

A The DJMIXED package

A.1 *Installing and using* DJMIXED

The provided software runs on SPSS versions 16 and higher (currently tested on 16, 17 and 18). The software makes use of Python framework via the ‘SPSS Programmability Extension’, which allows interleaving Python and SPSS code. The SPSS installation disks contain one convenient installer for the software that is required for this framework; the SPSS website contains copies of the relevant files if the disks are not available, see the installation guide for more details. Note that administrative rights are required to install DJMIXED, this limitation is imposed by SPSS.

The DJMIXED software itself consists of two files which need to be copied to specific locations (for SPSS 16 and 17) or of one ‘extension package’ which can be conveniently installed from within SPSS (version 18 and up). The website at djmixed.googlecode.com contains the files, example spss syntax and example output (in pdf format), and all further details.

Once installed, the packages provides a new SPSS command DJMIXED. There are a number of subcommands, only one of which can be specified at any time. Each subcommand takes one or more named options, all using the familiar `option=value` syntax.

MIXEDMODEL This command provides convenient shorthand for constructing a mixed model with random effects of words and participants. Required options are

- *dv=...* for the dependent variable,
- *items=...* to specify the variable which distinguishes items from each other (usually an item number),
- *pps=...* to specify the variable which distinguishes participants.

Optional command options include:

- *predictors=...* which takes a space-separated list of model predictors, including interactions,
- *model='...'* to give the model a name, which can be used later for model summaries and model comparison,
- *posthoc=...* to specify one variable (model term) for which post-hoc test (pair-wise comparisons) are run.
- *contrast= ..* to specify contrasts (planned comparisons). The syntax is **varname | weights1 | weights2 | ...**. The contrast list consists of two or more parts, separated by the character |. The first part is the name of the variable to run contrasts on. The following parts specify the weights to use in each contrast. For a factor with n levels, there should be $n - 1$ contrasts each containing n weights. Weights normally sum to zero and the contrasts should ideally be orthogonal.
- *output=...* Controls the amount of output and its destination. Currently three values are supported: The value **split** (default if no **output** is included) is to redirect the SPSS output to a secondary output window (djdetails) and run a **modelsummary** in the main output window.
The value **full** includes the full SPSS output in the main window, and does not run **modelsummary**. The value **none** suppresses all output.
- *plot=...* Plots to evaluate the fit of the model. Two options are currently supported: The value **residual** requests three graphs that show regression

diagnostic plots of residuals. As a side effect, the variables 'predicted', 'residual' and 'Zresidual' are added to the dataset. The value `equalvariance` requests a multiple-line histogram showing the distribution of residuals for each cell of the design. Design cells are determined by crossing all predictor variables and numbering the resulting cells. These histograms should overlap each other closely, indicating absence of heteroscedasticity.

MODELSUMMARY This command takes only one option, the name of the model to summarize (`model='...'`). The output is similar to table of results included in the paper.

COMPAREMODELS This command normally takes two options, `model1='...'` and `model2='...'`. The first model should be the simpler model, but the models will be automatically reversed if this rule is not obeyed. An overview of the properties and differences between the models is given, the LRT is run on the model specifications and the best model (assuming $\alpha = .05$ for the LRT) is indicated.

With the additional option `type=random`, this command is used for the special case of comparing two models that only differ in their random effects. See this appendix (section D) for details.

STARTMODEL This command takes one parameter, `model='...'`, which names the model that should follow this command. This command was designed for advanced users who want to use the SPSS MIXED syntax directly, but still want to make use of the convenience of COMPAREMODELS and MODELSUMMARY.

```

1  * SPSS syntax for Model 1: fixed effects and interaction .
2  MIXED rt BY priming morph
3  /FIXED= priming morph  priming*morph
4  /RANDOM=intercept | SUB(Participant) COVTYPE(VC)
5  /RANDOM=intercept | SUB(Word) COVTYPE(VC)
6  /METHOD=ML
7  /PRINT=COVB  SOLUTION  TESTCOV
8  /CRITERIA=CIN(95) MXITER(5000) MXSTEP(50) SCORING(1)
9      SINGULAR(0.000000000001) HCONVERGE(0,
10     ABSOLUTE) LCONVERGE(0, ABSOLUTE)
11     PCONVERGE(0.000001, ABSOLUTE)
12  /SAVE = PRED(predicted)  RESID(residual) .

