nature portfolio

https://doi.org/10.1038/s41562-021-01282-7

Supplementary information

Mental speed is high until age 60 as revealed by analysis of over a million participants

In the format provided by the authors and unedited

Mental speed is high until age 60 as revealed by analysis of over a million participants
Supplementary Information

Mischa von Krause[1], Stefan T. Radev[1], and Andreas Voss[1]

[1]Institute of Psychology, Heidelberg University, Heidelberg, Germany

Bayesian Workflow

In the following, we describe and visualize the details of our Bayesian pipeline, which we separate into four main steps, as advocated by the notion of a principled Bayesian workflow [1].

Domain Expertise Consistency

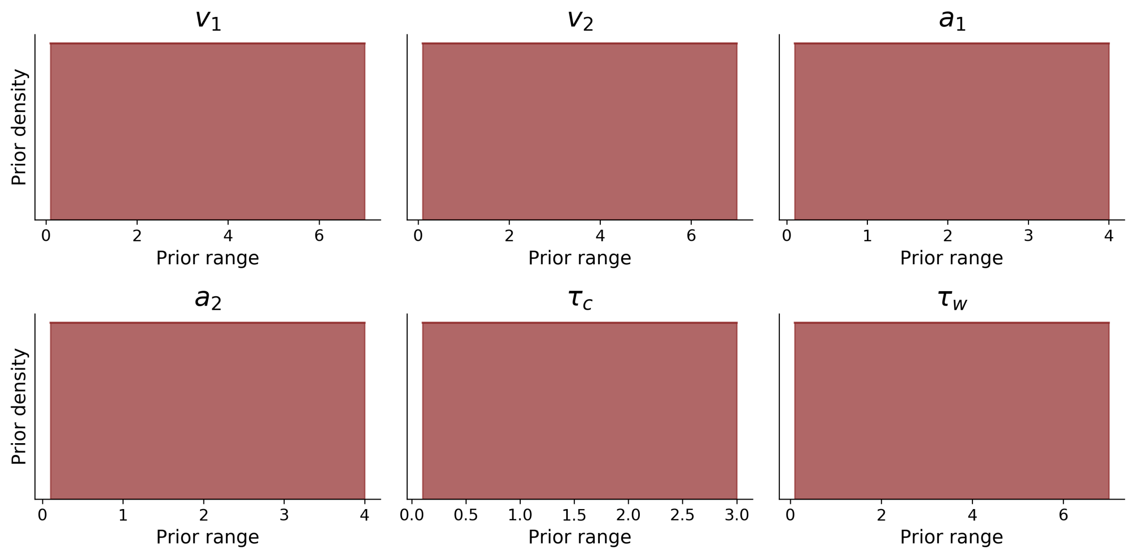
Prior predictive checks are designed to test whether a model is consistent with the relevant domain expertise. Typically, prior predictive checks are carried out either in i) prior space, or ii) in data space. In the latter case, more aptly termed a *prior pushforward check*, one obtains a random draw from the prior and simulates a synthetic data set using the sampled parameter configuration. We resort to the following plausibility considerations and visualizations to ensure consistency with domain expertise:

1. We place uniform priors over all diffusion model parameters. Concerning the choice of prior space, we select a broad prior domain which is sufficient to cover the range of plausible parameter values obtainable from human participants [2, 3, 4, 5]. The priors are visualized in Supplementary Figure 1. Our priors encode no initial beliefs in parameter differences between the experimental conditions (congruent/incongruent), so any differences emerging after estimation will be due to the information contained in the data. Moreover, the broader prior for non-decision times in incorrect trials compared to correct trials reflects the nature of the IAT task, as described in the main body of our paper.

