Supplementary Materials

Model Tables for Analysis of Full Data Set (*N*= 34)

PM Task

Table 1. Model results for the binomial probit models of response accuracy, and the mixed linear models of correct RT for both the lexical decision and PM tasks.

Effect	df	Accuracy		Response Time	
		χ^2	*p*	χ^2	*p*
PM	1	466.72	<.001	71.68	<.001
D	1	159.23	<.001	9.11	.01
PM x D	1	4.71	.03	0.22	.90

Lexical Decision Task

Effect	df	Accuracy		Response Time	
		χ^2	*p*	χ^2	*p*
S	1	134.99	<.001	0.05	.82
PM	2	5.00	.08	38.13	<.001
D	2	215.80	< .001	188.24	<.001
S x PM	2	17.09	<.001	12.94	<.001
S x D	2	5.44	.07	0.34	0.56
PM x D	4	4.64	.33	0.65	0.42
S x PM x D	4	12.26	.02	0.015	0.90

Simple effects analysis of the model data set (*N*=30)

Prospective Memory Task

Accuracy. PM responses were scored as correct if the participant pressed the PM response key instead of a lexical decision response key on the target trial. A condition by day repeated measures ANOVA was conducted on PM accuracy. PM accuracy was higher in the focal condition (*M* = 82%, *SE* = 3%), than the non-focal condition (*M* = 59%, *SE* = 3%), *F* (1, 29) =

104.67, $p < .001$, $\eta_p^2 = .78$, and there was an effect of day (day 1 $M = 77\%$, $SE = 4\%$; day 2 $M =$ 71%, $SE = 3\%$; day 3 $M = 63\%$, $SE = 4\%$), F (2, 58) = 18.34, $\varepsilon = .88$, $p < .001$, $\eta_p^2 = .39$. Condition and day did not interact, $F < 1$. The PM false alarm rate was 0.17% (ranging from 0 to 0.5% across participants).

Response Times. A condition by day repeated measures ANOVA was conducted on mean RTs to correct PM trials. Two participants had a block of trials with no correct PM responses, and thus were excluded from the ANOVA. Correct PM responses were faster in the focal condition ($M = 0.808$s, $SE = 0.023$s) than the non-focal condition ($M = 0.930$s, $SE = 0.026$s), F (1, 27) = 87.04, $p < .001$, $\eta_p^2 = .76$. The effect of day was not significant, F (2, 54) = 2.95, $\varepsilon =$ .73, $p = 0.08$, and condition and day did not interact, $F < 1$.

Lexical Decision Task

Accuracy. PM target trials, false alarms and the two lexical decision trials following each target trial or false alarm were excluded. Trials in which participants were presented repeated non-targets in the control condition and the two trials following were also excluded. A stimulus type by condition by day repeated measures ANOVA was conducted on mean lexical decision accuracy. There was an effect of day (day 1 $M = 94.0\%$, $SE = 0.93\%$; day 2 $M = 92.5\%$, SE =1.08%; day 3 $M = 91.6\%$, $SE = 1.15\%$), F (2, 58) = 8.76 , $\varepsilon = .88$, $p < .001$, $\eta_p^2 = .23$, and of condition, F (2, 58) = 5.40, $\varepsilon = .89$, $p < .01$, $\eta_p^2 = .16$ (non-focal = 93.2%, focal = 92.6%, control = 92.3%). There was an interaction between stimulus type and condition, F (2, 58) = 4.47, $\varepsilon =$ .91, $p < .05$, $\eta_p^2 = .13$ (see Figure 3). Planned comparisons revealed a small difference between both non-focal non-word accuracy (94.0%) compared to control non-word accuracy (92.7%), t (29) = 2.38, $p = .02$, $d = 0.43$, and for focal non-word accuracy (93.9%) compared to control non-word accuracy, t (29) = 2.61, $p = .01$, $d = 0.48$. Non-focal word accuracy (92.3%) was not

significantly different to control word accuracy (91.9%), t (29) = 1.12, p = .27, but there was a small difference between focal word accuracy (91.3%) and control word accuracy, t (29) = 2.27, p = .03, d = 0.41. No other effects or interactions were significant (smallest p = .19, ε = .95).

Response Times. A stimulus type by condition by day repeated measures ANOVA was conducted on mean RTs to correct lexical decision trials (Figure 1). There was an effect of day (day 1 M = 0.704s, SE = 0.016s; day 2 M = 0.657s, SE = 0.011s; day 3 M = 0.630s, SE = 0.011s), F (2, 58) = 24.44, ε = 0.71, p <.001, η_p^2= .46, which interacted with stimulus type, F (2, 58) = 5.32, ε = 0.91, p <.01, η_p^2 = .15. There was an effect of condition, F (2, 58) = 21.63, ε = 0.75, p < 0.001, η_p^2= .43, with non-focal RTs (M = 0.686s) longer than focal RTs (M = 0.658s) and control RTs (M = 0.648s). Condition interacted with stimulus type, F (2, 58) = 33.27, ε = 0.78, p < 0.001, η_p^2= .53. Word RTs were slower in non-focal PM blocks (0.696s) compared to control blocks (0.638s), t (29) = 9.28, p < 0.001, d = 1.69, and in focal PM blocks (0.660s) compared to control blocks, t (29) = 3.27, p < 0.01, d = 0.60, whereas non-word RTs were slower for non-focal blocks (0.675s) compared to control blocks (0.658s), t (29) = 2.26, p = 0.03, d = 0.41, but were not significantly different in control blocks compared to focal blocks (0.656s), t < 1. The non-focal PM slowing was larger for correct word responses than correct non-word responses, t (29) = 6.55, p <.001, d = 1.20. No other effects or interactions in the ANOVA reached significance (smallest p = .43, ε = .62).

BIC Selected LBA Analysis

The average fit of the BIC-selected LBA to the data is displayed in Figure 1. The most notable difference between the AIC-selected LBA and the BIC-selected LBA is that the BIC-selected LBA does not allow flexibility of thresholds in the interaction between PM condition and response type. BIC-selected parameter analysis is below.

For estimates of B there was a main effect of condition, F (2, 58) = 32.49, ε = .85, $p < .001$, η_p^2 = .53, (control M = 0.963, SE = 0.02; focal M = 0.989, SE = 0.02; non-focal M = 1.039, SE = 0.02). B was higher than control in the focal condition, t (29) = 2.63, $p < .05$, d = 0.48, and in the non-focal condition, t (29) = 6.90, $p < .001$, d = 1.26. Estimates of B decreased over days (day 1 M = 1.066, SE = 0.023; day 2 M = 0.987, SE = 0.015; day 3 M = 0.938, SE = 0.017), F (2, 58) = 23.86, ε = .75, $p < .001$, η_p^2 = .45. There was no interaction between day and condition ($F < 1$).

For v, there was a main effect of correspondence, F (1, 29) = 157.24, $p < .001$, η_p^2 = .84, with correct drift rates (M = 1.904, SE = 0.145) higher than incorrect drift rates (M = -0.672, SE = 0.184). The effect of stimulus type and its interaction with correspondence were not significant (Fs < 1). For sv, there was a significant effect of correspondence, with variability in rates lower for correct responses (M = 0.415, SE = 0.014) than for incorrect responses (which were fixed at 1), F (1, 29) = 847.12, $p < .001$, η_p^2 = .97.

For $t0$, there was no effect of stimulus type, F (2, 58) = 1.90, p = .18 (word M = 0.079, non-word M = 0.073).

DDM Analysis

As discussed in the manuscript, we favour the LBA because the DDM was a statistically inferior fit in terms of both AIC and LBA, and seemed to under-predict error rates to word responding (inadequate attribution to threshold). However, one benefit of the DDM was its slightly better fit to the trends in correct RTs than the LBA (Figure 3). Also note that the AIC-selected LBA over-predicted error rates for word responding in the non-focal condition, whereas the DDM under-predicted the error rates.

