Supplemental Materials

Posterior Predictive Checks of Model Fit

As parameter estimates from a descriptively inadequate model would be uninformative, posterior predictive checks (Gelman, Meng, & Stern, 1996) were used to evaluate the model's absolute goodness-of-fit. These checks compare the observed data to data sets predicted using the posterior distributions of the cognitive model parameters, which should closely approximate the observed data if the model fits well. As the predicted data sets are generated using the entire posterior distributions of model parameters, plots that describe the uncertainty of predicted data values are directly representative of uncertainty about model parameter estimates and the data generating process.

Joint cumulative distribution function plots (Figure S1a) display the cumulative probability of an "X" or "O" response over time for empirical data and posterior predictive data (averaged across participants) separately for each stimulus type and trial type ("go" vs. "signal" trials). Five RT quantiles (.10, .30, .50, .70, .90) of the empirical (circles) and predictive (dots) data are also plotted in order to better visualize misfit. In general, with the potential exception of "O" stimulus signal-respond trials in the ADHD group, the model appears to display excellent fit to correct RT. However, although the model describes overall accuracy rates well, fits to error RTs are much poorer than those to correct RTs, with marked over-prediction of slow error response quantiles. Given the relatively low number of error trials, this pattern of misfit is expected; the likelihood function used to fit the model applies more weight to the most common trial types, and the model can therefore be expected to describe these trials better than trials that are rare, such as errors.

However, the pattern of misfit indicates that the model is struggling to produce the fast error RTs found in this data set, which is consistent with two possibilities. First, it is possible that some of the children's RTs were fast guesses rather than products of the race process. As these trials would be fast and have chance accuracy, the race model is unable to account for them, and therefore predicts error RT values that are slightly longer than the true values, which are contaminated by the fast guessing process. Second, it is possible that the misfit is simply due to the fact that the τ_{go} parameter was not allowed to vary by match vs. mismatch. Regardless, as there were relatively few error responses in the task, the misfits in error trials are unlikely to impact inferences made with the model parameters of interest.

Figures S1b and S1c assess whether the model can account for two key ways in which performance on stop-signal trials varies with changes in SSD. Figure S1b displays the median signal-respond RTs (SRRTs) for both correct and error responses at eight bins of SSD. These bins were *absolute*, in that the quantile values that determine them were established by collapsing across the whole group's SSD values. Empirical data is displayed as black lines and dots, while posterior predictive data is displayed as violin plots, which are box plots of posterior predictive data values surrounded by kernel density plots of the same values. The density plots represents relative certainty about the location of the predicted medians. Consistent with assumptions of the race model, both groups' correct SRRTs increased as SSD increased. The posterior predictive data indicates that, despite some over-prediction of medians at short SSDs, the current model-based analysis generally captured this increasing pattern well. Error SRRTs, however, did not show the expected increase with SSD, although the model generally captured the absolute values of these SRRTs well.

Figure S1c displays the "inhibition functions" of both groups, which show the probability of a response, *P(response)*, at eight bins of SSD. Unlike the CE-RT plots, these bins were *relative*, in that the quantiles that determine them were established separately for each individual based on their unique distribution of SSDs. The relative binning procedure allows the expected pattern, in which *P(response)* increases as the SSD increases, and therefore leaves less time for the stop process to win the race, to be more clearly represented. The model appeared to capture both the general pattern of SSD-related increase and the absolute values of *P(response)* for most SSD bins well.

Overall, the ex-Gaussian race model captured general patterns in the behavioral data (e.g., accuracy, hallmark SSD effects) very well, but displayed a poorer description of error RTs, possibly because of the fact that we constrained τ_{go} to be the same between match and mismatch, or because of an un-modeled fast-guessing process. As the model effectively explains key features of behavioral performance, and as errors were relatively rare in this data set, it is reasonable to conclude that this misfit will not confound the main analyses of interest, which are aimed at assessing general parameter differences between the groups. Furthermore, as we found that there were no substantive differences between the results of the main analysis and our separate analysis of the high-accuracy subset of participants with a model that did not include error RTs (described below), we concluded that the effects of interest were unlikely to be confounded by the error misfits.

