  p-value: 0.6595

> # Problem 2 no predicate invention (in piP13)

> piP13$GP16_01 <- relevel(factor(piP13$GP16_01), "1")

> contrasts(factor(piP13$GP16_01))
   [T.0] [T.-2]
1      0      0
0      1      0
-2     0      1

> summary(lm(piP13$TIME_log_P2_Na ~ piP13$GP16_01))

Call:
lm(formula = piP13$TIME_log_P2_Na ~ piP13$GP16_01)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6901 -0.2430  0.0000  0.2921  0.3962

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)           1.3891     0.1117  12.432 2.09e-07 ***
piP13$GP16_01[T.0]    0.3500     0.2737   1.279     0.23
piP13$GP16_01[T.-2]  -0.3477     0.3706  -0.938     0.37
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3533 on 10 degrees of freedom
Multiple R-squared:  0.222,    Adjusted R-squared:  0.06642
F-statistic: 1.427 on 2 and 10 DF,  p-value: 0.285

> # Problem 3 predicate invention (in piP13)

> piP13$PP16_01 <- relevel(factor(piP13$PP16_01), "1")

> contrasts(factor(piP13$PP16_01))
   [T.0] [T.-2]
1      0      0
0      1      0
-2     0      1

> summary(lm(piP13$TIME_log_P3_Na ~ piP13$PP16_01))

Call:
lm(formula = piP13$TIME_log_P3_Na ~ piP13$PP16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.32017 -0.09083 -0.00066  0.13765  0.28659

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.681897   0.058865  28.572 6.42e-11 ***
piP13$PP16_01[T.0]   0.004559   0.144190   0.032   0.9754
piP13$PP16_01[T.-2] -0.681897   0.195235  -3.493   0.0058 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1861 on 10 degrees of freedom
Multiple R-squared:  0.5539,    Adjusted R-squared:  0.4647
F-statistic: 6.208 on 2 and 10 DF,  p-value: 0.01767

> # Problem 3 no predicate invention (in piP24)

> piP24$PP16_01 <- relevel(factor(piP24$PP16_01), "1")

> contrasts(factor(piP24$PP16_01))
  [T.0]
1     0
0     1

> summary(lm(piP24$TIME_log_P3_Na ~ piP24$PP16_01))

Call:
lm(formula = piP24$TIME_log_P3_Na ~ piP24$PP16_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27179 -0.21802  0.01165  0.12678  0.51433

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.2260     0.1464   8.376 6.79e-05 ***
```

```
piP24$PP16_01[T.0]   0.5105     0.1964   2.599   0.0355 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2928 on 7 degrees of freedom
Multiple R-squared:  0.4912,    Adjusted R-squared:  0.4185
F-statistic: 6.757 on 1 and 7 DF,  p-value: 0.03545

> # Problem 4 predicate invention (in piP24)

> piP24$AN18_01 <- relevel(factor(piP24$AN18_01), "1")

> contrasts(factor(piP24$AN18_01))
   [T.0] [T.-2]
1      0      0
0      1      0
-2     0      1

> summary(lm(piP24$TIME_log_P4_Na ~ piP24$AN18_01))

Call:
lm(formula = piP24$TIME_log_P4_Na ~ piP24$AN18_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.25935 -0.20500 -0.01693  0.01693  0.57739

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)           1.4354     0.1282  11.196 3.03e-05 ***
piP24$AN18_01[T.0]    0.1497     0.2564   0.584    0.581
piP24$AN18_01[T.-2]   0.1556     0.3392   0.459    0.663
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3141 on 6 degrees of freedom
Multiple R-squared:  0.07216,    Adjusted R-squared:  -0.2371
F-statistic: 0.2333 on 2 and 6 DF,  p-value: 0.7988

> # Problem 4 no predicate invention (in piP13)

> piP13$AN18_01 <- relevel(factor(piP13$AN18_01), "1")

> contrasts(factor(piP13$AN18_01))
   [T.0] [T.-2]
1      0      0
0      1      0
-2     0      1

> summary(lm(piP13$TIME_log_P4_Na ~ piP13$AN18_01))

Call:
lm(formula = piP13$TIME_log_P4_Na ~ piP13$AN18_01)

Residuals:
     Min       1Q   Median       3Q      Max
-0.32047 -0.16556 -0.01604  0.12226  0.58262
```

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.16556    0.09334  12.488 2.01e-07 ***
piP13$AN18_01[T.0]   0.35563    0.18667   1.905   0.0859 .
piP13$AN18_01[T.-2] -0.56350    0.29515  -1.909   0.0853 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.28 on 10 degrees of freedom
Multiple R-squared:  0.4635,    Adjusted R-squared:  0.3562
F-statistic:  4.32 on 2 and 10 DF,  p-value: 0.04443
```