The DDM (Figure 2) characterizes the performance of two-choice tasks as a noisy evidence accumulation process which begins at a starting point z and accumulates over time until response

A or B is reached (Boywitt & Rummel, 2012; Horn et al., 2011; Ratcliff et al., 2004). The tendency for evidence to accumulate towards response A or B and the speed of evidence accumulation are represented by the drift rate, which varies around the mean v and standard deviation η. Drift rates depend on stimulus features as well as the cognitive capacity devoted to stimulus processing. The threshold of evidence participants require to perform responses (criterion setting) is represented by the parameter a. A higher criterion setting indicates more cautious responding; response speed being traded for response accuracy. The starting point z, which is rectangularly distributed with range SZ, indicates the bias to make response A versus B. We report response bias in terms of Z (z/a). Higher values of Z indicate a bias towards making word responses (less evidence required); lower values indicate a bias against word responding. The DDM also includes time before and after information processing (encoding and motor response); non-decision time parameter *Ter*. The DDM assumes non-decision time is drawn from a uniform distribution.

As with our LBA analysis, we sequentially fit more and more complex DDMs, building to a most complex "top" model, which we allowed to be highly flexible. The top model included some reasonable restrictions to reduce computational effort. As a and Z are assumed to be set prior to onset of the stimulus, they were fixed over stimulus type. In order to test the delay theory, a and Z were free to vary over condition. To capture practice effects, a and Z were free to vary over days. SZ was fixed over all factors. In order to test the capacity sharing theories, we allowed mean drift rate v to vary over all possible factors; stimulus type, condition and day. We fixed variability in drift rates η over all factors. Non-decision time *Ter* was free to vary over condition and day, but variability in non-decision time was fixed. Upon the suggestion of a reviewer, our LBA analysis allowed non-decision time to vary over stimulus type. We attempted

the same flexibility with the DDM, but were not able to achieve it as the resulting model tree was too computationally expensive to fit (requiring around a year of fitting time on a 64-core server computer). We had the same issue attempting to include the variability in drift rates in our DDM models, and so had to rely on the LBA to test the attention lapse view. With our chosen top DDM, our sequential model fitting resulted in 11,265 model fits per subject, for a total of 337,950 fits. This was the same number of fits required for our LBA model tree, but each fit too considerably longer than the LBA. Table 1 shows our model selection results. Similar to the LBA, the AIC selected DDM retained full flexibility for *a* and *Z,* but dropped the condition factor for *v.* In contrast to the LBA, the AIC selected DDM retained full flexibility for *Ter.* The BIC selected model was simpler, retaining only the day factor for *a* and *Z*, retaining no factors for *v*, but again retaining full flexibility for *Ter.* The AIC model did not fit significantly worse than the top model, $\chi^2(360) = 395$, $p = 0.10$, but the BIC model did fit significantly worse than the top model, $\chi^2(870) = 3117$, $p < 0.001$. The BIC model was too inflexible to predict differences in response accuracy, whereas the AIC model provided a similar account to the top model of trends in accuracy and the RT distributions (Figure 3). Thus, we focus interpretation on the parameters of the AIC-selected model.

AIC Selected DDM Parameter Analysis

For *a*, a PM condition by day ANOVA indicated a main effect of PM condition, $F (2, 58) = 4.33$, $\varepsilon = 0.87$, $p < .05$, $\eta_p^2 = .13$, with no difference between control ($M = 0.124$, $SE = 0.005$) and focal conditions ($M = 0.124$, $SE = 0.004$), $t < 1$, but a significant difference between control conditions and non-focal conditions ($M = 0.131$, $SE = 0.005$), $t (29) = 2.51$, $p < .05$, $d = 0.46$. There was also a main effect of day, $F (2, 58) = 15.36$, $\varepsilon = 0.72$, $p < 0.001$, $\eta_p^2 = .35$ (day one $M =$

0.139, $SE = 0.006$; day two $M = 0.123$, $SE = 0.003$; day 3 $M = 0.117$, $SE = 0.003$). The interaction between day and PM was not significant, $F < 1$.

For Z, a condition by day ANOVA indicated a main effect of day, $F (2, 58) = 4.88$, $\varepsilon = .92$, $p < .05$ (day one $M = 0.489$, $SE = 0.010$; day two $M = 0.470$, $SE = 0.007$; day 3 $M = 0.464$, $SE = 0.008$). There was also a main effect of condition, $F (2, 58) = 16.67$, $\varepsilon = .88$, $p < .001$, $\eta_p^2 = .36$, the thresholds for word versus non-word responses similar under control conditions ($M = 0.492$, $SE = 0.009$), but relatively in favour of non-word responding under focal conditions ($M = 0.472$, $SE = 0.007$), $t (29) = 4.29$, $p < .001$, $d = 0.78$, and non-focal conditions ($M = 0.459$, $SE = 0.009$), $t (29) = 5.04$, $p < .001$, $d = 0.92$. We evaluate evidence for a global threshold increase in thresholds by separately examining evidence required for each response (word $= z$, non-word $=$ 1-z). Evidence required to respond word was significantly higher under non-focal conditions than control conditions, $t (29) = 5.94$, $p < 0.001$, $d = 1.09$, but not significant higher under focal conditions than control conditions, $t (29) = 1.67$, $p = 0.11$, $d = 0.30$. The evidence to respond non-word did not significantly increase under non-focal PM conditions, $t < 1$, or focal PM conditions, $t (29) = 1.45$, $p = 0.16$. Thus, the DDM threshold shifts caused by PM instructions were purely selective. The interaction between day and PM was not significant, $F = 1.97$, $\varepsilon = .65$, $p = .13$.

For v, a stimulus type by day ANOVA indicated a main effect of stimulus type, $F (1, 29) = 5.25$, $p < .05$, $\eta_p^2 = .15$, with word rates ($M = 0.281$, $SE = 0.007$) higher than non-word rates ($M = 0.269$, $SE = 0.007$), but no other effects or interactions (smallest $p = .85$).

For Ter, a PM by day ANOVA revealed a main effect of PM, $F (2, 58) = 21.39$, $\varepsilon = .79$, $p < 0.001$, $\eta_p^2 = .42$, with Ter lower in the control condition ($M = 0.354$, $SE = 0.006$) than in the focal condition ($M = 0.364$, $SE = 0.007$), $t (29) = 4.39$, $p < 0.001$, $d = 0.80$, and in the non-focal

condition ($M = 0.376$, $SE = 0.008$), t (29) = 6.03, $p < 0.001$, $d = 1.10$. There was also a main

effect of day, F (2, 58) = 5.82, $\varepsilon = .86$, $p < .01$, $\eta_p^2 = .17$, with *Ter* decreasing over days (day 1 M

= 0.377, $SE = 0.008$; day 2 $M = 0.364$, $SE = 0.006$; day 3 $M = 0.352$, $SE = 0.007$). The PM and

day effects did not interact, $F = 1.03$, $\varepsilon = .62$, $p = .38$.

BIC Selected DDM Analysis

For similar reasons as with our LBA modelling (failures to capture trends with the BIC

model, significant increase in deviance compared with the top model), we preferred the AIC-

selected DDM over the BIC-selected model. The fit of the BIC-selected DDM to the data

averaged across subjects is displayed in Figure 4.

A one way repeated measures ANOVA revealed that *a* decreased on later days, F (2, 58) =

12.50, $\varepsilon = .65$, $p < .001$, $\eta_p^2 = .30$ (day one $M = 0.136$, $SE = 0.004$; day two $M = 0.122$, SE

=0.002; day 3 $M = 0.116$, $SE = 0.002$). There was a non-significant trend towards Z decreasing

on later days, F (2, 58) = 2.88, $\varepsilon = .93$, $p = .07$, $\eta_p^2 = .30$ (day one $M = 0.492$, $SE = 0.007$; day two

$M = 0.478$, $SE = 0.005$; day 3 $M = 0.473$, $SE = 0.006$).