```

Fig. B.1. Normal SPSS syntax for mixed model 1 (line numbers added)

STOPMODEL Each **STARTMODEL** command should be followed by one **MIXED** command and this **STOPMODEL** command. The one option, *model*='...' can be set to '*' as shorthand for repeating the name of the model that was most recently started.

B SPSS syntax

The **DJMIXED**/**MIXEDMODEL** command will output the SPSS syntax that it generates. To demonstrate the convenience of using **DJMIXED** and explain some details about the original SPSS **MIXED** syntax, the syntax for Model 1 is discussed here. This discussion refers to the syntax as shown in Figure B.1. In line 2, the dependent variable **RT** is introduced, together with the factors

priming and **morph**. In line 3, the **FIXED** keyword describes the normal Anova-type model, which includes the two main effects and their interaction. The intercept is automatically included in the fixed model.

The two **RANDOM** keywords on line 4 and 5 describe the random effects part of the model. Line 4 indicates that there will be participant-specific adjustments, line 5 indicates additional item-specific adjustments. Because these adjustments modify all reaction times obtained for this item (or this participant), they technically modify the model intercept. In all, line 4 is the translation of the model term u_{0p} : The u is a **RANDOM** effect, the zero index signals adjustment of the intercept and the p index implies there is one adjustment value for each participant.

For both random effects, **covtype(vc)** is specified. This related to a technical issues which governs the potential correlations between the random effects. This specification requires a variance components (VC) structure, which is closest to the assumptions of normal Anova. One may occasionally see an unstructured matrix specified here, **covtype(un)**, but this is a further departure from Anova assumptions which should be necessary in rare cases only.

On line 6, the **METHOD** line instructs SPSS to use maximum likelihood estimation to compute the mixed model. Another popular choice is restricted maximum likelihood (**METHOD=REML**), which is computationally a little more efficient and has slightly different statistical properties than ML. When REML is used for computations, model comparison with the LRT is restricted to models which have identical fixed effect structures, which is the main reason why REML is not used above.

Lines 7 to 11 contain the **PRINT** and **CRITERIA** keywords which specify com-

```

MIXED rt
  /FIXED= | SSTYPE(3)
  /RANDOM=INTERCEPT | SUBJECT(Participant) COVTYPE(VC)
  /RANDOM=INTERCEPT | SUBJECT(Word) COVTYPE(VC)
  /METHOD=ML
  /PRINT=SOLUTION TESTCOV COVB
  /CRITERIA=CIN(95) MXITER(10000) MXSTEP(50) SCORING(1) SINGULAR(0.000000000001)
  HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE) .

```

Fig. B.2. SPSS syntax for null model

putational details that do not matter in daily practice, these can safely be copied and pasted from the example.

Line 12 instructs SPSS to save the predicted and residual values in two new variables that are called ‘predicted’ and ‘residual’. The syntax files that are provided with the DJMIXED package show how plots with regression diagnostics can be created from these variables. If this syntax is run for a second time, you should first remove these two variables. If the variable names exist, SPSS will save the predicted values and residuals under other names. Recommended syntax is `DELETE VARIABLES predicted residual .`

The DJMIXED null model presented in the paper is equivalent to the following SPSS syntax:

The DJMIXED syntax for the main-effects only model is equivalent to:

For the post-hoc tests on the factor ‘form’ shown in the paper (including correction for multiple comparisons), the following line of syntax is added to

```

MIXED rt
  BY morph priming
  /FIXED= priming morph | SSTYPE(3)
  /RANDOM=INTERCEPT | SUBJECT(Participant) COVTYPE(VC)
  /RANDOM=INTERCEPT | SUBJECT(Word) COVTYPE(VC)
  /METHOD=ML
  /PRINT=SOLUTION TESTCOV COVB
  /CRITERIA=CIN(95) MXITER(10000) MXSTEP(50) SCORING(1) SINGULAR(0.0000000000001)
  HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE) .

