Supplementary Material
for
Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel

Dave F. Kleinschmidt* T. Florian Jaeger

October 13, 2013

1 Behavioral data analysis

1.1 Mixed-effects modeling coefficients

We analyzed the data from our Mechanical Turk experiment using mixed-effects logistic regression (Breslow & Clayton, 1993; Jaeger, 2008). We modeled the proportion of /b/ responses using fixed effects for the amount of cumulative exposure (both linear and quadratic, using orthogonal polynomials, normalized so that they have unit standard deviation), the category of the video component of the adaptor (coded as +1 for /b/ and −1 for /d/), and the ambiguity condition for auditory component of the adaptor (with coding discussed below). All interactions between fixed effects were also included (except for linear-quadratic exposures interactions). We also explored log-transformed exposures as a predictor, which did not qualitatively change the results. Below we discuss the theoretically interesting fixed effects, but the full fixed effects estimates can be found in Table 1 for the conditions replicating Vroomen, Linden, Gelder, and Bertelson (2007), and Table 2 for the fixed effects for the full design.

In all models, the maximum random effects structure justified by the data was used:[1] by-subject random intercepts and slopes for exposures (both linear and quadratic), category, and exposure-by-category interaction (both linear and quadratic). We also repeated the same analysis with using the

*Corresponding author: dkleinschmidt@bcs.rochester.edu

[1] See http://hlplab.wordpress.com/2011/06/25/more-on-random-slopes/.

full design random effects structure, which did not qualitatively change the results.

In order to maximize statistical power (especially for the comparisons between the endpoint and intermediate conditions), we analyzed the data from both the first and second blocks of the experiment together. Even though there are strong carry-over effects (Figure 3), including the block number as a fixed effect (and its interactions with variables of interest) didn't qualitatively change the results of these regression analyses, and for the sake of brevity we do not report those coefficients.

1.1.1 Replication of Vroomen et al. (2007)

Before discussing the results from the critical, intermediate conditions, it is important to verify that our web-based paradigm can replicate the results of Vroomen et al. (2007).

The results of this analysis show that our experiment replicates the three major findings of Vroomen et al. (2007). First, there was a higher rate of adaptor-category responses in ambiguous blocks, reflecting early recalibration in the ambiguous condition (reflected by a significant category-by-ambiguity interaction, $\beta = 0.31$, $\mathrm{SE}(\beta) = 0.09$, $p < .001$). Second, there was a general trend towards selective adaptation with increasing exposure (a general decrease in adaptor-category responses, reflected by a significant exposures-by-category interaction, $\beta = -0.63$, $\mathrm{SE}(\beta) = 0.06$, $p < .0001$). Third, in the ambiguous condition, there was an initial rise in recalibration followed by a later trend towards selective adaptation, which is detected in the regression analysis as a downward quadratic trend in adaptor category responses in the ambiguous condition (significant three-way interaction between quadratic exposures, category, and ambiguity, $\beta = -0.08$, $\mathrm{SE}(\beta) = 0.04$, $p < .05$). This quadratic effect was complemented by a three-way interaction of linear exposures with category and ambiguity, which shows a weaker selective adaptation trend over all in ambiguous blocks ($\beta = 0.19$, $\mathrm{SE}(\beta) = 0.06$, $p < .01$). Finally, there was also a significant intercept, reflecting a slight /d/ bias ($\beta = -0.51$, $\mathrm{SE}(\beta) = 0.12$, $p < .0001$).

1.1.2 Intermediate conditions

In order to test the belief updating model's predictions that intermediate adaptors should produce intermediate adaptation, we used an extension of the regression analysis of the ambiguous and prototypical conditions above to test this prediction qualitatively, and the precise model predictions generated based on the data from the ambiguous and prototypical conditions.

Analyzing the data from this design is difficult: we wish to investigate the gradient effect of the two intermediate ambiguity conditions, but knowing that the most and least ambiguous conditions are themselves highly distinct. Treating the ambiguity level as a continuous, numerical predictor will reveal strong effects but these are likely due to the endpoints, the ambiguous and prototypical conditions. Using Helmert coding for the ambiguity conditions both respects the ordering of the conditions by ambiguity—important for testing the continuum hypothesis—and allows each condition to be evaluated separately. The Helmert coding scheme we used contrasts intermediate-ambiguous with ambiguous, intermediate-prototypical with the average of the ambiguous and intermediate-ambiguous conditions, and prototypical with the average of the other three conditions.

