Results

Table S1. *Wald Chi-square significance tests on mixed effects models of correct RTs and accuracies. Accuracy on each trial was modelled with generalized linear models and a probit link function. Participant mean correct RTs were modelled with a linear mixed effects model. S stands for ongoing-task stimulus type (word, non-word), Bias is the blocked manipulation (word caution, non-word caution), Day is the day of the experiment (one, two). Each test has one degree of freedom. Each term's significance test was calculated with a null model that included all other terms, except for higher order interactions involving the term (e.g., S x Bias would be ignored when testing S).*

	Ongoing Task				**PM Task**			
Factor(s)	*Accuracy*		*Response Time*		*Accuracy*		*Response Time*	
	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*
S	49.1	<.001	63.26	<.001	7.55	.006	0.01	.91
Bias	24.88	<.001	4.28	.039	1.36	.244	0.36	.55
Day	8.99	.003	49.92	<.001	117	<.001	52.07	<.001
S x Bias	836.59	<.001	34.54	<.001	2.03	.154	2.33	.13
S x Day	1.6	.205	0.41	.520	0.19	.661	0.32	.57
Bias x Day	0.04	.835	0.61	.433	0.31	.579	0.17	.68
S x Bias x Day	1.96	.162	0.02	.901	1.85	.174	0.19	.66

Table S2. *Wald Chi-square significance tests on a mixed effects model of participant mean ongoing task error RTs. S stands for ongoing-task stimulus type (word, non-word), Bias is the blocked manipulation (word caution, non-word caution), Day is the day of the experiment (one, two). Each test has one degree of freedom. Each term's significance test was calculated with a null model that included all other terms, except for higher order interactions involving the term (e.g., S x Bias would be ignored when testing S).*

Factor(s)	χ^2	*p*
S	3.45	.063
Bias	8.07	.004
Day	35.17	<.001
S x Bias	0.84	.360
S x Day	0.49	.484
Bias x Day	0.6	.437
S x Bias x Day	0.08	.772

Table S3. *Wald Chi-square significance tests on mixed effects models of PM miss type (word response, non-word response) and PM miss RTs. We fit a linear mixed effects model to mean participant miss type, and to mean participant miss RTs. S stands for ongoing-task stimulus type (word, non-word), Bias is the blocked manipulation (word caution, non-word caution), Day is the day of the experiment (one, two), Type is the type of PM miss. Each test has one degree of freedom. Each term's significance test was calculated with a null model that included all other terms, except for higher order interactions involving the term (e.g., S x Bias would be ignored when testing S).*

Factor(s)	*Proportion*		*Response Time*	
	χ^2	*p*	χ^2	*p*
S	1.1	.294	6.56	.010
Bias	0.26	.609	0.29	.590
Type	0.3	.586	10.01	.002
Day	20.09	<.001	15.37	<.001
S x Bias	0.29	.591	4.78	.029
S x Type	496.29	<.001	18.32	<.001
Bias x Type	8.92	.003	14.95	<.001
S x Day	0.09	.761	1.8	.179
Bias x Day	0.05	.827	1.12	.290
Type x Day	0.08	.779	0.1	.755
S x Bias x Type	1.81	.178	0.07	.796
S x Bias x Day	0.27	.605	0.08	.777
S x Type x Day	14.31	<.001	1.37	.241
Bias x Type x Day	1.07	.301	0.31	.575
S x Bias x Type x Day	0.01	.910	0.55	.457

Alternative Model with Varying Non-Decision Time by Day

A reviewer noted the possibility of non-decision time being affected by practice. To test this proposal, we fit an alternative PMDC model, which was identical to that reported in text except that it allowed non-decision time to vary across day one and day two of the experiment. We found that non-decision time was estimated to be larger on day two of the experiment (M = 0.17s, SD = 0.004) than on day one (M = 0.14s, SD = 0.004), inconsistent with a practice effect. Given this finding is inconsistent with a practice effect, we still present the model with a single non-decision time in text, which we believe provides more robust parameter estimation. However, for completeness, we report estimates from the alternative model here. The posterior mean of the A parameter was 0.47 (SD = 0.02). The mean of the standard deviation of matching accumulation was 0.59 (SD = 0.006). The thresholds are graphed in Figure S1, and estimated accumulation rates in Figures S2 and S3. They provide similar inferences to the reported model.

