Is There Phonological Feature Priming?

Supplementary Materials

Karthik Durvasula*

Michigan State University

durvasul@msu.edu

Alicia Parrish*

New York University

alicia.v.parrish@nyu.edu

November 7, 2018

1 Experiment 1: Extra Information

1.1 Stimuli

Word frequencies were obtained from SubtlexUS (Brysbaert and New, 2009). The mean $\log_e$ frequencies for experimental items did not differ significantly between the Related and Unrelated conditions for any of the POA conditions for either primes or targets (Table 1).

POA	Prime/Target	Related	Unrelated	Difference	Welch test
Bilabial	Prime	2.63	2.76	-0.13	[t(23)=-0.3, p=0.77]
Bilabial	Target	2.50	2.80	-0.30	[t(23)=-0.87, p=0.39]
Alveolar	Prime	2.86	2.71	0.15	[t(20.4)=0.35, p=0.73]
Alveolar	Target	3.13	3.11	0.02	[t(20.0)=0.07, p=0.95]
Velar	Prime	2.59	3.18	-0.59	[t(19.9)=-1.35, p=0.19]
Velar	Target	3.22	3.07	0.15	[t(18.6)=0.39, p=0.70]

Table 1: Mean $\log_e$ frequencies of the prime/target words for different POA.

1.2 Model selection procedure

We attempted to obtain the maximal random-effects structure that was justified by the data (Barr, Levy, Scheepers, and Tily, 2013). However, as is typical in psycholinguistic data (and in our own experience), the models with the most complex random-effects structures did not converge. It is important to note that the field and the statistical literature in general have not come to a consensus on how to best proceed in identifying the best random-effects structure, especially when a model with a particular random-effects structure does not converge (Bolker, 2014). In what follows, we describe the random-effects structure selection process that we used for our experiments by following other experienced linear mixed-effects modelers in psycholinguistics (Barr et al., 2013). We identified the appropriate random-effects structure by keeping the fixed-effects constant - we used the full fixed-effects model for the experiment (i.e., with interactions for all the fixed-effects). We started with the most complex random-effects structure. In the case of nonconvergence of the complex random-effects model, we systematically pared down the random-effects structure, by first removing varying slopes related to interaction terms and then removing those related to non-interaction terms, until convergence was reached. The least complex random-effects structure we entertained was one with a varying intercept for both subjects and items. When convergence was reached, the corresponding random-effects model was identified to be the maximal random-effects structure justified by the data.

Model comparison to identify the best combination of fixed-effects was performed through backwards elimination of nonsignificant terms, beginning with the interactions, through a chi-squared test of the log likelihood ratios. The most complex fixed-effects model entertained was the full model with all interaction terms, and the least complex model entertained was the model with only an intercept term and no fixed-effects. It is important to note that it is not appropriate to use likelihood tests for non-nested models; however, there is no simple way to present the actual comparisons made, so we chose to present it in the order of ascending degrees of freedoms for the models. For example, in Table 2, there is no p-value associated with the 1+POA row; the pairwise comparison with the previous model (namely, fixed-effects = 1 + TargetType) does not constitute a comparison of nested models, and is therefore left blank.

In Table 2, we present the pairwise comparisons between nested models comparing the model on the line that has particular p-value with the model in the previous line. As can be seen, the best model based on a chi-squared test of the log likelihood ratios is the model with just an intercept in the fixed-effects structure, since there was no significant improvement of the model when the various factors were added. Furthermore, the lowest BIC value is also the model with just an intercept in the fixed-effects structure. This suggests that none of the differences between the Unrelated and Related words were significant. This in turn suggests that there are no clear priming effects.

Fixed-effects	BIC	χ^2	df	Pr(> χ^2)
Intercept	-24			
1 + TargetType	-19.6	3.0	1	0.08
1 + POA	-14.8	2.7	1	
1 + POA + TargetType	-10.3	3.0	1	0.08
1 + POA + TargetType + POA*TargetType	3.0	1.7	2	0.42

Table 2: Comparison of linear mixed-effects models for Experiment 1. P-values correspond to legitimate pairwise comparisons between nested models comparing the model on the line that has particular p-value with the model in the previous line.