For *Ter*, a PM by day ANOVA revealed a main effect of PM, F (2, 58) = 29.01, $\varepsilon = .96$, $p <$

0.001, $\eta p2 = .50$, with *Ter* lower in the control condition ($M = 0.349$, $SE = 0.007$) than in the

focal condition ($M = 0.360$, $SE = 0.007$), t (29) = 3.06, $p < .01$, $d = 0.56$, and in the non-focal

condition ($M = 0.377$, $SE = 0.008$), t (29) = 6.92, $p < .001$, $d = 1.26$. There was also a main effect

of day, F (2, 58) = 6.84, $\varepsilon = .84$, $p < .01$, $\eta_p^2 = .19$, with *Ter* decreasing over days (day 1 $M =$

0.376, $SE = 0.009$; day 2 $M = 0.361$, $SE = 0.006$; day 3 $M = 0.349$, $SE = 0.007$). The PM and day

effects did not interact, $F < 1$.

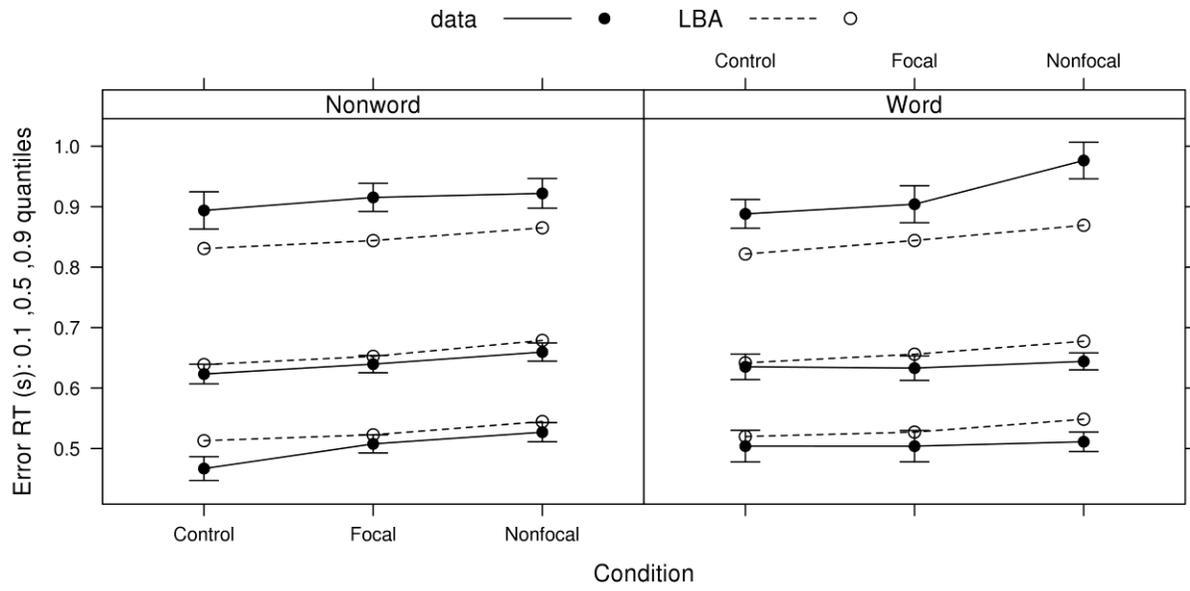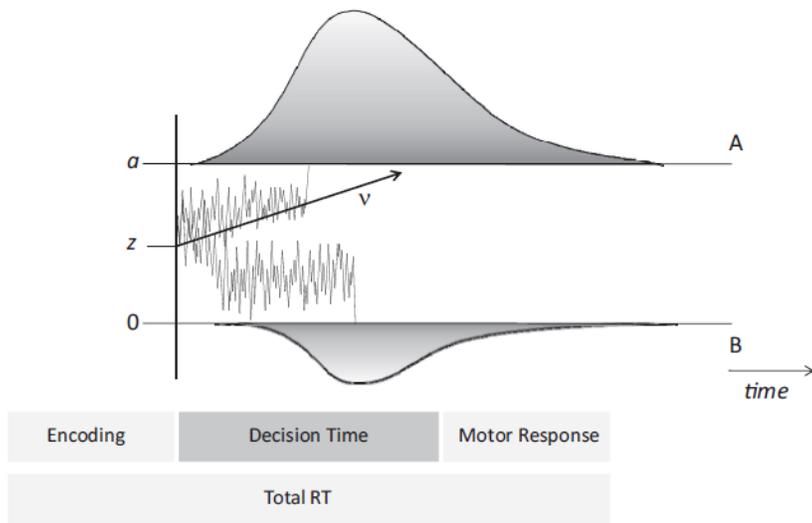*Figure 1.* BIC selected LBA model fits. Error bars calculated using the Morey (2008) bias-corrected method.
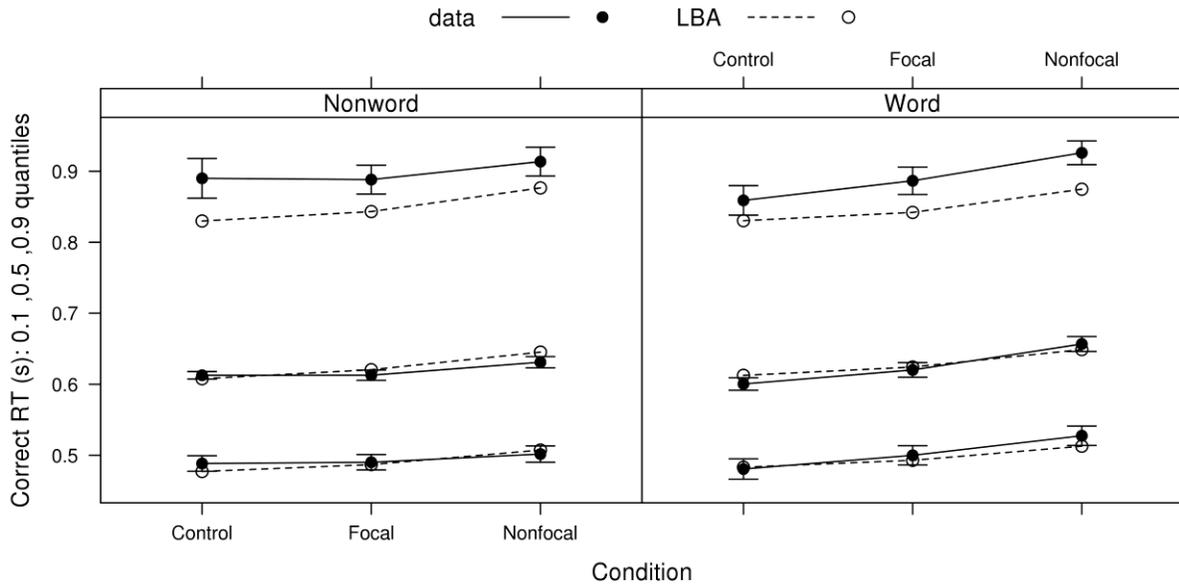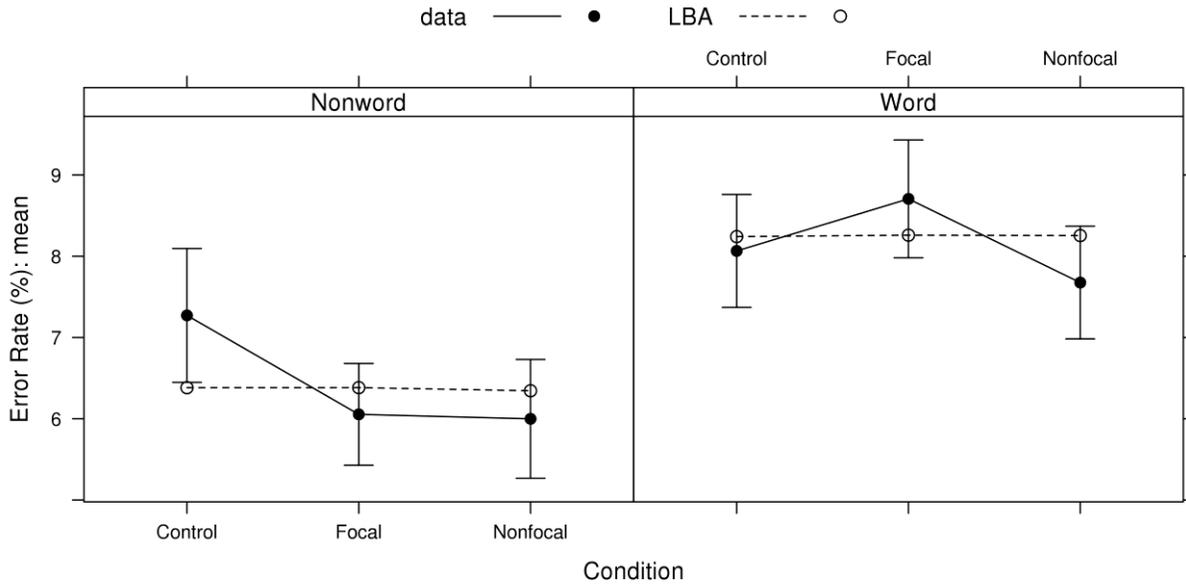
Table 2. DDM model selection. Model factors include condition (PM), stimulus type (S), and day (D).