```

Fig. B.3. SPSS syntax for main-effects only model

the model:

```

/EMMEANS = tables(form) compare adj(sidak)

```

For the contrasts on the factor ‘form’ shown in the paper, the following line is added to the model:

```

/TEST 'contrasts on form' form 0 1 -1; form 1 -0.5 -0.5

```

C Further investigations of homoscedasticity

The `PLOT equalvariance` subcommand to the `DJMIXED /MIXEDMODEL` command will show superimposed histograms (shown as lines) for each cell of the design. If these histograms are appreciably different, the distribution of residuals in each cell of the design can be further inspected with the following syntax:

```
sort cases by designcell.  
split file by designcell.  
descriptives Zresidual.  
split file off.
```

Design cells are created by crossing all predictor variables. The value of the variable ‘designcell’ is created by pasting the values of all predictors after each other. An overview of how the cells are named can be obtained by running this syntax:

```
dataset declare djmixed_legend.  
dataset activate tw.  
aggregate  
  /outfile='djmixed_legend'  
  /break=designcell  
  /morph=first(morph)  
  /priming=first(priming)  
  /Ncases=N.
```

If the design contains a factor that is nested in another factor, this definition overgenerates and empty design cells may show up. These can be safely ignored. Alternatively, the user can define their own design cells by simply assigning values to the ‘designcell’ variable. Once this is done, the syntax for the overlapped histogram plot can be copied from the output of a previous DJMIXED run.


```

1  PROC MIXED data=work.twset1 covtest method=ml noclprint ;
2      title 'Model 1';
3      class participant word morph priming ;
4      model RT = morph priming morph*priming / ddfm=sather solution ;
5      random Int / sub=participant type=vc ;
6      random Int / sub=word type=vc ;
7  run;

```

Fig. D.1. SAS syntax for Model 1 (line numbers added).

D SAS syntax

The SAS syntax for mixed models is exhaustively documented in (Littell, Milliken, Stroup, & Wolfinger, 1996). A more accessible paper (both in terms of content and because of its free availability) was written by (Singer, 1998). The syntax for Model 1 is shown in Figure refsas1. The similarities with the SPSS syntax are clear. In line 1, the dataset is declared and standard options are given to include significance tests on the random terms (also known as covariance parameters, option `covtest`), to use maximum likelihood (`method=ml`), and to suppress a listing of all levels of the categorical variables (`noclprint`).

In line 2 and 3, the model is named and the categorical variables are specified (similar to the BY clause in SPSS). Line 4 contains the fixed effect specification, followed by options to indicate use of the Satherthwaite method of computing the degrees of freedom and to include parameter estimates of the solution. Line 5 and 6 specify the random effects, adjustments to the intercept are made for groups, as defined by the variable Participant and Word, using

a Variance Components matrix (type=VC). In SAS, the values obtained for u_j can be output to a dataset using the OMS directive ‘solutionr’.

The DJMIXED package does not work on SAS.

E Covariates and random slopes

In this appendix, I will first discuss the addition of covariates to the mixed model and then discuss the addition of random slopes.

Adding a covariate is straightforward as it mixed models presented so far only have random intercepts. If this is true, a covariate can be added in the same way as it would be added to a normal regression model:

$$Y_{pi} = \beta_0 + \beta_4 C_i + u_{0p} + u_{0i} \\ + \beta_1 \text{Priming}_i + \beta_2 \text{Morph}_i + \beta_3 (\text{Priming}_i \times \text{Morph}_i) + \epsilon_{pi}.$$

Here, C_i is a covariate which is specific to each item, for example, the log frequency. If there are multiple item covariates or participant covariates, these can be added in the same way. The DJMIXED syntax for the model with log frequency as the covariate is shown in Figure E.1. If covariates are added to models with random slopes (to be discussed next), the covariate should always be added twice: To the intercept and to the slope (Nezlek, 2008, but see Raudenbush & Bryk, 2002 for exceptions).

In sum, because mixed modeling is a type of linear regression, it allows for effortless inclusion of covariates and extra factors. The mixed model distin-

```

1  * syntax for model with covariate .
2  DJMIXED /MIXEDMODEL
3          DV=rt
4          PREDICTORS = morph priming priming*morph logfreq
5          PPS=Participant
6          ITEMS=Word
7          MODEL='covariate' .