5 Naming Answers

5.1 Osnabrueck

```
> # Problem 1: p/2

> p1_p_correct <- cknow[cknow$GF16_01 == 1,]

> p1_p_incorrect <- cknow[cknow$GF16_01 == 0,]

> p1_p_empty <- cknow[cknow$GF16_01 == -1,]

> p1_p_dont <- cknow[cknow$GF16_01 == -2,]

> na.omit(p1_p_correct$GF08_01)
 [1] "check if X is the grandfather of Y (_father_ node of Z that is parent of Y)"
 [2] "grandfather"
 [3] "grandson"
 [4] "test if X is grandfather of Y"
 [5] "\"grandfather\", p(X,Y) is true, if X is the father of somebody, who is father
or mother of Y."
 [6] "it checks whether some node is the grandfather of some other node"
 [7] "looks for nodes with one degree of separation (granddad <-> grandson)"
 [8] "grandfather"
 [9] "X is grandfather of Y"
[10] "X is the grandfather of Y"
[11] "grandfather relation"
[12] "give node and grandchildren / grandmothers and node"
[13] "grandfather"
[14] "Grandfather"
[15] "it shows the grandson/grandparents nodes"
[16] "Grandfather"
[17] "checks if X is the grandmother node of Y"
[18] "checks, whether Y is a granddaughter node (?) of X"
[19] "grandfather"
[20] "Check for grandfather"
[21] "It tells you whether X is a grandfather of Y."
[22] "node-grandson relationship"
[23] "Checks if X is grandfather of Y"
[24] "x is grandfather of y"
[25] "grandfather"
[26] "grandfather"
[27] "grandfather"
[28] "grandfather"


> na.omit(p1_p_incorrect$GF08_01)
 [1] "parents\n"
 [2] "siblings and grandchildren"
 [3] "Find out if X is grandparent node of Y"
 [4] "grandparent\np/2 distinguishes whether x is a grandparent of y"
 [5] "finds grandchild / -parents pairs"
 [6] "one parent node"
 [7] "descendant"
 [8] "checks if X is a grandparent of Y"
 [9] "finding the middle-person in the relationship of grandparent and grandchild"
[10] "Assoziativgesetz"
```

```
[11] "First line defines grandfather relation\nSecond line defines father relation +
there is a mother documented in the tree"
[12] "sibling"
[13] "parent and child relation  and parent and child relationg and grandparent and
grandchild relation and parent and child relation"
[14] "has child with"
[15] "X related to Y?"
[16] "grandparent/grandchildren relation but only between \"true\" members of the
family tree  (descendants of aa, ab)"
[17] "doesn't take Z into account "
[18] "Gives a node and it's grandparent node"
[19] "grandparents search"
[20] "Grandchild"
[21] "detects the grandparents of a child"
[22] "relation grandparent or direct parent"
[23] "finds the child node of X in q1\nand the parent for q2"
[24] "They are looking for the grandchild node? I have no idea what this is called,
but something under two nodes"
[25] "parent"
[26] "check if x and y are connected via an internode z"
[27] "grandparents"
[28] "grandchild"
[29] "parents"
[30] "grandparent"
[31] "sibling nodes belonging to the same parent"
[32] "finds grandchildren Y is grandchild of X or Y is the strange not boxed sibling of X"
[33] "grandparents"
[34] "relationship between grandparent(?) nodes and grandchild(?) nodes. As in a is
parent of b, and b is parent of c therefore a is grandparent of c. "
[35] "2 children with same parent node"
[36] "fraternal grandparents "
[37] "grandchildren"
[38] "Grandchildren and Parents"
[39] "p checks relationship of nodes to \"grandchildren\"-nodes"
[40] "get neighbors of X"
[41] "p includes 2 elements"
[42] "connect nodes"
[43] "X is parent 1 of parent 1 of Y, or is parent 1 and there is a parent 2"
[44] "children/parents"
[45] "children from the same mother, children who have the same mother but different fathers."
[46] "gives child of child"


> na.omit(p1_p_dont$GF08_01)
[1] "i don't know"
[2] "sorry, is p/2 \"the code of the problem\" in general? i don't know ..."
[3] "don't know"



> # Problem 1: p1/2 >>> pt

> p1_pt_correct <- cknow[cknow$GF18_01 == 1,]

> p1_pt_incorrect <- cknow[cknow$GF18_01 == 0,]

> p1_pt_empty <- cknow[cknow$GF18_01 == -1,]

> p1_pt_dont <- cknow[cknow$GF18_01 == -2,]

> na.omit(p1_pt_correct$GF17_01)
 [1] "check if X is the parent node (mother or father) of Y"
 [2] "Find out if X is parent node of Y"
 [3] "finds parents of Y"
 [4] "child node"
 [5] "test if X is father or mother of Y"
 [6] "parent"
 [7] "parent/child relation"
 [8] "parent child relation"
 [9] "Mother or father"
[10] "Parent"
[11] "Parent"
[12] "if X and Y is parent node and childe node or not"