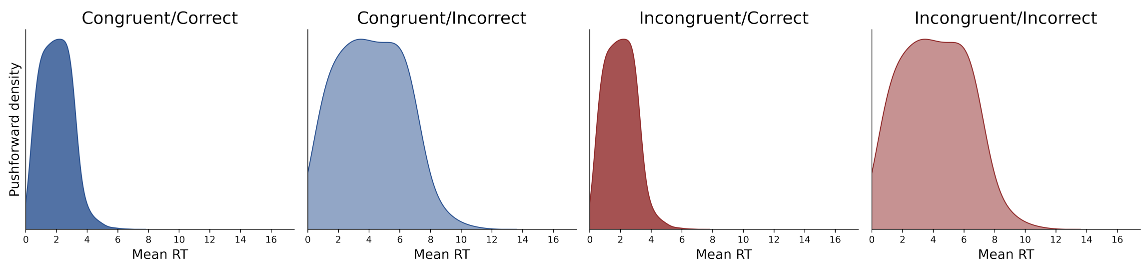
2. In order to explore the data space implied by our choice of priors, we perform a prior pushforward check with respect to mean response times. The resulting pushforward densities are depicted in Supplementary Figure 2 and comply with our expectation that incorrect trials are more likely to result in higher response times.

3. Further, we perform a prior pushforward check with respect to accuracy. The resulting pushforward densities are depicted in Supplementary Figure 3. These results imply a high expected accuracy prior to observing data (since the IAT is a relatively easy task) and no prior belief about differences in accuracies between conditions.

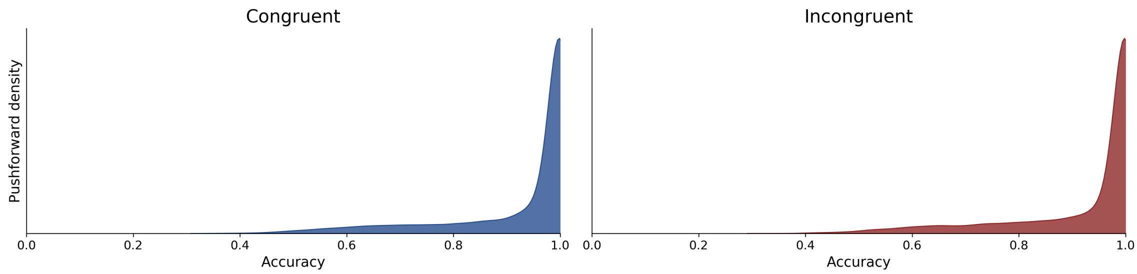
4. Finally, we perform prior pushforward checks with respect to higher moments of the implied RT distribution, namely, skewness and variance (standard deviation). The resulting pushforward densities are depicted in Supplementary Figures 5 and 4, respectively. These densities imply a positive expected skewness, as observed in numerous empirical RT studies. Moreover, our prior choice implies a moderate expected variability in individual RTs, which nevertheless is broad enough to include participants exhibiting very high variability in their responses. A larger variability is expected in incorrect responses.



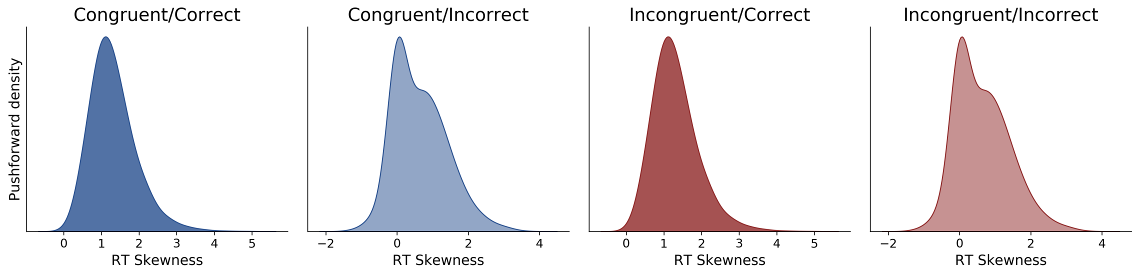
Supplementary Figure 1: Uniform prior distributions over diffusion model parameters.