```

Fig. E.1. DJMIXED syntax for mixed model with log frequency as a covariate (line numbers added)

guishes between three levels in the data (the observation, the item and the participant) and covariates and factors at each of these levels can be included. If multiple forms of the same words are tested, as in the example data, a lemma-level property would be represented as an item-specific adjustment to the intercept or to the slope of other factors. In a further developed statistical model, this item-specific adjustment, which amounts to unexplained variance, may be reduced or even replaced by introducing covariates that may help explain the differences between items, such as cumulative frequency, abstractness, etc. Given the apparent split between items that was found in the example dataset (see figures in the paper), one would look for a dichotomous predictor at the item level which can assign items to the flat slope versus steep slope groups, or for a continuous item-level predictor on which the items happen to cluster in two groups.

Another possible extension to the random effects model presented above is that of random effects modifying slopes. Above, it was assumed that fast or

slow participants (and easy and difficult items) can be modeled with an adjustment. Because the same adjustment applied to all observations involving one participant (or one item), the adjustment is technically described as modifying the intercept. Another way to conceptualize this is that an intercept adjustment modifies the base reaction time for a participant (item). The question can be raised whether the effect of a predictor, say Priming, is the same for all items and participants or whether the effect of Priming varies by participants and by items.

We will see that four statistical complications arise when the effects of predictors (slopes) are allowed to vary by participants or by items. The limited insights gained from this further analysis and the presence of statistical complications make this an option that should be considered with care, as the much simpler analysis presented above may be sufficient in almost all cases.

In the models above, the intercept β_0 was adjusted for each participant by a random term u_{0p} and a random term u_{0i} . These adjustments imply that the base reaction time (the intercept, β_0) varies by participants and by items. Similarly, one can think of situations in which the size of the effect of Priming, β_1 , should be adjusted for every participant by introducing a random term u_{1p} . For explanatory purposes, we will not yet include the matching random term for items u_{1i} and the fixed effects of Morph and the Morph by Priming interaction will also be left out. The model formula becomes

$$Y_{pi} = \beta_0 + u_{0p} + u_{0i} + [\beta_1 \text{Priming}_i + u_{1p} \text{Priming}_i] + \epsilon_{pi}.$$

Here, the intercept is adjusted by a participant-specific and item-specific adjustment, as before. The slope or influence of Priming is the part between square brackets, which is the sum of the overall slope β_1 and a participant-

specific adjustment to that slope, u_{1p} . Individual observed reaction times differ from expected reaction times by an error term ϵ , as before.

Often, priming effects are smaller for very fast participants (presumably because they already are operating near their physical limits for unprimed words) and such participants would have a relatively large and negative value for u_{1p} . The net effect of priming for this participant is then smaller than for most participants; numerically it would be β_1 minus a relatively large adjustment. Conversely, a drowsy participant who has a hard time activating their lexical items may benefit very much from priming, leading to a large, positive u_{1p} and a comparatively large net effect of priming, $\beta_1 + u_{1p}$. Because the values of the adjustments u_{1p} are distributed normally with mean zero and variance $\sigma_{u_{1p}}^2$, the average participant has a zero adjustment and the influence of priming will be exactly β_1 for them. As before, the larger the value of a participant-specific slope adjustment, the less likely it is to occur. As before, because all adjustments have an average and expected value of zero, the *expected* reaction time is simply:

$$E(Y_{pi}) = \beta_0 + \beta_1 \text{Priming}_i.$$

Mixed Model 5 The fifth model is an extension of Model 1 introduced in the main paper. Model 5 includes a participant-specific adjustment to the slope of Priming, and all three fixed effects. Item-specific adjustments to the slope of Priming will only be introduced in Model 6, below. For ease of reading, brackets have been introduced to separate the terms that sum up to the

intercept and the slope of Priming.

$$\begin{aligned}
Y_{pi} = & [\beta_0 + u_{0p} + u_{0i}] \\
& + [\beta_1 \text{Priming}_i + u_{1p} \text{Priming}_i] \\
& + \beta_2 \text{Morph}_i + \beta_3 (\text{Priming}_i \times \text{Morph}_i) + \epsilon_{pi}.
\end{aligned}$$

Applied to the example data, the analysis yields a model with eight parameters and fit indices -2LL=26384, AIC=26400. The fixed effect of Priming was highly significant, $F(1,64)=33.2$, $p=.000$, the effect of Morph reached significance, $F(1,182)=4.74$, $p=.031$, and so did the interaction, $F(1,179)=4.90$, $p=.028$. The random effects adjusting the intercept were significant: for u_{0p} , $Z=3.46$, $p=.000$; for u_{0i} , $Z=3.95$, $p=.000$. The random effect adjusting the slope of Priming was significant too, u_{1p} , $Z=2.30$, $p=.022$, with $r(u_{0p}, u_{1p}) = -.145$, so inclusion of all random effects can be warranted.