Finally, in response to a question from an anonymous reviewer, we also did the same analyses with the raw reaction time values as the dependent variable. Again, the best model based on a chi-squared test of the log likelihood ratios is one with just an intercept in the fixed-effects structure (Table 3), since there was no significant improvement of the model when the various factors were added. Furthermore, this model also had the lowest BIC value. And the model had a Bayes Factor of 12.1 compared to the next best model, again suggesting that it was indeed the best model for the data.

Fixed-effects	BIC	χ^2	df	Pr(> χ^2)
Intercept	24805			
1 + TargetType	24810	2.57	1	0.11
1 + POA	24815	1.88	1	
1 + POA + TargetType	24820	2.58	1	0.11
1 + POA + TargetType + POA*TargetType	24834	1.86	2	0.39

Table 3: Comparison of linear mixed-effects models for Experiment 1 with raw reaction times as the dependent measure. P-values correspond to legitimate pairwise comparisons between nested models comparing the model on the line that has particular p-value with the model in the previous line.

Although the analyses with both the $\log_e$ RTs and the raw RTs showed the same pattern of support for the model with just an intercept in the fixed-effects structure, we suggest that the results from the analysis with $\log_e$ RTs are more reliable, as the residuals of the final model were more normally distributed.

1.3 Normality testing

While standard tests for checking the normality of residuals allow for easy rejection of the null hypothesis (that the residuals are drawn from a population that is normally distributed), we can use graphical checks to see if the $\log_e$ transform results in residuals that are closer to being normally distributed than the raw values (Anglim, 2016). In Table 1, we plotted the standardized residuals of the best models based on both the $\log_e$ transformed values and the raw values. The black line represents the abstract standard normal distribution. As can be seen, the residuals for the best model based on the $\log_e$ transformed values are much closer to being normal than the residuals for the best model based on raw reaction time values.