| Model	a	Z	v	Ter	p	Deviance	AIC	BIC
Top Model	PM, D	PM, D	S, PM, D	PM, D	1440	-88820	-85940	-72163
AIC Model	PM, D	PM, D	S, D	PM, D	1080	-88425	-86265	-75932
BIC Model	D	D	1	PM*D	570	-85703	-84563	-79110

Figure 2. Illustration of the DDM (illustration taken from Horn et al., 2011)

RT = Accumulation time+ t0

Figure 3. DDM AIC selected model fits to error rates (left column), RT distribution for correct responses(10th, 50th, and 90th percentiles, middle column) and the 50th percentile (median) for error responses. Error bars calculated using the Morey (2008) bias-corrected method.
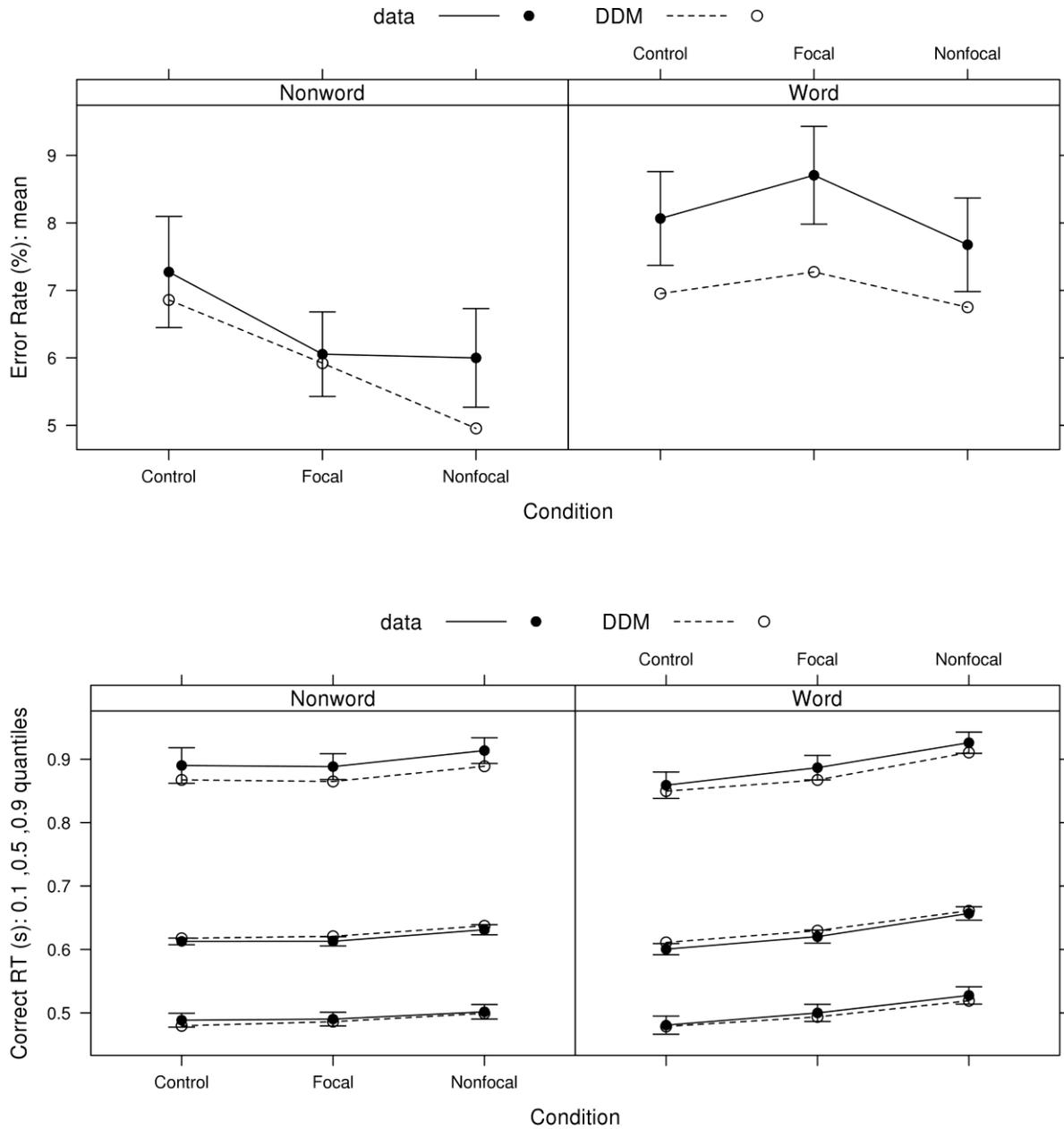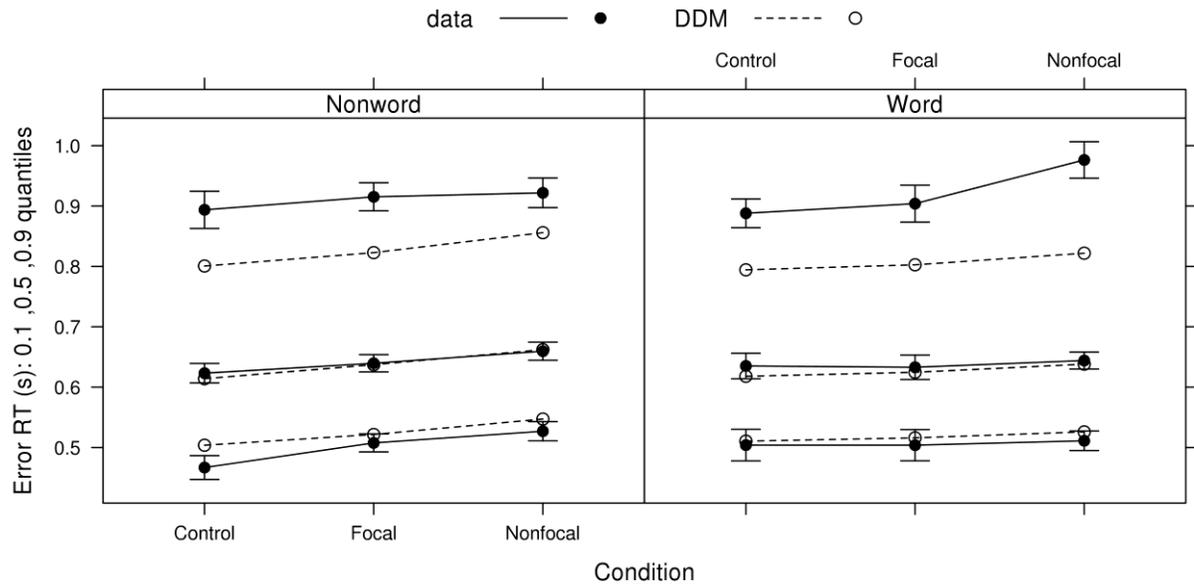
Figure 4. BIC selected DDM model fits. Error bars calculated using the Morey (2008) bias-corrected method.
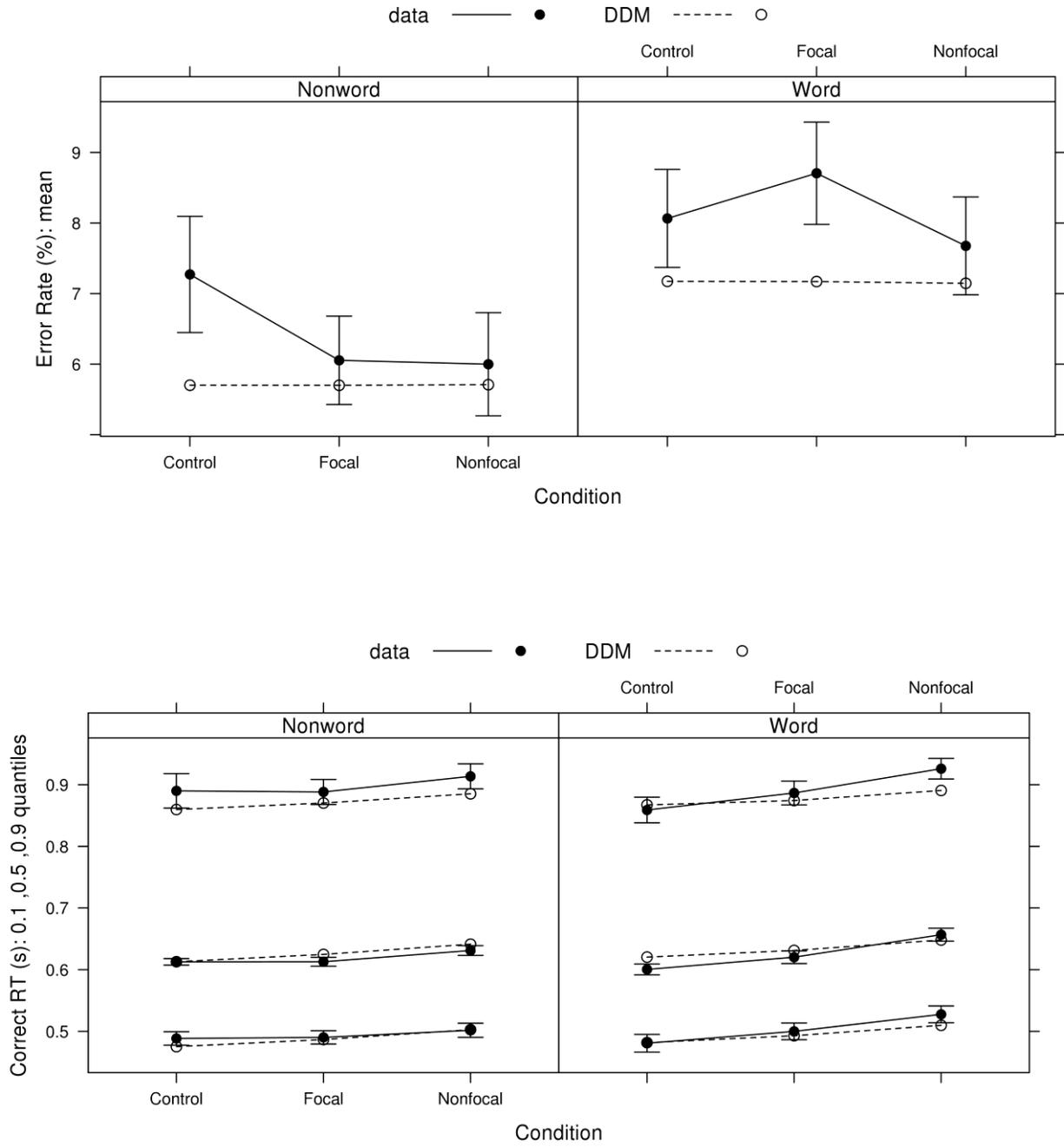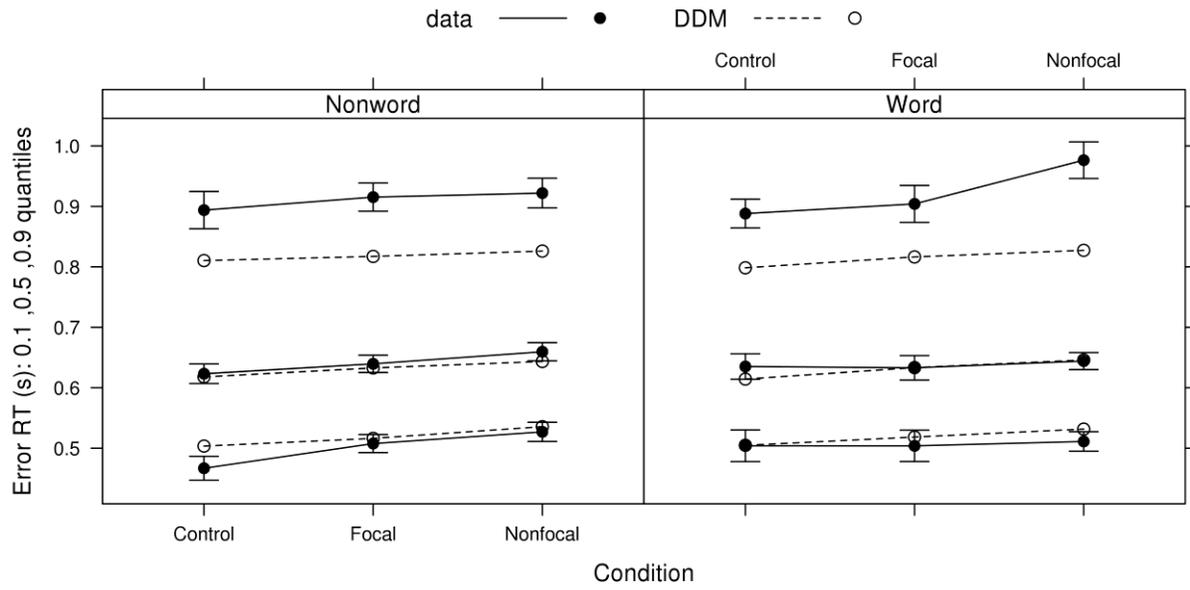