```

```
[13] "parent"
[14] "check if a direct connection for x and y exists"
[15] "Check for parent"
[16] "parent"
[17] "son or daughter"
[18] "relationship between parent nodes and child nodes."
[19] "parents"
[20] "children"
[21] "parent"
[22] "parent"
[23] "gets child"


> na.omit(p1_pt_incorrect$GF17_01)
 [1] "sub parents"
 [2] "check for specific grandson in q1, else check in q2"
 [3] "checks if Y is a child of X and if X has a parent"
 [4] "p1 has to be a statement of q1 or q2\n"
 [5] "where is p/2 ? "
 [6] "True if there are the same pair twice in p1 or the same pair in p1 and p2"
 [7] "searchspace"
 [8] "detects relative relationships"
 [9] "checks if X is the mother node of Y"
[10] "returns all granddaughter nodes/ grandmother nodes"
[11] "grandfather"
[12] "node-son relationship"
[13] "grandparents"
[14] "parent of parent with 2 children node"
[15] "It gets the Y for which there exists a statement with X."
[16] "connect p1 and 2"
[17] "grandparents"


> na.omit(p1_pt_empty$GF17_01)
[1] "no"

> # Problem 3: p/2

> p3_p_correct <- cknow[cknow$PP16_01 == 1,]

> p3_p_incorrect <- cknow[cknow$PP16_01 == 0,]

> p3_p_empty <- cknow[cknow$PP16_01 == -1,]

> p3_p_dont <- cknow[cknow$PP16_01 == -2,]

> na.omit(p3_p_correct$PP08_01)
 [1] "check if X is great grandparent node of Y (X is parent of U who is parent
of Z who is parent of Y)"
 [2] "greatgrandchildren and lots of unused statements"
 [3] "Find out if X is great-grandparent node of Y"
 [4] "grandgrandparent\ndistinguishes whether X is a grandgrandparent of Y"
 [5] "greatgrandparent"
 [6] "finds pairs of grandgrandparent and -child"
 [7] "grand grand parent"
 [8] "parent of grandparent relation from X to Y"
 [9] "test if X is great grandparents of Y"
[10] "\"great grandparent\": p(X,Y) is true, if X is the parent of somebody, who is
the parent of somebody, who is the parent of Y.\n\n(Note: The definition of p/2 can be optimized. )"
[11] "defines the grand-grandparent relation"
[12] "grandgrandparent and grandgrandchild\n"
[13] "Greatgranparent"
[14] "Two generation relations between X and Y"
[15] "greatgrandmother or greatgrandfather"
[16] "X is a great-grandparent of Y"
[17] "X is a great-grandparent of Y"
[18] "Great-grandparents"
[19] "great-grandparent/grandchild relation"
[20] "Great grandparent nodes"
[21] "grandgrandparents relation"
[22] "grand grand parents"
[23] "Is x greatgrandparent of y?"
[24] "gives the names of the great-grandparents and their great-grandchildren"
```

```
[25] "grandgrandparent"
[26] "grandparent-child for any nodes (boxed or unboxed)"
[27] "Great Grandparent"
[28] "it show grandgrandpartents nodes"
[29] "Great-Grandparent"
[30] "if it is three nodes down from X"
[31] "grand-grandparent or parent of a parent of a parent of Y"
[32] "greatgrandparent"
[33] "checks if two nodes x and y are connected via two internodes u and z"
[34] "Check for great-grandparent"
[35] "It tells you whether X is a greatgrandparent of Y."
[36] "grand grand parent"
[37] "Checks if X is great-grandparents of Y"
[38] "X is the great-grandparent of Y, can be used to determine who are X's great-grandchilds,
to determine who is the great-grandfather/mother of Y or to check if the relation between X and
Y is great-grandparent - great-grandchild"
[39] "great grand children"
[40] "greatgrandparents or greatgrandchildren"
[41] "grand-grandparent (from either side of the family)"
[42] "finds great grandchildren"
[43] "parents of grandparents, "
[44] "relationships between great-grandparent nodes and great-grandchild nodes."
[45] "greatgrandparent"
[46] "grandgrandchildren"
[47] "Checks for Great Grandchildren"
[48] "great-grandchild"
[49] "nodes to greatgrandchildren"
[50] "greatgrandparent"
[51] "great grandparent"
[52] "great_grandparent"
[53] "grandchildren...maybe"
[54] "find grand children, from either grandparent."


> # Problem 2: p/2

> p2_p_correct <- cknow[cknow$GP16_01 == 1,]

> p2_p_incorrect <- cknow[cknow$GP16_01 == 0,]

> p2_p_empty <- cknow[cknow$GP16_01 == -1,]

> p2_p_dont <- cknow[cknow$GP16_01 == -2,]

> na.omit(p2_p_correct$GP08_02)
 [1] "check if X is a grandparent node (father/mother of Z that is
father/mother of Y) of Y"
 [2] "grandchildren"
 [3] "Find out if X is grandparent node of Y"
 [4] "grandparent\ndistinguishes whether  x is a grandparent of y"