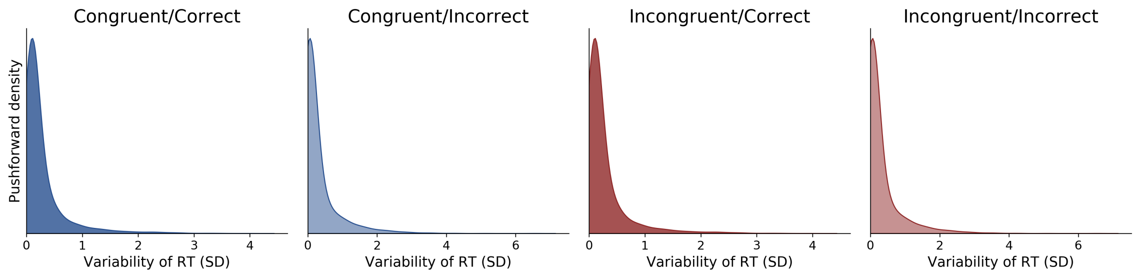
Supplementary Figure 2: Simulated distributions of mean response times (RTs), as implied by our prior specification.



Supplementary Figure 3: Simulated distributions of accuracies, as implied by our prior specification.



Supplementary Figure 4: Simulated distributions of skewness, as implied by our prior specification.



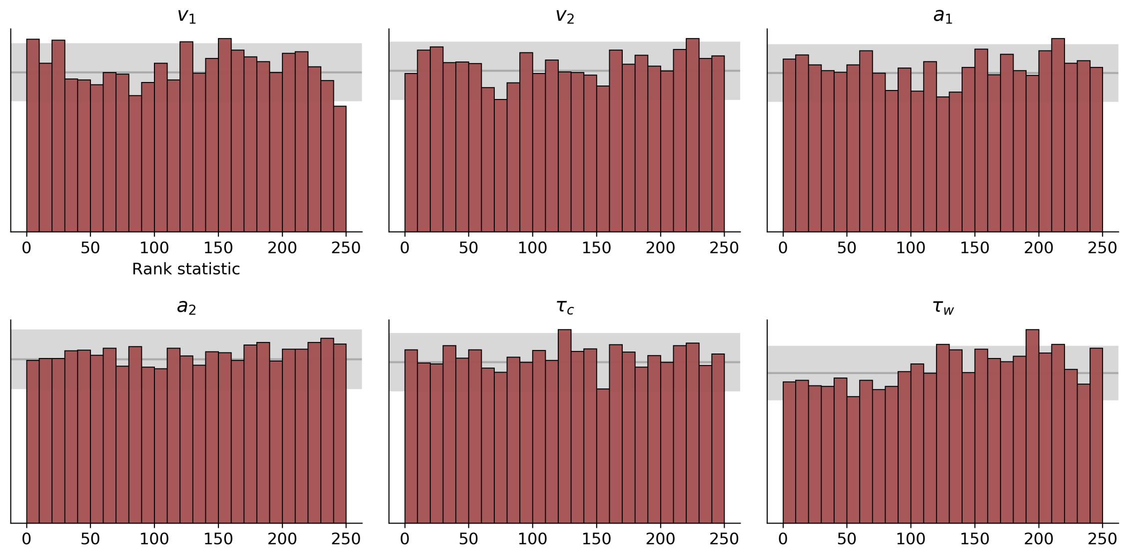
Supplementary Figure 5: Simulated distributions of variability (standard deviations from the mean RT), as implied by our prior specification.