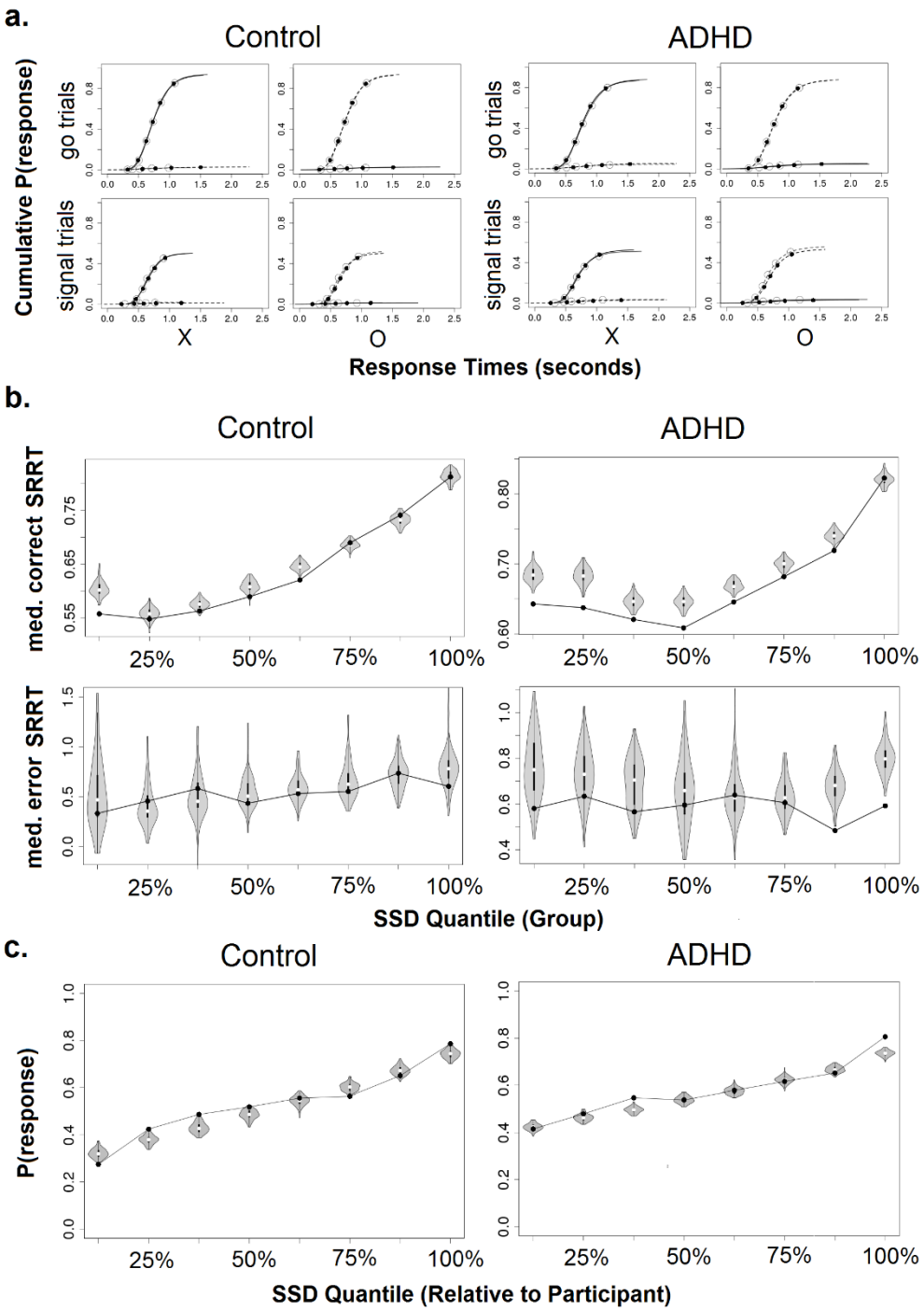
Secondary Model Analysis with High-Accuracy Participants

In the secondary analysis, we applied the traditional ex-Gaussian race model, in which only correct RTs are considered and mismatch RT distribution parameters are not estimated (Matzke, Dolan, Logan, Brown, & Wagenmakers, 2013; Matzke, Love, & Heathcote, 2016), to a

high-accuracy (>.95 accuracy rate) subset of participants. The conservative accuracy exclusion criteria removed 42% of the ADHD group (n = 88) that was included in the main analysis, and 11% of the Control group (n = 11). The same priors and estimation procedures as those described above were used to fit this model, and the analysis strategies were the same as those described in the text.