As before, we used a mixed-effects logistic regression model, with ambiguity condition, video category, and linear and quadratic exposures as fixed effects and the maximum random effects structure justified by the data. The full list of fixed-effects coefficients can be found in Table 2. In the analysis of the conditions replicating Vroomen et al. (2007), stronger selective adaptation trends overall in the prototypical condition were indicated by a three-way interaction between category, (linear) exposures, and condition. Analogously, stronger selective adaptation trends were found with decreasing ambiguity (for intermediate-ambiguous, relative to fully ambiguous, $\beta = -0.10$, $\mathrm{SE}(\beta) = 0.05$, $p < .05$; for intermediate-prototypical, relative to the two more ambiguous conditions, $\beta = -0.11$, $\mathrm{SE}(\beta) = 0.03$, $p < .001$; for prototypical relative to the other three conditions, a marginal trend $\beta = -0.04$, $\mathrm{SE}(\beta) = 0.02$, $p = 0.09$). As with the stronger recalibration effect in the ambiguous vs. prototypical condition above, there were also overall fewer adaptor category responses—less recalibration—in the prototypical condition compared to the other three (category-by-prototypical interaction, $\beta = -0.13$, $\mathrm{SE}(\beta) = 0.03$, $p < .001$), with the other two conditions non-significant but trending in the expected direction. No significant quadratic effects or interactions were detected in this analysis, which is unsurprising given that these effects are only weakly detectable, if at all, when considering the maximally-distinct ambiguous and prototypical conditions above.

	Coef β	SE(β)	**z**	**p**
(Intercept)	−0.51	0.12	−4.1	**<.0001**
Exposures	−0.01	0.04	−0.2	0.8
Exposures^2	0.07	0.04	1.9	0.1
Ambiguity	0.00	0.12	0.0	1
Category=B	0.35	0.09	3.7	**<.001**
Exposures : Ambiguity	−0.04	0.04	−1.0	0.3
Exposures^2 : Ambiguity	−0.01	0.04	−0.2	0.8
Exposures : Category	−0.63	0.06	−10.5	**<.0001**
Exposures^2 : Category	0.04	0.04	1.1	0.3
Ambiguity : Category	0.31	0.09	3.4	**<.001**
Exp. : Ambig. : Cat.	0.19	0.06	3.2	**<.01**
Exp.^{2} : Ambig. : Cat.	−0.08	0.04	−2.0	**<.05**

Table 1: Mixed-effects logistic regression results of our web-based study, only including the subset of conditions replicating Vroomen et al. (2007) (ambiguous and prototypical). Table shows fixed effects coefficients given maximum random effects structure justified by the data (see also main text).

1.2 Carry-over effects between blocks

For each participant in their experiment, Vroomen et al. (2007) collected data for four repetitions each of the four conditions in their design (ambiguous and prototypical crossed with /b/ and /d/), but we fit our model to only the first block from each subject. The reason for this is that there are strong *carryover* effects from one block to the next. Many of these blocks were separated by only a short break, and so it is likely that the first few test trials will show effects not only of the current block but also previous blocks. In fact, the correlation between the first test (after 1 exposures) of block n and the last test block (256 exposures) of block $n-1$ is $r = 0.46$ (bootstrap 95% confidence interval of $[0.37, 0.54]$).[2]

Such carry-over effects might explain the presence of an initial positive aftereffect in *both* conditions (Figure 2, left). Because of the Latin-square ordering of conditions employed by Vroomen et al. (2007), any given block

[2]While part of this correlation is driven by a number of transitions where /b/ response is at floor or ceiling both before and after the transition (Figure 1), the correlation remains even after removing points where either previous or next block response was at floor or ceiling (about 35% of the data), albeit slightly weaker ($r = 0.34$, 95% CI of $[0.23, 0.44]$).