Computational Faithfulness

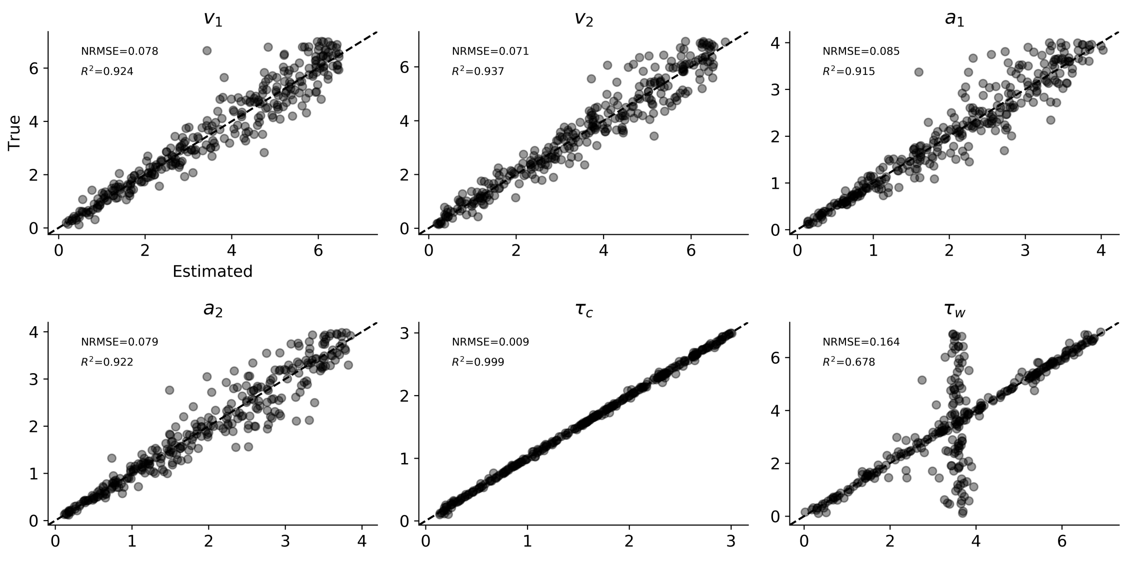
Computational faithfulness refers to the ability of a Bayesian method to recover the correct target posterior in a particular modeling scenario. Since BayesFlow enables fully amortized inference, we can efficiently compute simulation-based calibration (SBC, [8]), which allows us to visually detect potential biases in posterior estimation. After the training phase of BayesFlow, we performed 5000 simulations from the diffusion model and obtained 250 posterior samples for each simulated data set. SBC histograms of the resulting rank statistics are depicted in Supplementary Figure 6 and indicate no apparent global biases in posterior location and dispersion. The marginal histograms for non-decision times in incorrect trials imply a slight underestimation of the *true* parameter values by the posterior means. This small distortion is probably caused by simulated data sets with zero incorrect trials in both conditions, which render estimation of τ_w impossible.