Figure 5. Top LBA parameter estimates of mean drift rates. In general, minimal effects of PM condition, on the overall quality of processing (correct – error rate). Panels broken in half vertically (lower range on the top array, upper range on the bottom array), so that we could account for the different scales of different participants.
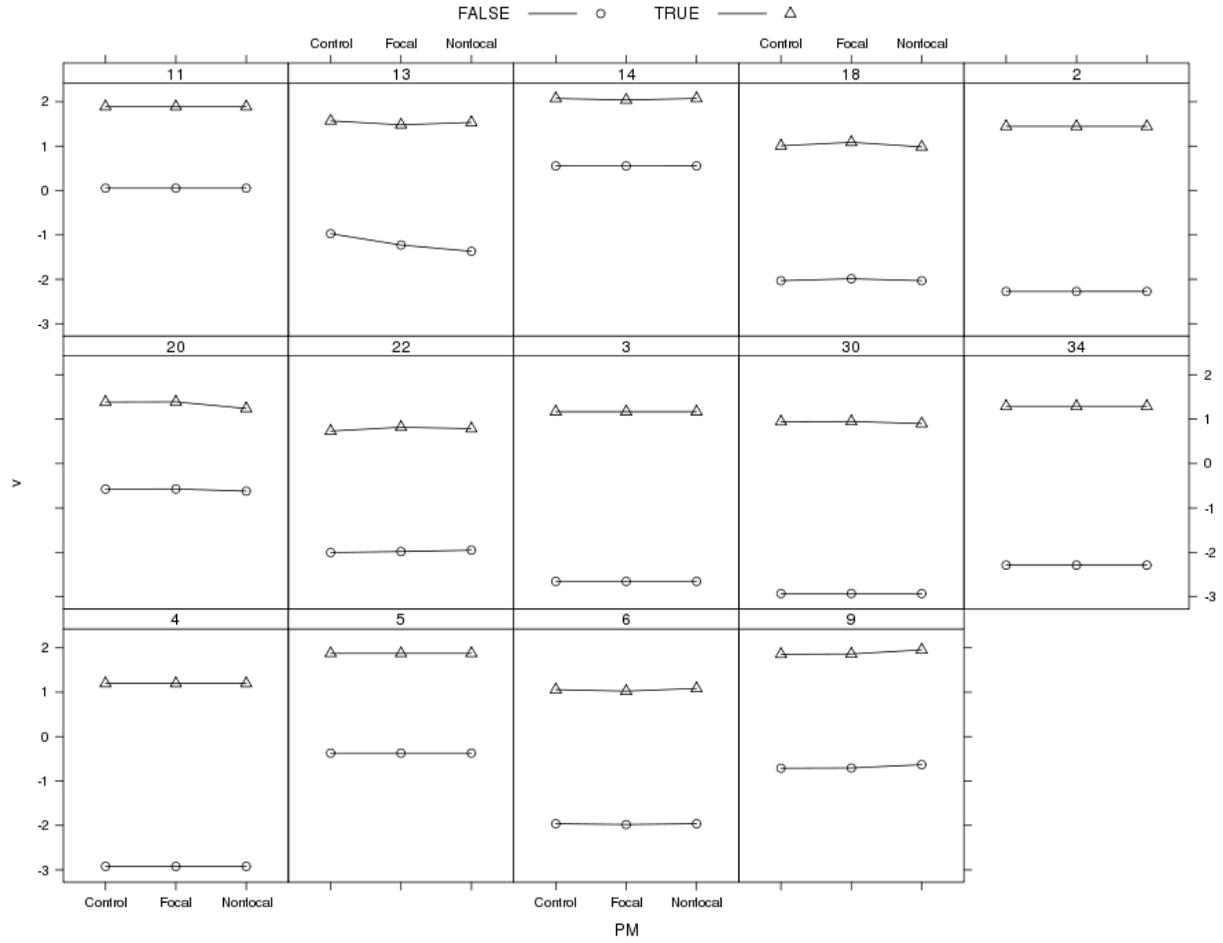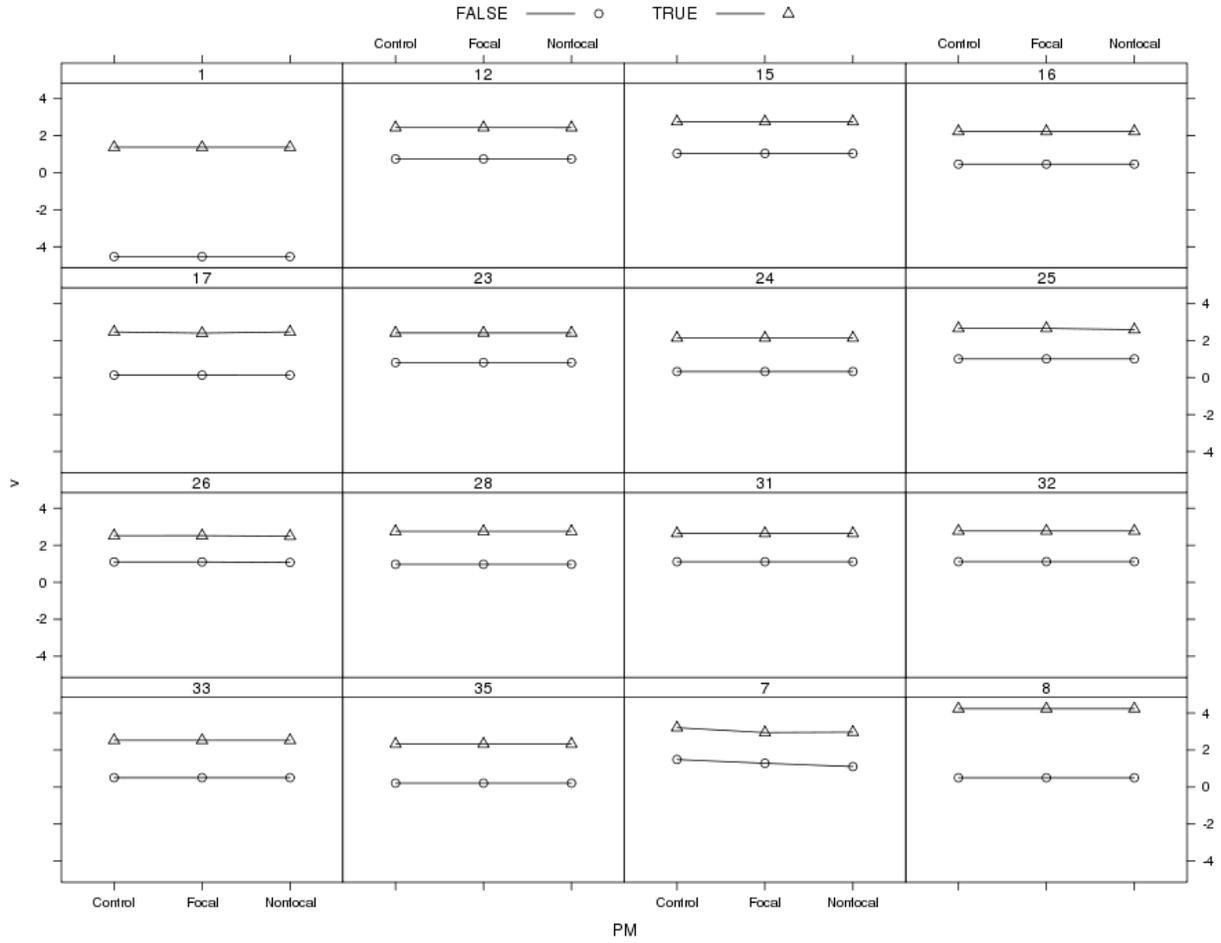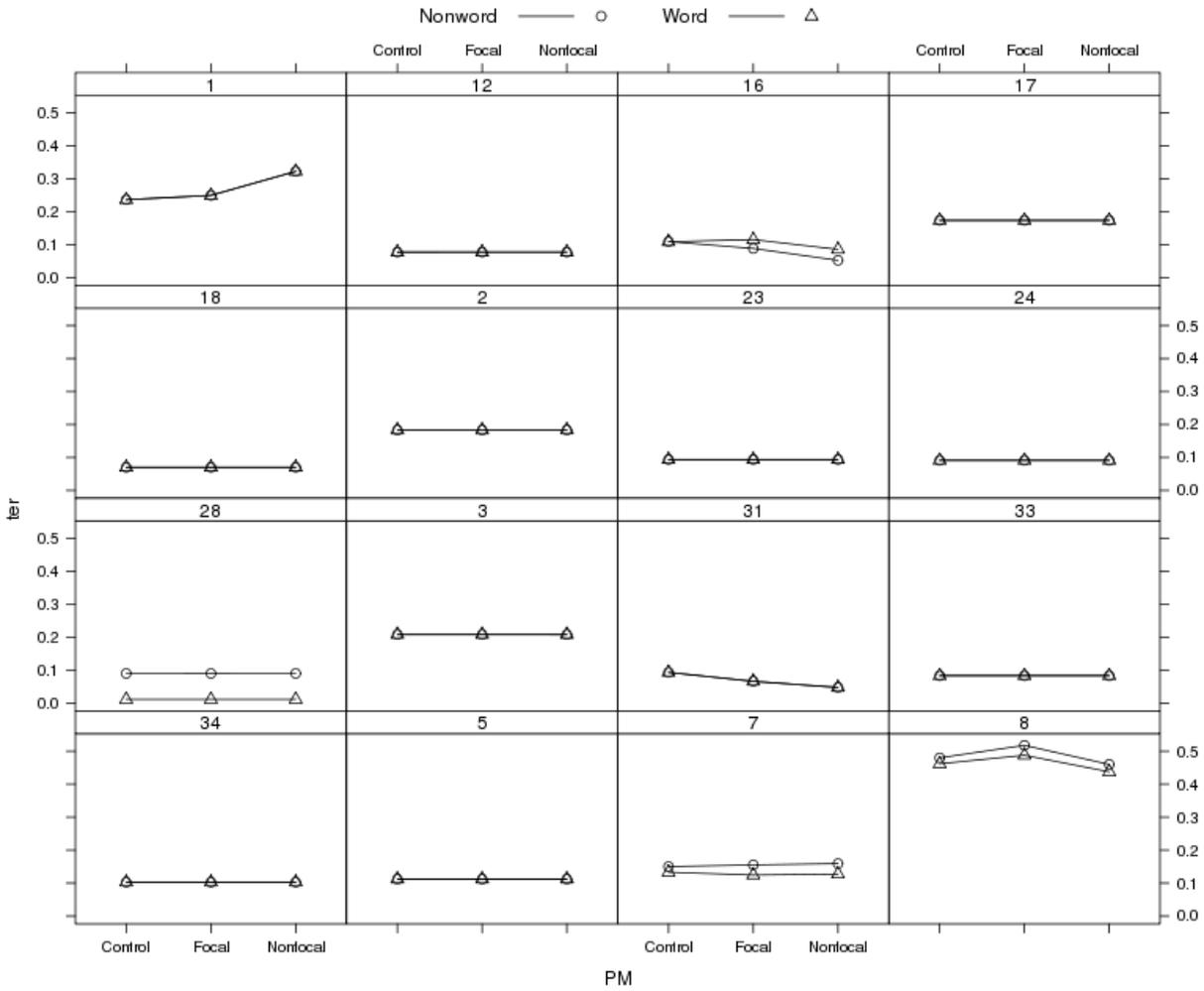