The AIC of this model is 13 points lower than that of Model 1 (Example 1), which indicates this model fits much better. A formal comparison using the LRT of Model 1 and Model 5 yields $LRT(1) = 14.8, p = .00012$. However, this p -value is incorrect due to a technical issue, having to do with the comparison of two models with different random effects. We will discuss this issue next.

The one difference between Model 1 and Model 5 is the presence of the random factor u_{1p} in Model 5. The model formula for Model 5 displayed above can also be used for Model 1 with the only difference in the variance of u_{1p} :

$$\text{Model 1 : } s^2(u_{1p}) = 0,$$

$$\text{Model 5 : } s^2(u_{1p}) > 0.$$

Because the variance of u_{1p} is zero in Model 1, this term has effectively been removed from the model. More precisely, a random effect with a variance and a mean of zero is not really a random effect, as it is always zero. Model 1 has no participant-specific adjustment to the slope of Priming, whereas Model 5 has.

Having captured both models in one formula, the difficulty with this model comparison can be explained: Because a variance can never be smaller than zero, the parameter space (set of possible values) of s^2 is the range from zero to positive infinity. The model comparison between Model 1 and Model 5 is effectively a test for $s^2(u_{1p}) = 0$ and this test is about a boundary case in the parameter space. In the statistical literature, the test between Model 1 and 5 is said to be ‘on the boundary’. Normal formulas, including our standard likelihood-ratio test, do not hold on the boundary and Stram and Lee (1994) recommend to use the following modification to the LRT: Instead of using the normal degrees of freedom, which is the difference in the number of parameters between the models, they suggest to use a 50/50 mixture of two chi-squared distributions. The two chi-squared distributions have degrees of freedom identical to the number of *random effects* in each model. Model 1 has two and Model 5 has three random factors and the chi-squared distribution that Stram and Lee recommend is $\chi^2_{(2+3)} = 0.5\chi^2_2 + 0.5\chi^2_3$. Using this distribution, the LRT value of 14.8 obtained above is evaluated as follows

$$\begin{aligned}
P[LRT(2+3) = 14.8] &= 0.5 P[\chi^2(2) = 14.8] + 0.5 P[\chi^2(3) = 14.8] \\
&= 0.5 \cdot 0.0006112528 + 0.5 \cdot 0.00199579 \\
&= 0.001303521 \\
&\approx 0.0013,
\end{aligned}$$

Table E.1

Fixed and Random effects for Model 5

Fixed Effects					
Model term		Category	β	F	p
Priming		Unprimed	63.88	33.2	.000
Morph		Inflection	-.12	4.74	.031
Priming×Morph		Unpr+Infl	40.87	4.90	.028
Random Effects					
Model Term		Adjustment for	Variance	Z	p
u_{0p}	Intercept	Participants	5159	3.46	.001
u_{0i}	Intercept	Words	2185	3.95	.000
u_{1p}	Priming	Participants	1030	2.30	.022
ϵ	Error	—	24127	30.39	.000

which introduces the *ad hoc* notation $\text{LRT}(n+m)$ to denote a test of the LRT against a 50/50 mixture of chi-squared distributions with n and m degrees of freedom. The outcome is virtually identical to what was obtained with the uncorrected test, using the difference in number of parameters as the degree of freedom. In general, however, the normal LRT is too liberal (too optimistic) in the case of tests on the boundary and the adjusted value should be used.

The model comparison and the significance estimate for the u_{1p} parameter

reveal that participants do exhibit differing slopes of Priming, or in other words, that the effect of Priming is not the same for each participant.