Model Adequacy/Sensitivity

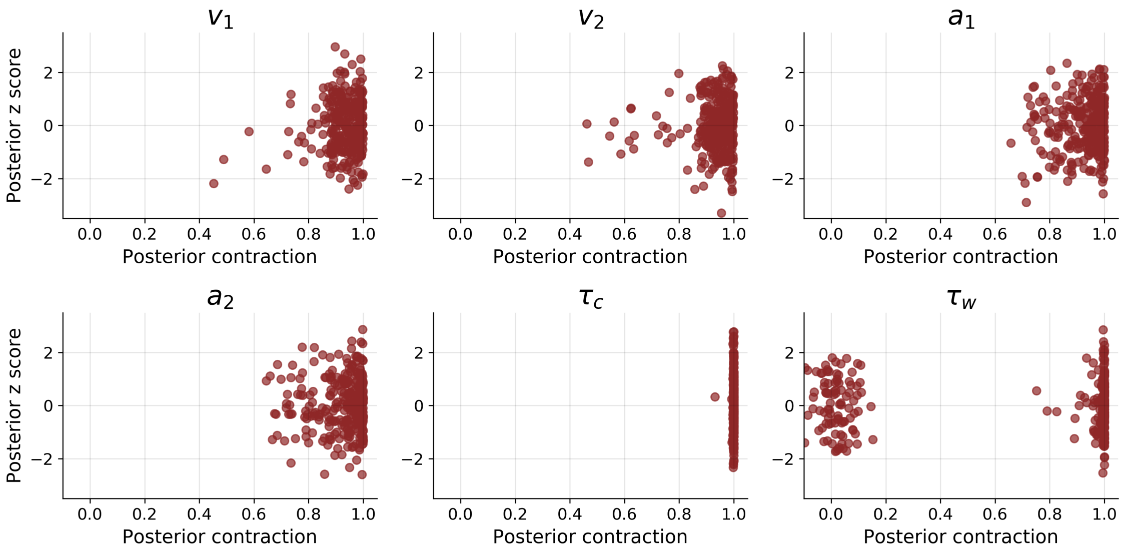
Model sensitivity asks whether the parameters of a model can be recovered given the model's prior specification, generative scope, and particular algorithmic from. To evaluate model sensitivity, we first perform a simulation-based recovery study and plot the known true parameters vs. the corresponding posterior means (as summaries of the full posteriors). The recovery result obtained from 300 simulations are depicted in Supplementary Figure 7. The plots indicate very good point-estimate recovery, with R^2 metrics ranging from 0.999 (non-decision time in correct trials) to 0.678 (non-decision time in incorrect trials). The low R^2 for the non-decision time parameter in incorrect trials is due to simulated data sets having zero or very few errors. These data sets thus provide no information for Bayesian updating and BayesFlow returns the prior. Second, we compute posterior contraction and posterior z-score on 300 simulated data sets, and visualize these on a 2D Cartesian plane [1]. Accordingly, an adequate model exhibits high posterior contraction and a posterior z-score symmetrically distributed around 0. Indeed, such behavior is depicted in Supplementary Figure 8, which also represents the estimation problems for the non-decision time parameter in incorrect trials posed by data sets with zero or very few errors.



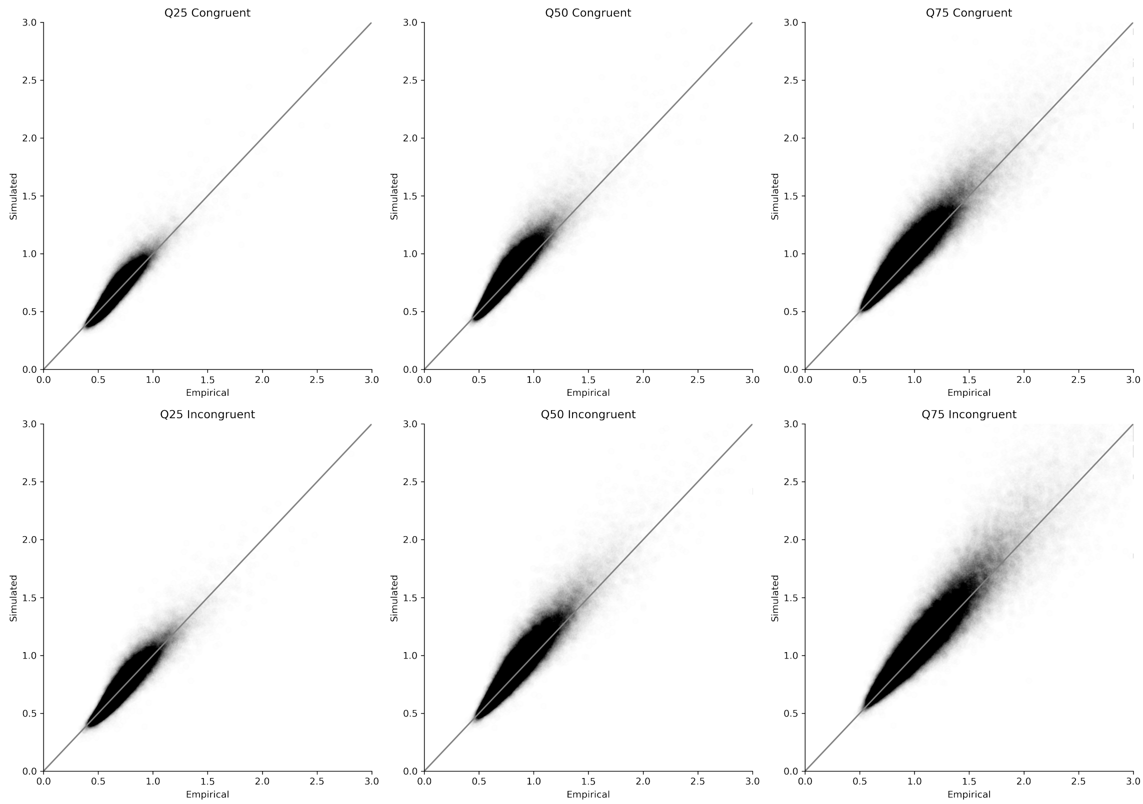
Supplementary Figure 6: SBC histograms for each marginal parameter posterior. Horizontal gray bars depict confidence intervals for a binomial distribution, as recommended by [8].