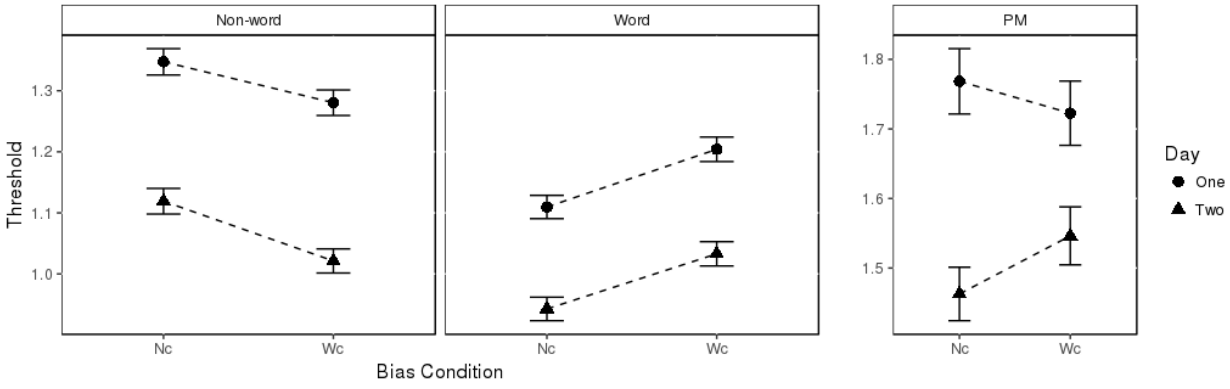

Figure S1. Participant-averaged thresholds for a model that allowed non-decision time to vary by day. The symbols correspond to the posterior mean, and the error bars correspond to plus and minus one standard deviation from the mean. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions.

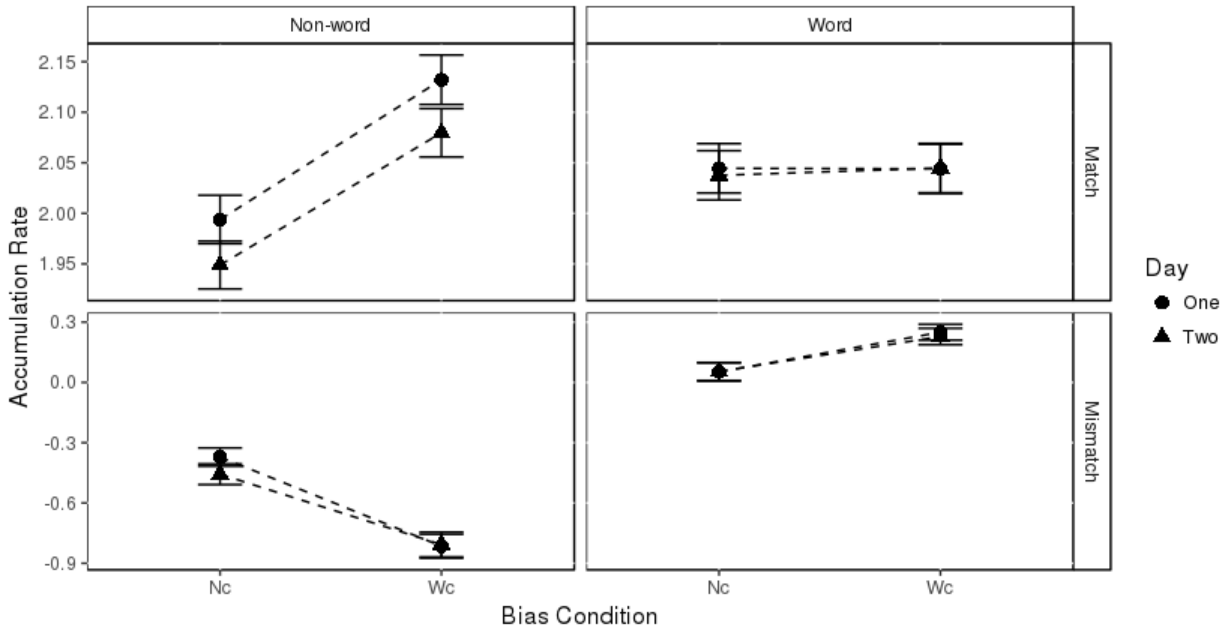

Figure S2. Participant-averaged non-PM trial accumulation rates for word and non-word stimuli (columns). Obtained from a model that allowed non-decision time to vary by day. The symbols correspond to the posterior mean, and the error bars correspond to plus and minus one standard deviation from the mean. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions. Accumulation to PM decisions on non-PM trials (M = -2.82, SD = 0.11), that is PM false alarm accumulation, is not plotted.