 [5] "grandparent"
 [6] "finds pairs of grandparent and grandchild"
 [7] "looking for grandchilds"
 [8] "grandparent relation from X to Y"
 [9] "test if X is grandparents of Y"
[10] "\"grandparent\". p(X,Y) is true, of X is the parent of somebody,
who is the parent of Y."
[11] "defines the grandparent relation. So p(x,?) returns all Enkel from x,
while p(?,x) returns all grandparents from x"
[12] "grandparent and grandchild relation"
[13] "grandparent"
[14] "it takes connections from q1 and q2 or from only q1, only q2.
2 degrees of separation through the whole tree"
[15] "grandmother or grandfather"
[16] "X is a grandparent of Y"
[17] "X is a grandparent of Y"
[18] "Grandparents number two"
[19] "grandparent/grandchild relation (this time between all possible nodes)"
[20] "Grandparent nodes\n"
[21] "grandparents relation"
[22] "grandparents"
[23] "Is x any grandparent of y ?"
[24] "gives the name of grandparents and their grandchildren"
```

```
[25] "grandparent"
[26] "Grandparent"
[27] "Grandparent"
[28] "if it is those grandchild node"
[29] "grandparent"
[30] "grandparent"
[31] "check if two nodes x and y are connected via an internode z"
[32] "Check for grandparent"
[33] "It tells you whether X is a grandparent of Y."
[34] "Checks if X is a grandparent of Y"
[35] "x is grandparent, either male or female, y is grandchild\n\np splits
 x and y in such way that they are connected over a mutual value/person z"
[36] "grandchild"
[37] "grandparents or grandchildren"
[38] "grandparents"
[39] "grandparent (from either side of the family)"
[40] "Evaluate if Y is a grandchild of X"
[41] "finds grandchildren"
[42] "relationship between grandparent nodes and grandchild nodes"
[43] "grandparent"
[44] "Grandchildren"
[45] "grandparent"
[46] "grandparent"
[47] "grandparent"
[48] "grandparent"
[49] "find grandparents, returns grandparents of children."
[50] "gives child of child"


> na.omit(p2_p_incorrect$GP08_02)
 [1] "Mismatch"
 [2] "descendant"
 [3] "grandson for q1 or q2"
 [4] "reversing"
 [5] "Assoziativgesetz und kommutativgesetz auf q1 und q2"
 [6] "branch alternatives"
 [7] "p/2 returns all parent nodes of a single node x and all nodes
for which x is a parent node"
 [8] "X grandson of Y?"
 [9] "X=grandmother node\nY=grandchildren"
[10] "ancestor relation"
[11] "x is the father of z and z is the mother of of y "
[12] "grandmother node relation"
[13] "for q1 and q2 they both must be connecting p in the tree"
[14] "the same as the one before, but now it includes those nodes
who are not direct descendents of aa or ab"
[15] "who are the partens"
[16] "relation node-son son-node"
[17] "parents"
[18] "parent of parent with 2 children "
[19] "Check for Parents"
[20] "same as problem 1"
[21] "related or not"
[22] "implies two elements"
[23] "descendent"


> na.omit(p2_p_empty$GP08_02)
[1] "/"


> na.omit(p2_p_dont$GP08_02)
[1] "don't know"
[2] "no idea sorry"
[3] "no idea"
[4] "dont know"


> # Problem 2: p1/2 >>> pt

> p2_pt_correct <- cknow[cknow$GP18_01 == 1,]

> p2_pt_incorrect <- cknow[cknow$GP18_01 == 0,]
```

```
> p2_pt_empty <- cknow[cknow$GP18_01 == -1,]

> p2_pt_dont <- cknow[cknow$GP18_01 == -2,]

> na.omit(p2_pt_correct$GP17_02)
 [1] "parent\ndistinguishes whether X is a father or a mother of Y"
 [2] "parent"
 [3] "of kind q1 or q2"
 [4] "\"parent\". p1(X,Y) is true, if X is father or mother of Y."
 [5] "p1 defines the parent relation. "
 [6] "it checks whether some node is a parent node of another node"
 [7] "selectes either q1 or q2"
 [8] "X is a parent of Y"
 [9] "X is a parent of Y"
[10] "parent"
[11] "parent"
[12] "x is the father or the mother of y"
[13] "It tells you whether X is a parent of Y."
[14] "who are the parents"
[15] "Checks if X is father or mother of Y"
[16] "parent (from either side of the family - you can reach married persons by stepping the tree downwards and upwards again)"
[17] "finds children"
[18] "Check for one Parent"
[19] "checks if X is father/mother from Y"
[20] "parent"
[21] "parent"


> na.omit(p2_pt_incorrect$GP17_02)
 [1] "nodes of tree"
 [2] "accepting to check tree in reversed direction"
 [3] "one brance alternative selected"
 [4] "Relations regardless of heritage"
 [5] "Exchenging granddad and grandma"
 [6] "parent if parent is boxed"
 [7] "p1/2 must take the arguments of q1/2 and q2/2"