Supplementary Figure 7: True parameter values vs. estimated posterior means. The plots indicate excellent recovery of the parameters, with the vertical line in the last plot indicating unrecoverable non-decision time parameters for data sets with zero or very few errors.



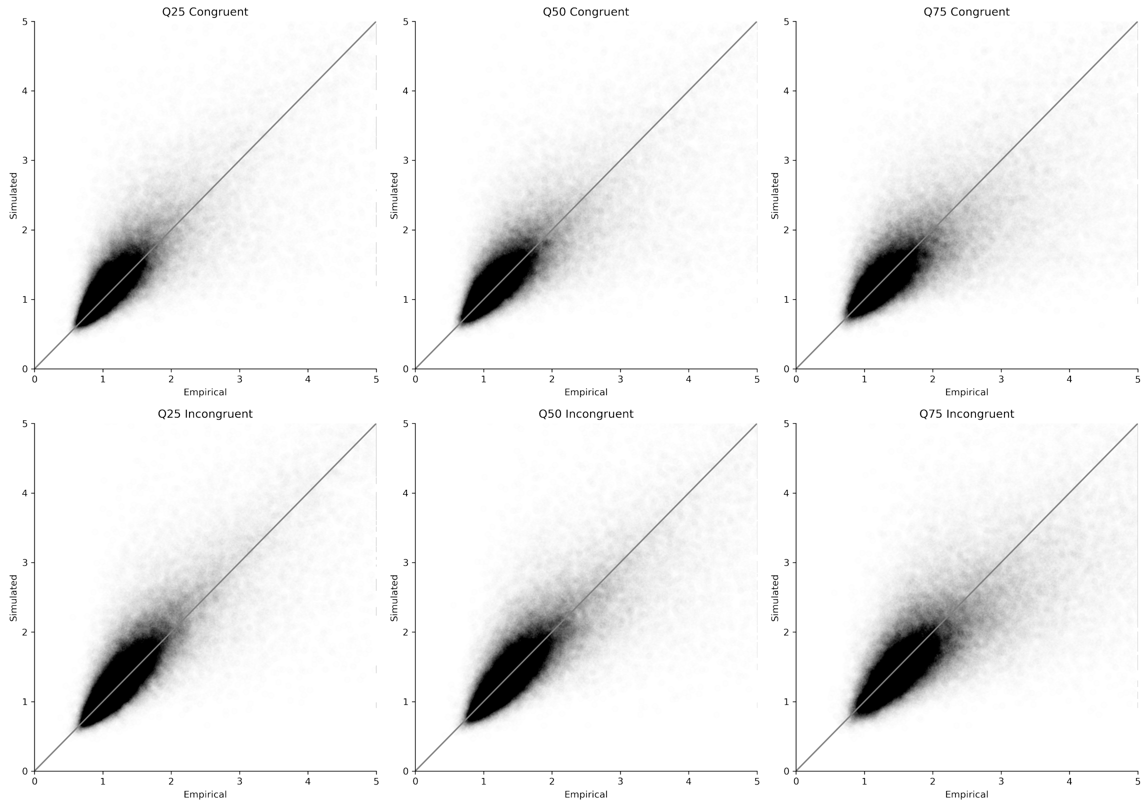
Supplementary Figure 8: Posterior z-scores and posterior contraction for each marginal posterior obtained from simulations from the Bayesian model. The plots indicate high posterior contraction and a z-score centered around 0, but also no contraction for τ_w in data sets with zero or very few errors.