	Coef β	SE(β)	**z**	**p**
(Intercept)	−0.51	0.07	−7.0	**<.0001**
Exposures	−0.03	0.03	−1.1	0.3
Exposures^2	0.05	0.02	1.8	0.1
Intr.Ambig.	−0.07	0.10	−0.7	0.5
Intr.Proto.	0.03	0.06	0.6	0.6
Prototypical	0.02	0.05	0.3	0.8
Category=B	0.41	0.06	6.4	**<.0001**
Exp. : Intr.Ambig.	−0.03	0.04	−0.7	0.5
Exp. : Intr.Proto.	0.03	0.02	1.1	0.3
Exp. : Prototypical	0.02	0.02	1.3	0.2
Exp.^{2} : Intr.Ambig.	−0.02	0.03	−0.5	0.6
Exp.^{2} : Intr.Proto.	−0.01	0.02	−0.4	0.7
Exp.^{2} : Prototypical	0.02	0.01	1.1	0.3
Exp. : Category	−0.69	0.04	−18.4	**<.0001**
Exp.^{2} : Category	0.05	0.03	1.7	0.1
Intr.Ambig. : Cat.	−0.04	0.09	−0.5	0.6
Intr.Proto. : Cat.	−0.07	0.05	−1.3	0.2
Prototypical : Cat.	−0.13	0.03	−3.9	**<.001**
Exp. : Intr.Ambig. : Cat.	−0.10	0.05	−2.0	**<.05**
Exp. : Intr.Proto. : Cat.	−0.11	0.03	−3.6	**<.001**
Exp. : Prototypical : Cat.	−0.04	0.02	−1.7	0.1
Exp.^{2} : Intr.Ambig. : Cat.	0.01	0.04	0.3	0.8
Exp.^{2} : Intr.Proto. : Cat.	0.04	0.02	1.9	0.1
Exp.^{2} : Prototypical : Cat.	0.02	0.02	1.4	0.2

Table 2: Mixed-effects logistic regression results of our web-based study for all four conditions (see also main text).

is twice as likely to be followed by a block with a different category adaptor: there are three conditions that could follow any given condition, and two of those have a different category adaptor, and two of them have a different prototypical/ambiguous condition. The sign of the aftereffect on the current block depends on the adaptor category of the *current* block, and so if the last block had an adaptor of the other category, any carryover would be in the opposite direction. That is, the carryover influence of a negative aftereffect at the end of a /d/ block would appear as a positive aftereffect at the beginning of a following /b/ block. On average, the aftereffect at the end of a block is negative: it is essentially 0 for the ambiguous/recalibration blocks, and is strongly negative for the prototypical/selective adaptation blocks. This, combined with the fact that on two out of three blocks the adaptor category—and hence direction of the aftereffect—switches from the last block, implies that carry-over effects would result in a slight, positive aftereffect at the beginning of each block, which is precisely what is observed.

1.2.1 The role of carry-over effects in the intermediate condition data

One potential issue in analyzing the data from our web-based replication and extension of Vroomen et al. (2007) is that each participant completed two blocks, one with a /b/ adaptor and one with a /d/ adaptor. This lead to strong carry-over effects between blocks (Figure 3). In order to ensure that these carry-over effects did not affect our analysis, the belief updating model was fit to only the first block (because it assumes no prior experience besides general language background). Additional regression modeling was performed, including block as a fixed effect, interacting with all other fixed effect predictors. Controlling for block did not change the sign, magnitude, or significance of any of the effects of interest, and so this expanded regression model is not reported because of the large number of fixed effects in the expanded model.

References

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(421), 9–25.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of memory and language*, *59*(4), 434–446.

Vroomen, J., Linden, S. van, Gelder, B. de, & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia*, *45*(3), 572–7.