 [8] "x is grandparent, either male or female, y is grandchild\n\np1 checks if x is the father of y (in this case, it checks if
 [9] "Evaluate a relation between X and Y"
[10] "siblings"
[11] "implies half of the element"
[12] "the parents of parents"


> na.omit(p2_pt_empty$GP17_02)
[1] "/"


> na.omit(p2_pt_dont$GP17_02)
[1] "nothing"
[2] "don't know"
[3] "no idea "


> na.omit(p3_p_incorrect$PP08_01)
 [1] "grandgrandson"
 [2] "one more step in between than before"
 [3] "Assoziativgesetz + Assoziativgesetz"
 [4] "it checks whether two nodes are connected with 2 nodes in between ...
relation = grand grandfather?"
 [5] "finds nodes with 3 degrees of seperation (grandgrand father)"
 [6] "parents"
 [7] "give grand grandmother and grand grandchildren of a node"
 [8] "combiner"
 [9] "ancester relation (great grand mother)"
[10] "grand-granddaughter checker"
[11] "node-greatson relation"
[12] "parents"
[13] "Evaluate if X is a great grandfather of Y"
[14] "parent of parent of parent with 2 children"
[15] "Has X a grandfather."
[16] "implies element not fully"
[17] "connect nodes"
```

```
[18] "gives children"


> na.omit(p3_p_empty$PP08_01)
[1] "no" "/"


> na.omit(p3_p_dont$PP08_01)
[1] "Phone Tag"
[2] "don't know"
[3] "no idea"
[4] "I dont understand the relation between the tree graph and the output right
next to it. For the example it made sense to me, now it doesnt anymore."

> # Problem 3: p1/2 >>> pt

> p3_pt_correct <- cknow[cknow$PP18_01 == 1,]

> p3_pt_incorrect <- cknow[cknow$PP18_01 == 0,]

> p3_pt_empty <- cknow[cknow$PP18_01 == -1,]

> p3_pt_dont <- cknow[cknow$PP18_01 == -2,]

> na.omit(p3_pt_correct$PP17_01)
 [1] "check if X is either mother or father of Y"
 [2] "Find out if X is parent node of Y"
 [3] "finds pairs of parent and child"
 [4] "parent relation"
 [5] "test if X is parent of Y"
 [6] "p1 has to be of q1 or q2"
 [7] "Parent"
 [8] "parent"
 [9] "parent/child relation"
[10] "parents relation"
[11] "Is x parent of x?"
[12] "Parent"
[13] "Parent"
[14] "if it is q1 or q2, so p1 makes them same (no differences between q1 and q2)"
[15] "parent"
[16] "parent"
[17] "check if there is a direct conntection from x to y"
[18] "Check for parent"
[19] "parents"
[20] "mother or father"
[21] "relationships between parent nodes and child nodes."
[22] "parent"
[23] "children"
[24] "parent"
[25] "gets the parent of Y."
[26] "parent"
[27] "children"


> na.omit(p3_pt_incorrect$PP17_01)
[1] "grand grand child"
[2] "checking if specific grandgrandson exist"
[3] "Assigns a p1 node with a pair of p1 or p2"
[4] "searchspace q1 and q2"
[5] "the same"
[6] "checks if X mother node of Y"
[7] "lists all grand-granddaugthers/-mothers"
[8] "node-son relation"
[9] "greatgrandparents\n"
[10] "connect nodes"
[11] "searches nodes"


> na.omit(p3_pt_dont$PP17_01)
[1] "Bottom of the barrel" "dont know"
[2] "i dont know"

> # Problem 4: p/2
```

```
> p4_p_correct <- cknow[cknow$AN18_01 == 1,]

> p4_p_incorrect <- cknow[cknow$AN18_01 == 0,]

> p4_p_empty <- cknow[cknow$AN18_01 == -1,]

> p4_p_dont <- cknow[cknow$AN18_01 == -2,]

> na.omit(p4_p_correct$AN08_01)
 [1] "check if X is in any way related to Y (is X parent of Y or parent of Z who is
parent of Y, recursive tree traversal from root to leaf nodes maximally)"
 [2] "if node X is an ancestor of node Y"
 [3] "Is Y reachable from X"
 [4] "descendant"
 [5] "parent grandparent or parent of grandparent relation"
 [6] "test if X is an ancestor of Y"
 [7] "\"ancestor\". p(X,Y) is true, if X is the parent of Y of if X is the grandparent
of Y or if X is the grand-grandparent of Y or..."
 [8] "Recursively defines a heritance-relation. Succeeds if X is antecedent of Y
but not if Y is antecedent of X (antecedent = Vorfahre I hope) "
 [9] "only parent child relations of those that are 'true' anticidents. so parent
who is 'married' to the 'family' is part of a parent child relation. but all children
are part of a parent child relation."
[10] "it checks whether some node preceeds some node y regardless the number of
nodes in between"
[11] "Descendents"
[12] "X is a descendant of Y. "
[13] "ancestor/descendant relation (over any number of generations)"
[14] "ancestors"
[15] "give all following nodes of X"
[16] "gives the name of all descendants of a person"
[17] "ancestor"
[18] "ancestor relation"
[19] "Ancestor"