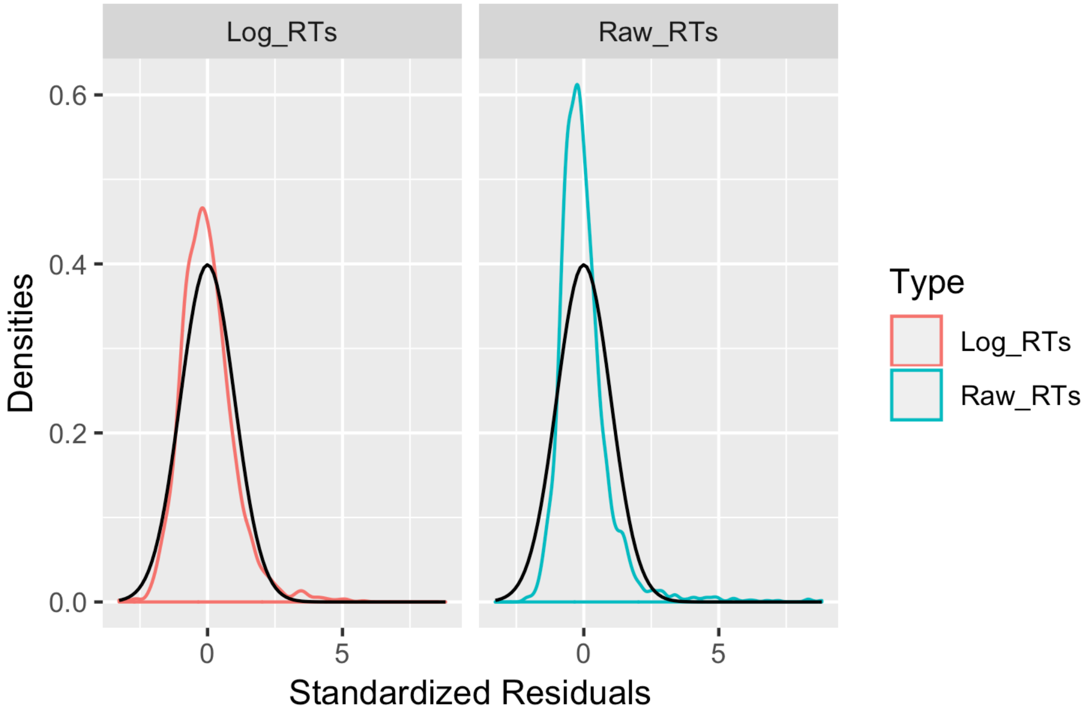

Figure 1: Comparison of standardized residuals of the raw reaction time values and the $\log_e$ transformed values with the standard normal distribution for Experiment 1.

Stimuli for Experiment 1

Velar stimuli

Related		Unrelated	
Prime	**Target**	**Non-Prime**	**Target**
game	came	mart	cart
gore	core	joy	coy
call	gall	forge	gorge
curl	girl	chide	guide
cap	gap	lit	kit
coat	goat	lord	gourd
cold	gold	pang	gang
gave	cave	dill	kill
gab	cab	base	case
come	gum	ding	king
gain	cane	van	can
gable	cable	rule	cool

Alveolar stimuli

Related		Unrelated	
Prime	**Target**	**Non-Prime**	**Target**
dart	tart	gape	tape
sip	zip	coo	zoo
tore	door	park	dark
zoom	doom	hair	dare
sigh	tie	room	tomb
toon	soon	gauge	sage
sink	zinc	burn	turn
sent	tent	vote	tote
zing	sing	vole	soul
time	dime	fog	dog
den	zen	cone	zone
dummy	tummy	more	tour

Bilabial stimuli

Related		Unrelated	
Prime	**Target**	**Non-Prime**	**Target**
pat	bat	tang	bang
ball	mall	leak	peek
bale	pale	rink	pink
pill	bill	cord	bored
peat	beet	torn	born
mean	bean	node	bode
pore	bore	tell	bell
bear	pear	dent	pent
bad	mad	job	mob
bun	pun	runt	punt
balm	palm	zest	pest
patch	batch	shark	bark
pen	men		

2 Experiment 2: Extra Information

2.1 Stimuli

Again, the mean $\log_e$ frequencies for target words did not differ significantly between the Related and Unrelated conditions for any of the POAs for either primes or targets (Table 4).

POA	Prime/Target	Related	Unrelated	Difference	Welch test
Bilabial	Prime	2.90	3.01	-0.11	[$t(22)=-0.3$, $p=0.77$]
Bilabial	Target	2.86	2.84	0.02	[$t(11.6)=0.08$, $p=0.94$]
Alveolar	Prime	2.80	2.64	0.16	[$t(19.3)=0.46$, $p=0.65$]
Alveolar	Target	2.97	2.97	0	[$t(20.8)=0$, $p=1$]
Velar	Prime	3.04	2.94	0.1	[$t(22.2)=0.2$, $p=0.79$]
Velar	Target	2.92	2.88	0.04	[$t(23)=0.1$, $p=0.91$]

Table 4: Mean $\log_e$ frequencies of the prime/target words for different POAs.

The neighborhood densities for pairs of words were controlled for as much as possible. The $\log_{10}$ Kucera-Francis (KF) frequency weighted densities were obtained from the Irvine Phonotactic Online Dictionary v. 1.4 (Vaden, Halpin, and Hickok, 2009). These values did not differ significantly between the Related and Unrelated conditions for each POA for either primes or targets (Table 5). It is worth pointing out that the difference in values for the Velar primes and targets are marginally significant, suggesting that there is the possibility of a sufficient departure from the null hypothesis expectation of "no difference" for each comparison; however, foreshadowing the results somewhat, this does not seem to have affected the pattern of results, as ultimately we did not find any effect of priming for any of the POA conditions.