Figure 7. Top LBA parameter estimates of t0. Top panels low t0 participants, bottom panels high t0 participants. Minimal effects of PM condition, although some significant variation in participants 1, 8, 15, and 31.
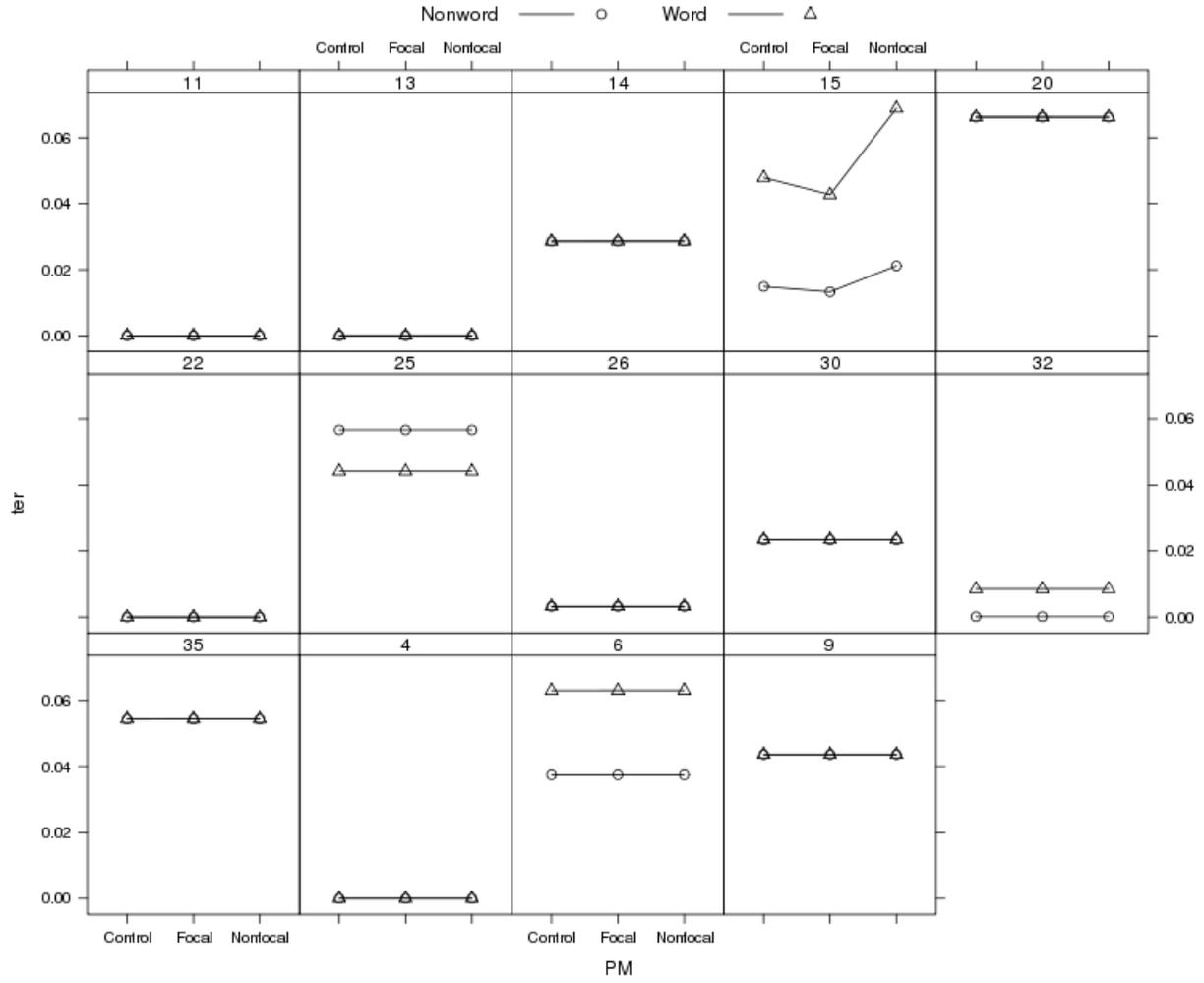
Figure 6. Top LBA parameter estimates of the standard deviation of each participants' drift rates. In general, minimal effects of PM condition. If anything, *sv* was reduced under PM conditions, which is the opposite of the prediction of the attention lapse view.