[20] "ancester relation, true when either mother node or grandmother node"
[21] "checks wheter Y is a descendent of X"
[22] "ancestor"
[23] "Check for direct line in familiy tree, with X higher up than Y"
[24] "It tells you whether X is a parent/grandparent/greatgrandparent/... of Y."
[25] "Checks in the first case if X is a parent of Y, second case if X is in any
relation a parent of Y (grandparents, great grandparents, etc....)"
[26] "true if X is parent grandparent or greatgrandparent of Y and vice versa"
[27] "decendent"
[28] "Children, Grandchildren, Grandchildren and so on"
[29] "ancestor"
[30] "ancestor"
[31] "p/2 checks if there is an ancestor relation (parent, grandparent, greatgrandparent
in the example) or finds all ancestors/decendants"
[32] "ancestor"


> na.omit(p4_p_incorrect$AN08_01)
 [1] "To be recursive or not to be"
 [2] "parentOrGrandparent\nDistinguishes whether X is a parent or a grandparent of Y"
 [3] "parent or grandparent"
 [4] "finds pairs of related people"
 [5] "allowing one middle step searching for relations"
 [6] "It checks firstly whether q1 and q2 exists. therefore, it should return true
regardless of the other constraints"
 [7] "one step function"
 [8] "Either a parent or a grandparent"
 [9] "recursive search through the tree"
[10] "Relation by son or grandson"
[11] "parent or grandparent"
[12] "X is either a parent OR grandparent of Y"
[13] "Parents and grandparents"
[14] "Searches if pair exists"
[15] "Is x parent or greatparent of Y?"
[16] "check for relation between two nodes"
[17] "Parent and grandparent"
[18] "it shows paretns and or grand parent nodes"
[19] "p must be the end node of a parent p1"
```

```
[20] "p(X, Y) is parent node/child node\np(X, Y) is grand parent node/grandchild
node(is this term exists?)"
[21] "parent/grandparent"
[22] "true if there ist a direct connection from x to y or a connection via an internode z"
[23] "real parents"
[24] "relation node-son grandson-son"
[25] "checks if X and Y are in any way related. first checks if there is a direct
relationship (father/mother), "
[26] "gradn children"
[27] "parent or grandparent"
[28] "Evaluate if Y is a child of X."
[29] "finds child or grandchild"
[30] "parents or themselves"
[31] "relationships between parent or grandparent nodes and child or grandchild nodes\n"
[32] "parent of parent of 2 children"
[33] "Checks if people are related"
[34] "children or grandchildren"
[35] "Closely related or not."
[36] "connect nodes"
[37] "children in law"
[38] "find cousins, looks at the parents siblings and their children."
[39] "searches if node exist or if related"


> na.omit(p4_p_empty$AN08_01)
[1] "no"
[2] "?"
[3] "no"
[4] "-"
[5] "/"


> na.omit(p4_p_dont$AN08_01)
[1] "no idea"
[2] "do not remember, sorry, confused"


> # Problem 4: p1/2 >>> pt

> p4_pt_correct <- cknow[cknow$AN16_01 == 1,]

> p4_pt_incorrect <- cknow[cknow$AN16_01 == 0,]

> p4_pt_empty <- cknow[cknow$AN16_01 == -1,]

> p4_pt_dont <- cknow[cknow$AN16_01 == -2,]

> na.omit(p4_pt_correct$AN17_01)
 [1] "parent\nDistinguishes whether X is a parent of Y"
 [2] "parent"
 [3] "\"parent\". p(X,Y) is true, if X is the father or the mother of Y."
 [4] "defines the parent relation again"
 [5] "only true parents and childrens are allowed"
 [6] "it checks whether some node x is a parent node of node y"
 [7] "chose q1 or q2"
 [8] "X is a parent of Y"
 [9] "X is a parent of Y"
[10] "parents"
[11] "parent"
[12] "direct parent relation"
[13] "it shoes the parent nodes"
[14] "It tells you whether X is a parent of Y."
[15] "real parents"
[16] "Checks if X is father or mother of Y"
[17] "parent"
[18] "find child"
[19] "parent "
[20] "Checks for Parents"
[21] "is X father/mother of Y"
[22] "parent"
[23] "parent"
```

```
> na.omit(p4_pt_incorrect$AN17_01)
[1] "all nodes of tree"
[2] "checking direct neighbours"
[3] "Ascendents"
[4] "Relation regardless of heritage"
[5] "Female male change"
[6] "is a bidirectional move in the tree"
[7] "when p fails and there is no direct relationship: checks if x is
grandparent of Y, then if x is great-grandparent of Y (....)"
[8] "find sibling nodes"


> na.omit(p4_pt_empty$AN17_01)
[1] "no"
[2] "/"


> na.omit(p4_pt_dont$AN17_01)
[1] "don't know"
[2] "p 1/2  - did it say anywhere what that is? "
[3] "do not remember, sorry, this test is confusing *_*"
[4] "can't explain, don't know."
```