POA	Prime/Target	Related	Unrelated	Difference	Welch test
Bilabial	Prime	29.2	24.3	4.9	[t(22.8)=-0.72, p=0.48]
Bilabial	Target	34.0	27.3	6.7	[t(20.9)=1.00, p=0.33]
Alveolar	Prime	28.3	30.8	-2.5	[t(18.7)=-0.36, p=0.72]
Alveolar	Target	23.1	29.7	-6.6	[t(17.7)=-1.00, p=0.33]
Velar	Prime	39.0	28.9	10.1	[t(22.9)=2.05, p=0.06]
Velar	Target	33.0	24.4	8.6	[t(22.9)=2.03, p=0.06]

Table 5: Neighborhood densities ($\log_{10}$ Kucera-Francis (KF) frequency weighted densities) of the prime/target words for different POA.

2.2 Model selection procedure

As can be seen in Table 6, the best model, based on a chi-squared test of the log likelihood ratios, is one with just an intercept and no other fixed-effects, since there was no significant improvement of the model when the various factors were added.[1]

Fixed-effects	BIC	χ^2	df	Pr(> χ^2)
Intercept	113.2			
1 + TargetType	120.9	0.0002	1	0.99
1 + POA	128.2	0.38	1	
1 + POA + TargetType	135.9	0.002	1	0.96
1 + POA + TargetType + POA*TargetType	150.2	1.07	2	0.59

Table 6: Comparison of linear mixed-effects models for Experiment 2. P-values correspond to legitimate pairwise comparisons between nested models comparing the model on the line that has particular p-value with the model in the previous line.

---

[1] Similar to the results for Experiment 1, in Table 6, there is no p-value associated with the 1+POA row; the pairwise comparison with the previous model (namely, fixed-effects = 1 + TargetType) does not constitute a comparison of nested models, and is therefore left blank.

Finally, we also did the same analyses with the raw reaction time values as the dependent variable. Again, the best model based on a chi-squared test of the log likelihood ratios and the lowest BIC value is one with just an intercept in the fixed-effects structure (Table 7). And the model had a Bayes Factor of 42.5 compared to the next best model, again suggesting strong evidence in favor of the simplest model (Raftery, 1995).

Fixed-effects	BIC	χ^2	df	Pr(> χ^2)
Intercept	92.3			
1 + TargetType	99.8	0.2	1	0.65
1 + POA	106.9	0.53	1	
1 + POA + TargetType	114.4	0.24	1	0.63
1 + POA + TargetType + POA*TargetType	128.5	1.34	2	0.51

Table 7: Comparison of linear mixed-effects models for Experiment 2 with raw reaction times as the dependent measure. P-values correspond to legitimate pairwise comparisons between nested models comparing the model on the line that has particular p-value with the model in the previous line.

2.3 Normality testing

As with Experiment 1, we can use informal visual checks to see if the $\log_e$ transform result in residuals that are more normally distributed than with the raw values (Anglim, 2016). In Table 2, we plotted the standardized residuals of the best models based on both the $\log_e$ transformed values and the raw values. The black line represents the abstract standard normal distribution. As can be seen, the residuals for the best model based on the $\log_e$ transformed values are much closer to being normal than the residuals for the best model based on raw reaction time values. This suggests, as in Experiment 1, that the results for the analysis with $\log_e$ transformed reaction times are more trustworthy.