A correlation of -0.145 was found between the participant-specific adjustment of the intercept u_{0p} and the participant-specific adjustment of the slope of Priming u_{1p} . This means that fast participants, who are associated with negative adjustments to the intercept, tend to show a larger Priming effect, as they have positive adjustments to the slope of Priming. The interpretation just given crucially depends on the coding of the dummy factors, which is why in Table E.1 the effect of Priming is listed as a β of 63.9 for the category of Unprimed words. In other words, primed words are coded with a zero and unprimed words are coded with a one. The intercept of the model (β_0) captures the expected reaction time for primed words and an additional 63.9 ms (β_1) has to be added for the expected reaction time of unprimed words. If the correlation was stronger and positive, faster participants (negative u_{0p}) would be strongly associated with reduced priming effects (negative u_{1p}) and it would be prudent to investigate whether fast participants show a reaction time floor effect, limiting the potential effect of priming.

When comparing the variance estimates for the three random effects, it is obvious that the newly introduced participants-specific adjustment to the effect of Priming u_{1p} is of smaller magnitude as the random effects adjusting the intercept. With a variance of 1030 and therefore $s = 32\text{ms}$, the majority of participant-specific adjustments should be between $[-2s; +2s] = [-64; +64]$.

Mixed Model 6 If the slope of Priming varies significantly between participants, the next question to ask is whether it varies significantly between items too. The mixed model analysis can easily be extended with a fourth

random effect, to yield the following full model

$$\begin{aligned}
Y_{pi} = & [\beta_0 + u_{0p} + u_{0i}] \\
& + [\beta_1 \text{Priming}_i + u_{1p} \text{Priming}_i + u_{1i} \text{Priming}_i] \\
& + \beta_2 \text{Morph}_i + \beta_3 (\text{Priming}_i \times \text{Morph}_i) + \epsilon_{pi}.
\end{aligned}$$

The top line shows the three effects that contribute to the intercept, the second line shows the three effects that contribute to the slope of Priming. The final line shows the effect of Morph, which does not have a participant or item-specific adjustments, and the interaction.

A minor complication arises with this model: So far we have used the actual word that was presented as the level of the factor Item. However, the base form of the word was always primed by the derived or the inflected form which means that Words are really nested within Priming. When estimating u_{1i} with Words as the item designator, about one-third of the data is missing for each u and SPSS cannot complete that analysis. Therefore, for model 6 the variable Base was used as the item designator, which has the same value for all three variants of the word. The theoretical claim that is implicitly made by using Base is that the locus of word-specific effects is at an abstract word representation like the lemma. Whether that claim is appropriate for this data will be not be discussed here.

Applied to the example data, the analysis yields a model with nine parameters and fit indices $-2LL = 26380$, $AIC = 26398$. The difference in AIC with the previous model is very small, 2 points. In the current model, the fixed effect of Priming was highly significant, $F(1,35)=50.1$, $p=.000$, the effect of

Morph was significant, $F(1,1915)=7.53$, $p=.006$, and so was the interaction, $F(1,1931)=7.82$, $p=.005$. Three of the four random effects were significant: For u_{0p} , $Z=3.47$, $p=.001$; for u_{0i} , $Z=2.17$, $p=.030$. The participant-specific adjustment to the slope of Priming, u_{1p} , yielded $Z=2.25$, $p=.024$, but the item-specific adjustment u_{1i} yielded $Z=1.56$, $p=.119$. Participant-specific adjustments were not strongly correlated, $r(u_{0p}, u_{1p}) = -.142$, but item-specific adjustments were, $r(u_{0i}, u_{1i}) = -.356$.

An informal model comparison with the Model 5 shows that the AIC value of this model is smaller, but only by 2 points. Although AIC is not normed, a two point difference is much less than what substantial model modification usually produce, which is in the 10 to 50 range. However, a formal comparison between this model and model 5 cannot be made because the Base factor was used in Model 6 and the Word factor in Model 5.

Only models that are *nested* can be tested with the LRT, where nested means that one model can be obtained from the other by restricting the value of one or more parameters. The null model (without any fixed predictors) is nested within Model 1 as it can be derived from Model 1 by assuming that β_1 , β_2 and β_3 are zero. As we saw above, Model 1 itself is nested within Model 5 because Model 5 allows $s^2(u_{1p})$ to be larger than zero. However, because Model 6 uses Base to index items whereas Model 1 uses Word, there is no nesting relationship between the two. For a proper comparison of the fit of this model, the analyzes should have used Base as the index for items from the start.