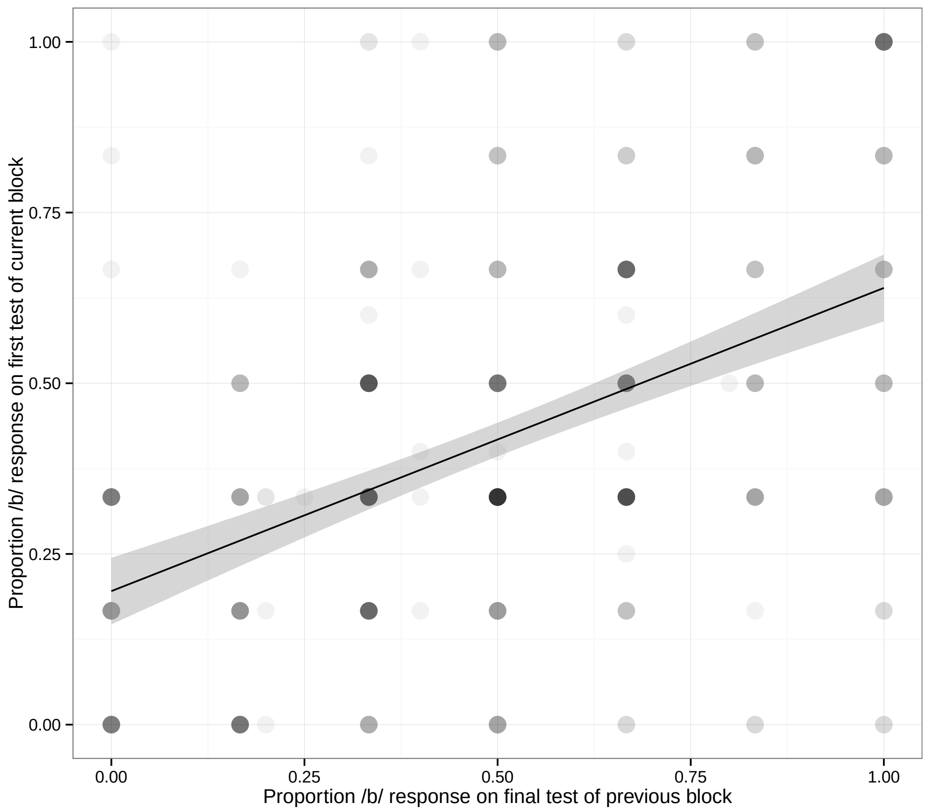

Figure 1: Carryover effects between blocks from Vroomen et al. (2007). The x-axis plots the proportion of /b/ responses at the end of block $n-1$ (last six test trials), and the y-axis shows the proportion of /b/ responses at the beginning of block n (first six test trials). Blocks occasionally had missing data in the data files, which is why not all data lines up with the 6×6 grid of the majority of the data.

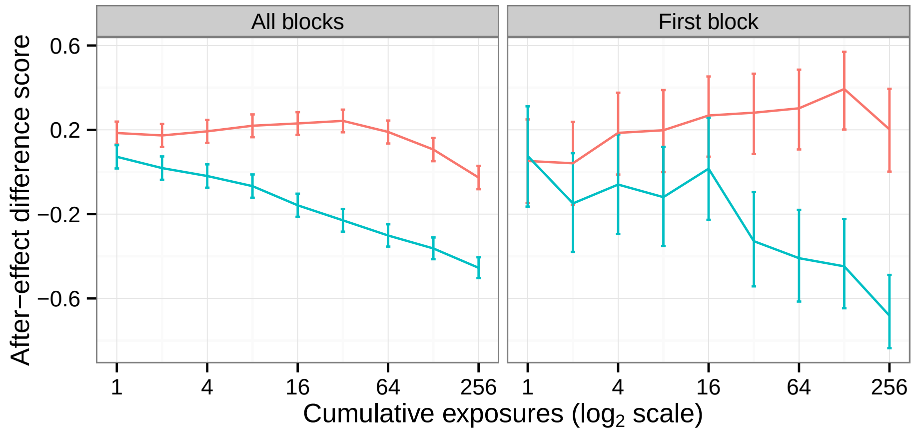

Figure 2: Aftereffect for all blocks of data from Vroomen et al. (2007), vs. just the first block from each subject. If all blocks are considered, there is an initial positive aftereffect in both the prototypical (selective adaptation; blue) and ambiguous (recalibration; red) data. This is not present if only the first block is considered.

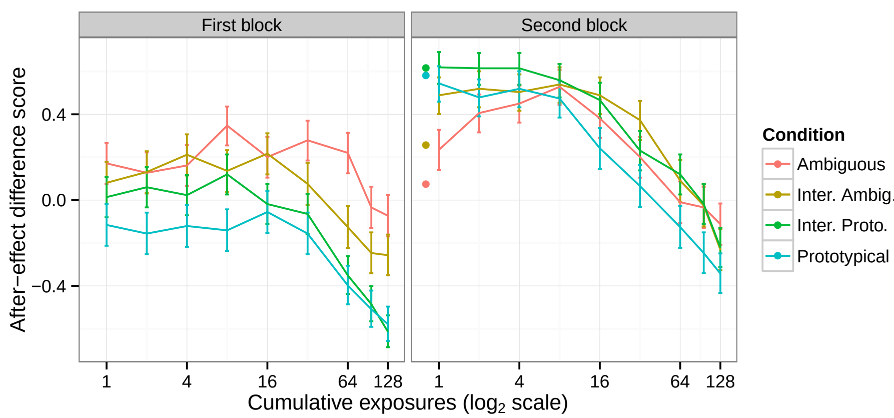

Figure 3: Strong carry-over effects between blocks in our web-based experiment.Since the two blocks seen by a participant were always in the same ambiguity condition but with opposite adaptor categories, the strong negative aftereffects at the end of the first block show up as strong positive aftereffects at the beginning of the second block (indicated by points in second panel).