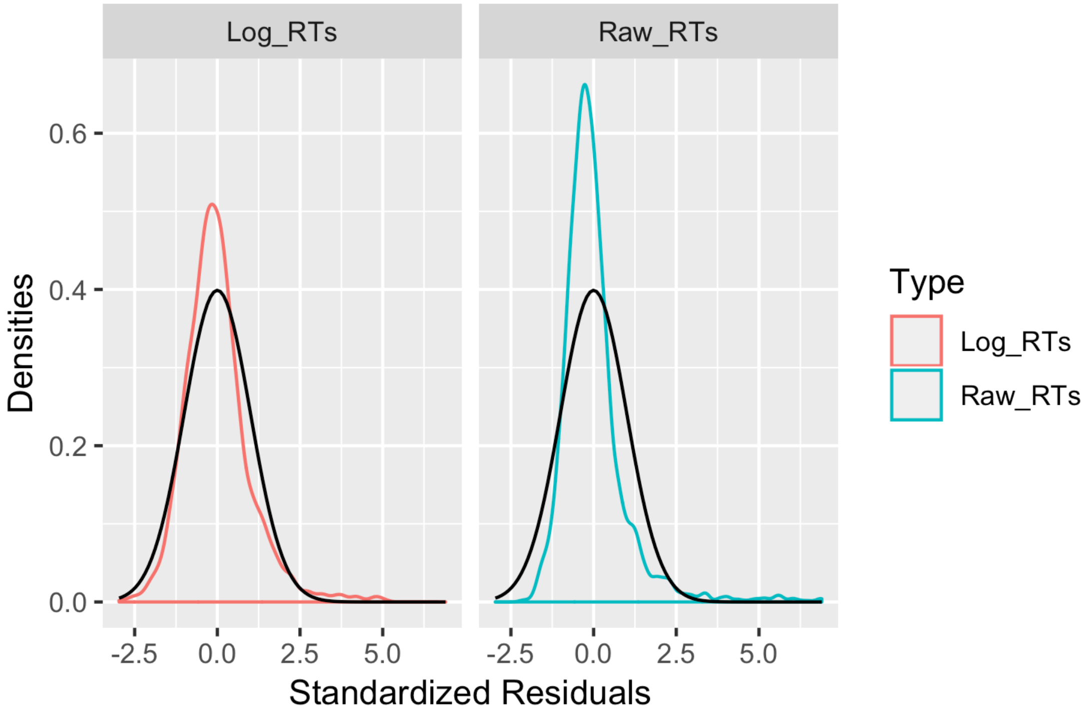

Figure 2: Comparison of standardized residuals of the raw reaction time values and the $\log_e$ transformed values with the standard normal distribution for Experiment 2.

Stimuli for Experiment 2

Velar stimuli

Related		Unrelated	
Prime	**Target**	**Non-Prime**	**Target**
code	goad	pang	gang
call	gall	bit	kit
curl	girl	till	gill
cap	gap	pun	gun
cane	gain	tape	gape
coat	goat	duel	cool
cold	gold	tide	guide
game	came	base	case
gore	core	bite	kite
gave	cave	bowl	coal
gab	cab	boy	coy
gage	cage	pate	gate
		ding	king

Alveolar stimuli

Related		Unrelated	
Prime	**Target**	**Non-Prime**	**Target**
time	dime	cog	dog
tuck	duck	beer	tear
tip	dip	pier	dear
tore	door	cot	dot
tomb	doom	cope	dope
tame	dame	curt	dirt
die	tie	gown	town
dart	tart	goal	toll
dank	tank	bang	tang
dent	tent	boat	tote
den	ten	gag	tag
dean	teen	bone	tone

Bilabial stimuli

Related		Unrelated	
Prime	**Target**	**Non-Prime**	**Target**
park	bark	toss	boss
path	bath	toil	boil
pelt	belt	ton	bun
pour	bore	tell	bell
patch	batch	cord	board
pill	bill	torn	born
peg	beg	guest	pest
ban	pan	gale	pale
bin	pin	dine	pine
bear	pair	guy	pie
bad	pad	geek	peak
box	pox	deep	peep
balm	palm		

References

Anglim, J. (2016). *Is normality testing 'essentially useless'?* Cross validated (StackExchange). Retrieved from `https://stats.stackexchange.com/questions/2492/is-normality-testing-essentially-useless`

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 1–26. doi: 10.1016/j.jml.2012.11.001

Bolker, B. M. (2014). *Random- and fixed-effects structure in linear-mixed models.* Cross validated (StackExchange). Retrieved from `http://stats.stackexchange.com/questions/130714/random-and-fixed-effects-structure-in-linear-mixed-models`

Brysbaert, M., and New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977-990. Retrieved from `http://subtlexus.lexique.org/moteur2/index.php` (Online; accessed October-2014)

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. Retrieved from `http://www.jstor.org/stable/271063`

Vaden, K., Halpin, H., and Hickok, G. (2009). *Irvine phonotactic online dictionary, version 1.4. [data file].* Available from http://www.iphod.com.