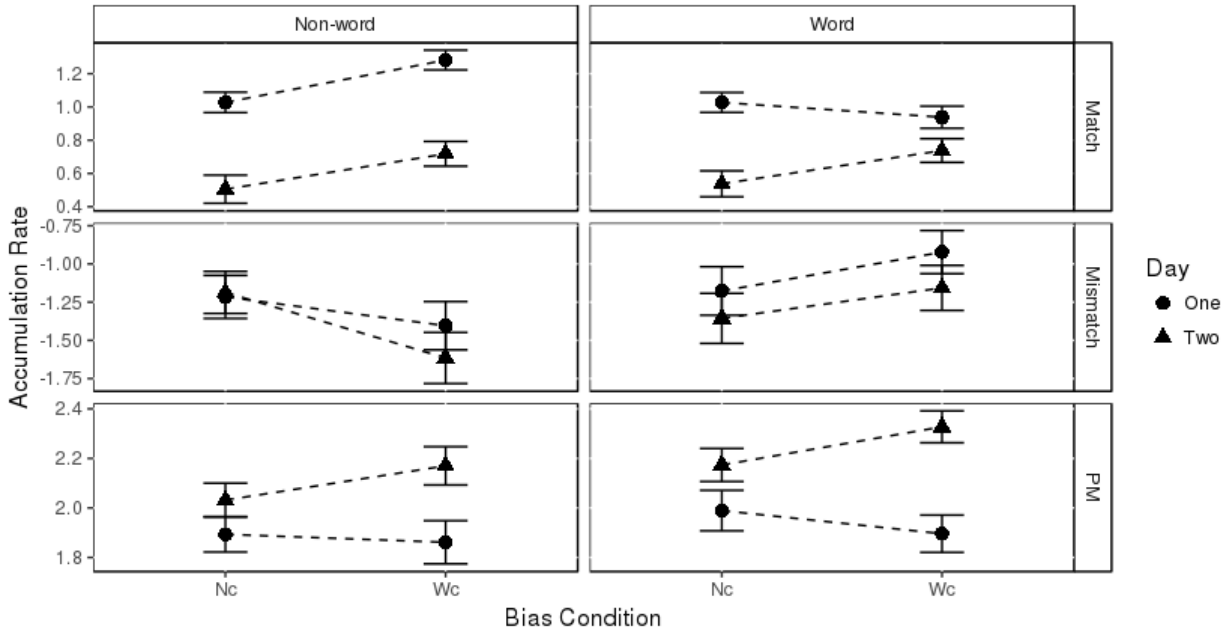

Figure S3. Participant-averaged PM trial accumulation rates from a model that allowed non-decision time to vary by day. The symbols correspond to the posterior mean, and the error bars correspond to plus and minus one standard deviation from the mean. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions.

Posterior Exploration

To evaluate how model mechanisms contributed to the observed effects of the bias manipulation on ongoing-task responses, we systematically removed mechanisms from the model by averaging parameters across the bias conditions and examined the resulting effects on simulated performance. We simulated a model with no effects of bias condition on ongoing-task thresholds (ongoing-task thresholds averaged across bias conditions), a model with no effects of bias condition on accumulation rates (ongoing-task accumulation rates averaged across bias conditions), and a model with no effect of bias condition on either. Because the bias manipulation did not induce substantial shifts in PM accuracy, we do not examine how the model predicts differences in PM accuracy across conditions. Instead we examined predictions of shifts in ongoing-task response proportions and RTs across bias conditions. We discuss the percentage of each effect produced by the posterior mean predictions of the models and provide more detailed summaries of fit in Figures S4 and S5. Note that predicted effects can be larger than the observed, and so this percentage can exceed 100%.

We first examined the cognitive processes underlying shifts in ongoing-task response type. To summarize shifts in response rate, we averaged the decreases in non-word response proportion in the Nc condition and the decreases in word response proportion in the Wc condition. On non-PM trials, this averaged shift in response proportion was 0.07 in the data. The full model predicted this shift in response rates well, with an effect 102% the size of the observed. Averaging ongoing-task thresholds over bias condition caused severe mis-fit, resulting in predicting 53% of the observed effect. Averaging ongoing-task rates over bias condition also caused substantial mis-fit, resulting in predicting 60% of the observed effect. Averaging both ongoing-task rates and ongoing-task thresholds out of the model resulted in predicting only 12.3% of the effect.