5.2 Bamberg

```
> # Problem 1: p/2

> p1_p_correct <- cknow[cknow$GF16_01 == 1,]

> p1_p_incorrect <- cknow[cknow$GF16_01 == 0,]

> p1_p_empty <- cknow[cknow$GF16_01 == -1,]

> p1_p_dont <- cknow[cknow$GF16_01 == -2,]

> na.omit(p1_p_correct$GF08_01)
 [1] "grandfather"
 [2] "grandfather"
 [3] "Grandfather, because father of Z and Z is father or mother of Y"
 [4] "Grandfather"
 [5] "grandfather"
 [6] "returns, if X is the grandfather of Y"
 [7] "X grandfather of Y"
 [8] "grandfather"
 [9] "grandfather"
[10] "GRANDFATHER"


> na.omit(p1_p_incorrect$GF08_01)
[1] "checks the sons"
[2] "grandparent"
[3] "ancestors"
[4] "siblings number"
[5] "mother/father of my father"
[6] "restricton of first problem"
[7] "grant parent on fathers side"
[8] "parents"
[9] "grandparent"
[10] "grandparent"
[11] "grandparent"


> na.omit(p1_p_dont$GF08_01)
[1] "dont know"


> # Problem 1: p1/2 >>> pt

> p1_pt_correct <- cknow[cknow$GF18_01 == 1,]

> p1_pt_incorrect <- cknow[cknow$GF18_01 == 0,]
```

```
> p1_pt_empty <- cknow[cknow$GF18_01 == -1,]

> p1_pt_dont <- cknow[cknow$GF18_01 == -2,]

> na.omit(p1_pt_correct$GF17_01)
[1] "searchs for the father or the mother"
[2] "parent"
[3] "parent"
[4] "p1 tells whether X is a parent of Y"
[5] "returns father or mother"
[6] "parent"
[7] "Checks, if X is a parent of Y"
[8] "X parent of Y"
[9] "parent"


> na.omit(p1_pt_incorrect$GF17_01)
[1] "Father or mother of the grandchild"
[2] "p1 is father, p2 is mother"
[3] "GRANDMOTHER"


> na.omit(p1_pt_dont$GF17_01)
[1] "dont know"

> # Problem 2: p/2

> p2_p_correct <- cknow[cknow$GP16_01 == 1,]

> p2_p_incorrect <- cknow[cknow$GP16_01 == 0,]

> p2_p_empty <- cknow[cknow$GP16_01 == -1,]

> p2_p_dont <- cknow[cknow$GP16_01 == -2,]

> na.omit(p2_p_correct$GP08_02)
[1] "grandparent"
[2] "Grandfather or Grandmother"
[3] "grandparents"
[4] "grandparent"
[5] "oma/opa von x"
[6] "grandparent"
[7] "grandparent"
[8] "checks, if X is a grandparent of Y"
[9] "X grandparent of Y"
[10] "grandparent"
[11] "grandparent"
[12] "returns if X is a grandfather/grandmother of Y"
[13] "grandparent"
[14] "GRANDCHILDREN"
[15] "grandparent"


> na.omit(p2_p_incorrect$GP08_02)
[1] "I think it searches if they have the same father or mother"
[2] "who is son/daugther of x and father/mother of y"
[3] "find grandparents child, who has children of its own"
[4] "children number"
[5] "same as before"
[6] "parents"


> na.omit(p2_p_dont$GP08_02)
[1] "dont know"


> # Problem 2: p1/2 >>> pt

> p2_pt_correct <- cknow[cknow$GP18_01 == 1,]

> p2_pt_incorrect <- cknow[cknow$GP18_01 == 0,]

> p2_pt_empty <- cknow[cknow$GP18_01 == -1,]
```

```
> p2_pt_dont <- cknow[cknow$GP18_01 == -2,]

> na.omit(p2_pt_correct$GP17_02)
[1] "find parent or son/daughter"
[2] "parents"
[3] "check parent"
[4] "parents"
[5] "parent"
[6] "returns if X is a parent of Y"
[7] "parent"
[8] "parent"


> na.omit(p2_pt_empty$GP17_02)
[1] "0"