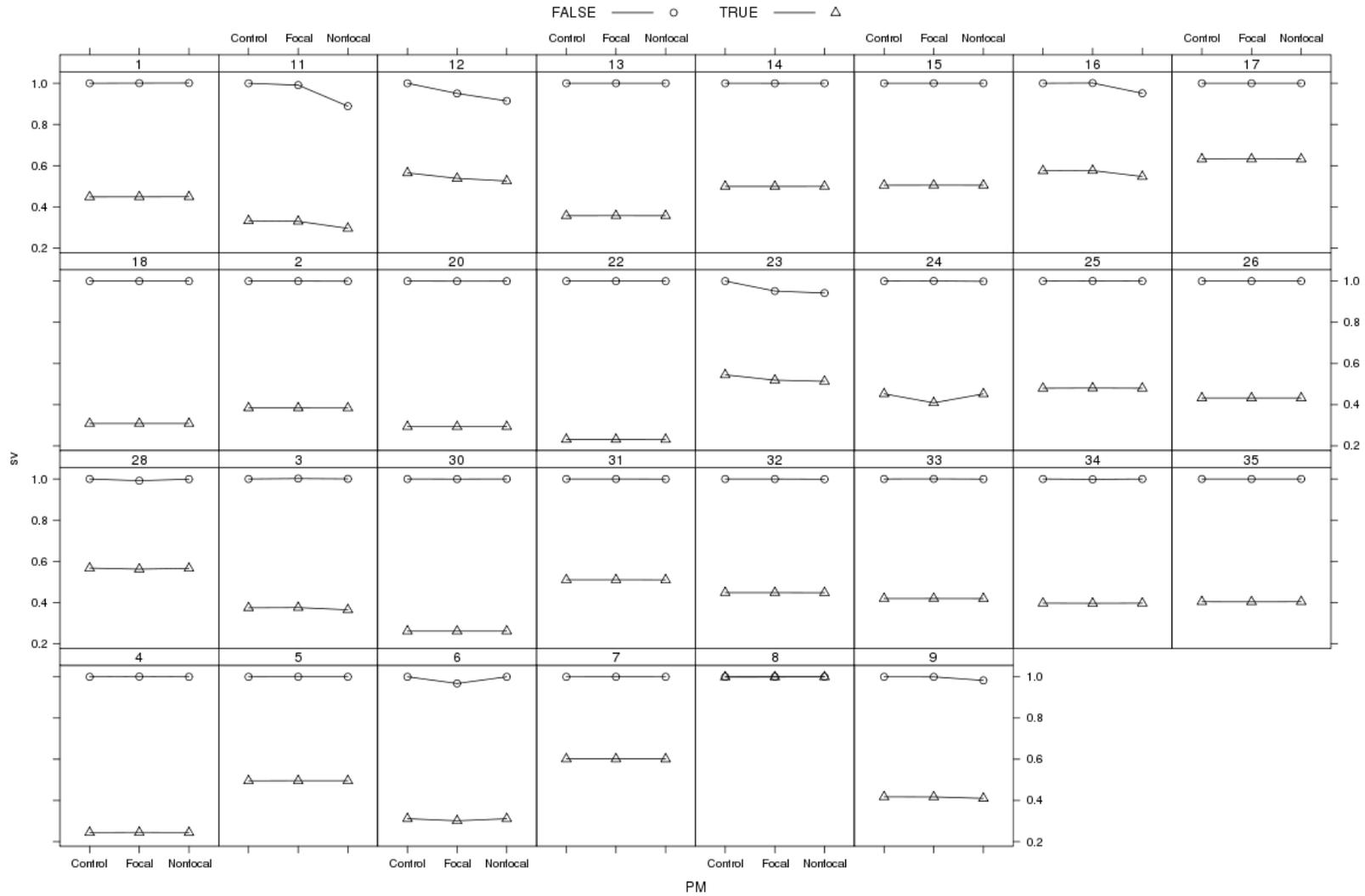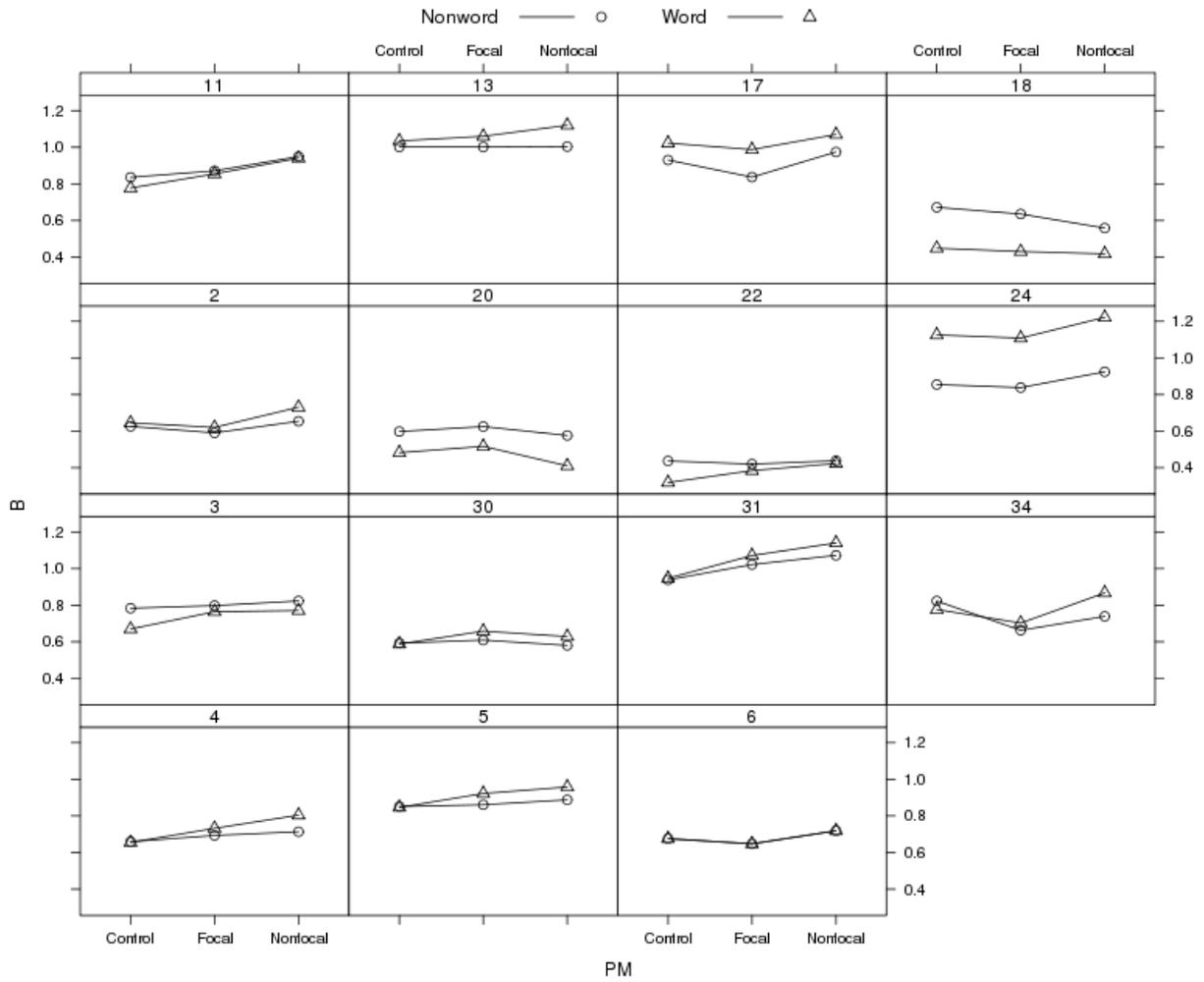
Figure 8. Top LBA parameter estimates of response threshold. Top panels low threshold participants, bottom panels high threshold participants. Participants 13, 34, 4, 5, 12, 14, 16, 23, 25, 28, 32, 33, and 9 showed a larger increase to the word threshold to some extent, whereas there was no participant for which the reverse pattern was obvious.