We next examined shifts in ongoing-task response type on PM trials. We examined the sum of the non-word response proportion decreases in the Nc condition and the word response proportion decreases in the Wc conditions (0.05). The full model predicted 104% the size of the effect, the averaged ongoing-task thresholds model predicted 50% of the effect, and the averaged ongoing-task rates model predicted 47% of the effect. The model with rates and thresholds did not predict any effect at all, instead predicting a slight negative difference (-5% of the effect).

We also examined the cognitive processes underlying shifts in ongoing-task RTs on non-PM trials. To calculate an overall measure of shifts in RT, we averaged the RT increases to non-word responses in the Nc condition with the increases to word responses in the Wc conditions (0.05s). The full model accounted for shifts in RT with bias accurately, predicting an effect 105% the size of the observed. By contrast, averaging ongoing-task thresholds across bias conditions resulted in predicting only 32% of the observed shifts in RT. Averaging rates across bias conditions resulted in predicting 77% of observed shifts. Averaging both ongoing-task thresholds and ongoing-task rates across conditions resulted in predicting effectively none of the observed shifts in RT (3.3%). We also examined shifts in ongoing-task RTs on PM trials (0.051s). The full model predicted 95% of the observed effect. Averaging thresholds led to predicting 36% of the observed effect, whereas

averaging rates led to predicting 92% of the observed effect. Averaging both thresholds and rates led to predicting 28% of the observed effect.