> # Problem 3: p/2

> p3_p_correct <- cknow[cknow$PP16_01 == 1,]

> p3_p_incorrect <- cknow[cknow$PP16_01 == 0,]

> p3_p_empty <- cknow[cknow$PP16_01 == -1,]

> p3_p_dont <- cknow[cknow$PP16_01 == -2,]

> na.omit(p3_p_correct$PP08_01)
 [1] "grandgrandfather/mother"
 [2] "parent of the grandparent"
 [3] "grand-grandparents/grand-grandchildren"
 [4] "Great-grandfather or great-grandmother "
 [5] "X is GranGrandparent of Y"
 [6] "urgroßvater/mutter"
 [7] "urgroßeltern"
 [8] "greatGrandParent"
 [9] "Checks, if X is a great-grandparent of Y"
[10] "X greatgrandparent of Y"
[11] "grandgrandparent"
[12] "grandgrandparent"
[13] "returns if X is a grand-grandfather/grand-grandmother of Y"
[14] "grandgrandparent"


> na.omit(p3_p_incorrect$PP08_01)
[1] "The sons of X must be the parents of ( the parents of Y)"
[2] "ancestors"
[3] "ancestors number"
[4] "try all combinations"
[5] "relations between parents"
[6] "grand children"
[7] "(Urgroßvater) what das it mean in englisch, maybe grandgrandparent?"


> na.omit(p3_p_dont$PP08_01)
[1] "dont know"


> # Problem 3: p1/2 >>> pt

> p3_pt_correct <- cknow[cknow$PP18_01 == 1,]

> p3_pt_incorrect <- cknow[cknow$PP18_01 == 0,]

> p3_pt_empty <- cknow[cknow$PP18_01 == -1,]

> p3_pt_dont <- cknow[cknow$PP18_01 == -2,]

> na.omit(p3_pt_correct$PP17_01)
[1] "Tells you the father or the mother."
[2] "parent"
[3] "parent"
[4] "Father or mother"
[5] "\"X is a parent of Y\""
```

```
[6] "mutter oder vater von x"
[7] "parent"
[8] "X parent of Y"
[9] "parent"


> na.omit(p3_pt_incorrect$PP17_01)
[1] "father"
[2] "great-grandparent"
[3] "grandgrandchildren"


> na.omit(p3_pt_dont$PP17_01)
[1] "dont know"

> # Problem 4: p/2

> p4_p_correct <- cknow[cknow$AN18_01 == 1,]

> p4_p_incorrect <- cknow[cknow$AN18_01 == 0,]

> p4_p_empty <- cknow[cknow$AN18_01 == -1,]

> p4_p_dont <- cknow[cknow$AN18_01 == -2,]

> na.omit(p4_p_correct$AN08_01)
[1] "Gives you all the ancestors"
[2] "find ancestor"
[3] "Ancestor"
[4] "gibt alle Vorfahren zurück"
[5] "ancestor"
[6] "recursive check one parent"
[7] "ancestor"
[8] "Checks, if X is an ancestor of Y"
[9] "ancestor"
[10] "ancestor"
[11] "ancestor"
[12] "returns if X is an ancestor of Y"
[13] "vorfahr"
[14] "descendants"
[15] "ancestor"


> na.omit(p4_p_incorrect$AN08_01)
[1] "parent or grandparent"
[2] "parent or grandparent"
[3] "recursive"
[4] "X is parent or grandparent of Y"
[5] "parents"


> na.omit(p4_p_dont$AN08_01)
[1] "not sure"
[2] "dont know"


> # Problem 4: p1/2 >>> pt

> p4_pt_correct <- cknow[cknow$AN16_01 == 1,]

> p4_pt_incorrect <- cknow[cknow$AN16_01 == 0,]

> p4_pt_empty <- cknow[cknow$AN16_01 == -1,]

> p4_pt_dont <- cknow[cknow$AN16_01 == -2,]

> na.omit(p4_pt_correct$AN17_01)
[1] "find parent"
[2] "parents"
[3] "check if parent"
[4] "parents"
[5] "parent"
[6] "returns if X is father/mother of Y"
```

```
[7] "parent"
[8] "parent"


> na.omit(p4_pt_dont$AN17_01)
[1] "not sure"
```