Posterior medians and Bayesian *p*-values for differences in all group-level model parameters are displayed in Supplemental Table 1. As demonstrated by these values, results from the secondary analysis are similar to those of the main analysis. This analysis replicated group effects of P_{GF}, P_{TF} and τ_{stop}, although evidence for τ_{stop} differences was weaker, p=.165. The weakness of evidence for this effect, and the generally lower estimates of P_{GF} and P_{TF} in the ADHD group relative to the main analysis, are likely due to the fact that a large portion of this group, and potentially the individuals with the most severe impairment, were excluded. Consistent with the findings of greater $\sigma_{go\text{-}match}$ and τ_{go} in the main analysis, children with ADHD also displayed higher σ_{go} and τ_{go}. Finally, in contrast with the main analysis, there was only weak evidence for greater between-subjects variability in τ_{stop} in the ADHD group, p=.127, and little evidence for greater P_{TF}, p=.443. Removal of participants with the most severe impairment would be expected to reduce between-subject variability in these processes, however, and this pattern therefore suggests that the between-subject variability effects in the main analysis were driven by individuals with the greatest impairment.

Supplemental Figure 1. Empirical data plotted against posterior predictive data from the model. **a)** Joint cumulative distribution function plots of the probability of X (solid line) and O (dotted line) responses for empirical (gray, circles) and posterior predictive (black, dots) data, by stimulus type and trial type, **b)** Medians of correct and incorrect signal-respond RTs (SRRTs) by group average SSD quantile for empirical (black) and posterior predictive (gray violin plots) data. **c)** Probability of response by relative SSD quantile (i.e., quantile for each individual participant) for empirical (black) and posterior predictive (gray violin plots) data.



Supplemental Table 1. Posterior medians of all group-level parameter values and Bayesian *p*-values for each group difference in the secondary analysis. Group mean parameters for P_{TF} and P_{GF} have been transformed back to the probability scale for clarity, and technically reflect the medians, rather than the means, of the group distributions. For the $\sigma_{stop,}$ and τ_{stop} parameters, population means and SDs (denoted with "Pop.") are reported in addition to the means and SDs for the truncated normal distributions. Due concerns about severe truncation at 0 (outlined in the text) the population parameters of $\sigma_{stop,}$ and τ_{stop} were used to make inferences about group differences.

Group Means

Parameter	Control Post. Median	ADHD Post. Median	Bayesian *p*-value
μ_{go}	.611	.644	.034
σ_{go}	.110	.129	.003
τ_{go}	.169	.190	.033
μ_{stop}	.164	.175	.043
σ_{stop}	.016	.015	.456
τ_{stop}	.029	.029	.499
σ_{stop} Pop.	.038	.036	.444
τ_{stop} Pop.	.073	.089	.165
P_{GF}	.019	.034	<.001
P_{TF}	.128	.208	.001

Group SDs

Parameter	Control Post. Median	ADHD Post. Median	Bayesian *p*-value
μ_{go}	.118	.124	.323
σ_{go}	.035	.051	.001
τ_{go}	.070	.076	.257
μ_{stop}	.035	.028	.074
σ_{stop}	.038	.038	.486
τ_{stop}	.076	.097	.105
σ_{stop} Pop.	.026	.026	.461
τ_{stop} Pop.	.051	.064	.127
P_{GF}	.410	.483	.084
P_{TF}	.673	.686	.443

Supplemental References

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, 733-760.

Matzke, D., Dolan, C. V., Logan, G. D., Brown, S. D., & Wagenmakers, E.-J. (2013). Bayesian parametric estimation of stop-signal reaction time distributions. *Journal of Experimental Psychology: General*, *142*(4), 1047.

Matzke, D., Love, J., & Heathcote, A. (2016). A Bayesian approach for estimating the probability of trigger failures in the stop-signal paradigm. *Behavior research methods*, 1-15.