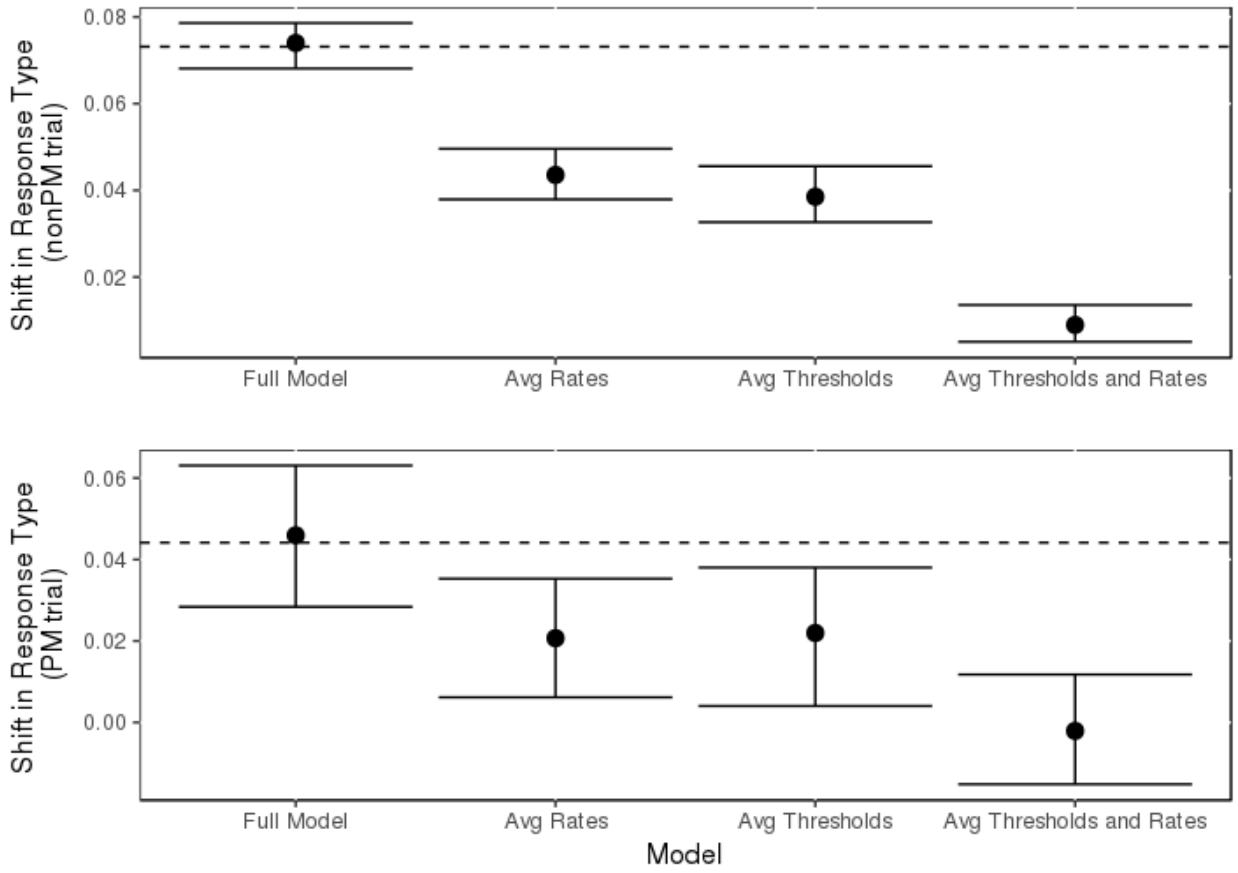

Figure S4. Posterior exploration of the effects of the bias manipulation on shifts in ongoing task response type. To summarize shifts in response type, we averaged the decreases in non-word response proportion in the non-word caution condition with the decreases in word response proportion in the word caution condition. The dashed lines indicate the observed data. Full model corresponds to the full posterior prediction of the estimated model. Avg stands for averaged. These models are simulated based on averaging ongoing task accumulator parameters across bias conditions, for example in the "Avg" threshold model the ongoing task thresholds were averaged across the word caution and non-word caution conditions. The black dots indicate the posterior mean predictions. The black bars indicate the 95% credible intervals of the posterior predictions. For the purposes of these graphs, we concatenated all participants' data into a single data frame, and posterior predictions for all participants into a single data frame, and then calculated summary statistics on this data frame. "