Even without the LRT test, it is possible to reject Model 6 as overly complicated without adding much in the way of data fit to Model 5. This decision

is based on the combination of a very small increase in AIC, a non-significant item-specific adjustment to the slope of Priming, and negligible difference in the significance and β -values for the predictors Morph, Priming and their interaction.

Multiple complications arose in the discussion of this extended example. We have only dealt with further random effects modifying the slope of Priming here, but similar model extensions can be made to allow for item- and participant-specific adjustments to the effect of Morph and to the interaction. A total of 8 random effects can theoretically be included this way.

As the number of random factors increases, estimation of each single one of them will be less robust. An assumption that mixed models make is that the same error term can be applied to all cases, regardless of condition, participant or item they were measured on: In the formulas, there is a separate value for each ϵ_{pi} (with p indexing participants, and i indexing items), but all values of the scalar ϵ are drawn from one normal distribution, $\epsilon \in \mathcal{N}(0, \sigma_\epsilon^2)$. This assumes that the errors in the observations stemming from a fast participant have the same variance as the errors produced by a slow participant; an assumption which is probably only approximately correct for most datasets. In the face of this obvious violation of the modeling assumption, extracted random slope parameters should probably be viewed with much caution as they may well reflect participant or item-specific differences in variability instead of differences in slopes (Rouder, Tuerlinckx, Speckman, Lu, & Gomez, 2008).

It should also be kept in mind that the exact structure of the random effects is rarely a psycholinguistic goal in itself. Rather, the random effects help to account for the variability between items and participants, which may otherwise

disturb our statistics. It may well be that a model incorporating participant-specific and item-specific adjustments to the intercept will suffice in many practical cases. If so, researchers can do well to adhere to this practice and only investigate a more complex random structure if the data require this (such data can be identified by the usual regression diagnostics, such as the distribution of the error term after fitting a random-intercepts model).

In Model 5, the presence of a significant random effect of participant, modifying the influence of Priming should be interpreted as evidence that priming differs between participants. However, this result does not give us any information about which participants are faster than others, nor does it explain this difference. It is often more productive to capture variation in priming effects by introducing an additional participant-level factor (such as reading ability, reading speed, fluency, aptitude on a related task) as this improves the statistical model and adds to our theoretical understanding. Random effects are designed to capture sources of noise in the data and they should not be used to mask systematic differences between items or participants.

The major drawback of exploring additional random effects with most psycholinguistic datasets is the lack of statistical ‘carrying capacity’ (Nezlek, 2008) to explore all possibilities. Although our datasets are sizable, they are usually not sufficient to do an in-depth analysis of every potential random effect without there being a real chance that some effects are masked by others. Consider again the estimation of a single observation in Model 6: This observation is said to have the predicted value (constructed from the intercept, Priming, Morph and their interaction) plus or minus an individual error (ϵ_{pi}), participant and item specific adjustments to the intercept (u_{0p}, u_{0i}), and participant and item specific adjustments to the effect of Priming (u_{1p}, u_{1i}).

This amounts to five factors that essentially model noise in the data and which are, in real-life cases, almost all confounded and correlated. Obviously, reliable estimation is difficult under these circumstances.

A more complex analysis should preferably be done on large datasets, with sizable numbers of participants and items (several 100s) and a design that is not overly contrived. In the Structural Equation Modeling literature, it is customary to validate new structural models with a second dataset to avoid introducing sweeping proposals that are due to chance (Tabachnick & Fidell, 2007). It can similarly be recommended that an analysis introducing participant-specific and item-specific adjustments to the slope of a predictor is reapplied to the data of an independent second experiment, for which a second set of model comparisons is made, to convince the reader that the original model was not a result of overfitting the data.

References

- Littell, R., Milliken, G., Stroup, W., & Wolfinger, R. (1996). *Sas system for mixed models*. Cary, NC, USA: SAS Institute Inc.
- Nezlek, J. B. (2008). An introduction to multilevel modeling for social and personality psychology. *Social and Personality Psychology Compass*, 2(2), 842-860.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. (2nd ed.). Thousand Oaks, CA, USA: Sage Publications.
- Rouder, J. N., Tuerlinckx, F., Speckman, P., Lu, J., & Gomez, P. (2008). A

hierarchical approach for fitting curves to response time measurements.

Psychonomic Bulletin & Review, 15(6), 1201–1208.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24(4), 323–355.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (fifth edition)*. Boston, USA: Pearson.