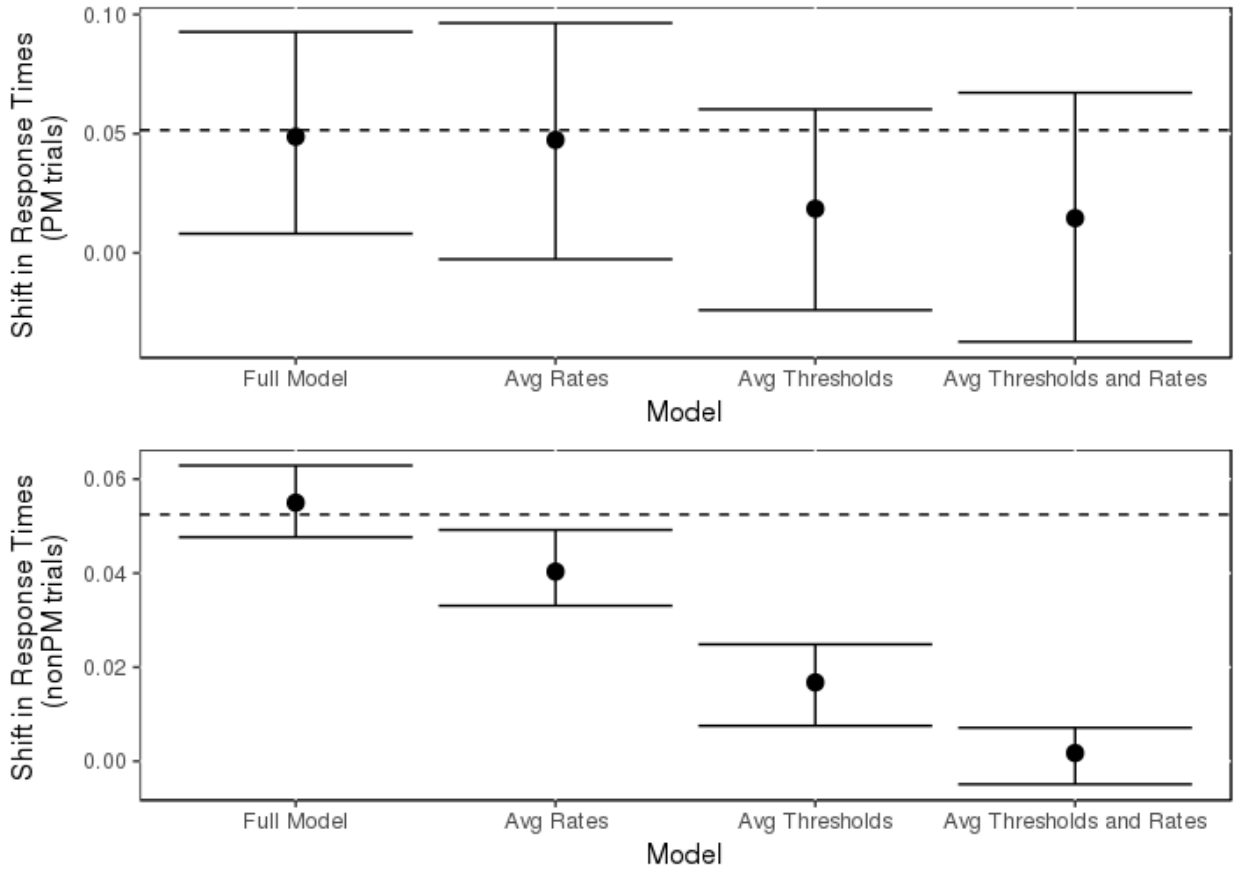

Figure S5. Posterior exploration of the effects of the bias manipulation on shifts in ongoing task response times. To summarize shifts in response time, we averaged the increases in non-word response time in the non-word caution condition with the increases in word response time in the word caution condition. The dashed lines indicate the observed data. Full model corresponds to the full posterior prediction of the estimated model. Avg stands for averaged. These models are simulated based on averaging ongoing task accumulator parameters across bias conditions, for example in the "Avg" threshold model the ongoing task thresholds were averaged across the word caution and non-word caution conditions. The black dots indicate the mean predictions. The black bars indicate the 95% credible intervals of the posterior predictions. For the purposes of these graphs, we concatenated all participants' data into a single data frame, and posterior predictions for all participants into a single data frame, and then calculated summary statistics on this data frame. "

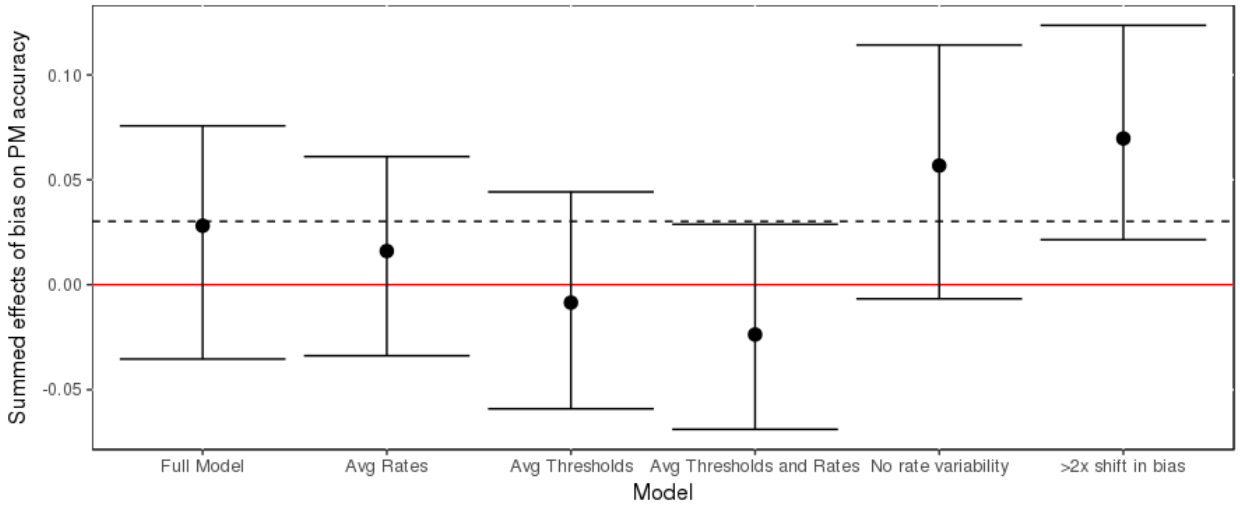

Figure S6. Posterior exploration and simulation exploring whether bias could conceivably produce a shift in PM accuracy. To calculate an overall measure of the effect of bias condition on PM accuracy, we summed the increase in PM accuracy to PM non-words in the non-word caution condition with the increase in PM accuracy to PM words in the word caution condition. The dashed lines indicate the observed data. The solid red line indicates an effect of 0. Full model corresponds to the full posterior prediction of the estimated model. Avg stands for averaged. These models are simulated based on averaging ongoing task accumulator parameters across bias conditions, for example in the "Avg" threshold model the ongoing task thresholds were averaged across the word caution and non-word caution conditions. The "no rate variability" model was simulated by effectively setting the variability in evidence accumulation rates to zero. The ">2x shift in bias" model was simulated by adding 0.1 to the word threshold in the word caution condition and 0.1 to the non-word threshold in the non-word caution condition. The black dots indicate the mean predictions. The black bars indicate the 95% credible intervals of the posterior predictions. For the purposes of these graphs, we concatenated all participants' data into a single data frame, and posterior predictions for all participants into a single data frame, and then calculated summary statistics on this data frame. "