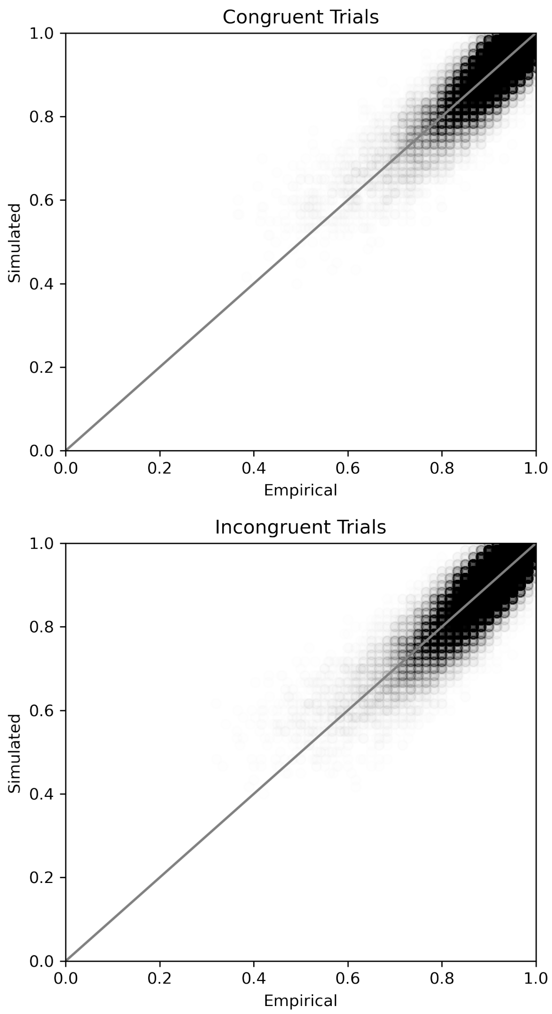
Supplementary Figure 9: Posterior predictive checks for correct response times. The plot shows the quantiles of the correct response RTs for empirical and simulated data plotted against each other in a scatter plot, separated by experimental condition. The opacity level in the graph is very low because of the great number of plotted points.

Posterior Predictive Checks

Finally, posterior predictive checks assess whether the model captures the relevant structure of the assumed true data generating process. We randomly selected $100,000$ cases from the IAT sample to simulate response times and accuracy data based on the diffusion model parameter posterior means. Due to the large sample size studied, it was sufficient to use posterior means as representatives of the parameter values. 60 trials each were generated for both experimental conditions per person, as is also the case in the empirical data. Supplementary Figures 9, 10, and 11 show scatter plots of the empirical response time quartiles and accuracy rates plotted against the respective empirical data for both experimental conditions. Model fit is good on average, although simulated slow error response times show some larger deviations from a perfect correlation, explainable by unstable summary statistics for both empirical and simulated data due to the low number of error trials available for analysis.



Supplementary Figure 10: Posterior predictive checks for error response times. The plot shows the quantiles of the error response RTs for empirical and simulated data plotted against each other in a scatter plot, separated by experimental condition. As accuracy generally was high (median accuracy 95%), error quantiles are based on very low numbers of trials and thus unstable. The opacity level in the graph is very low because of the great number of plotted points.



Supplementary Figure 11: Posterior predictive checks for accuracy rates. The plot shows the proportions of correct responses for empirical and simulated data plotted against each other in a scatter plot, separated by experimental condition.

. The opacity level in the graph is very low because of the great number of plotted points.

Additional Analyses

In Supplementary Figure 12 we show our main analyses for mean correct RTs, drift rates, and boundary separations for the congruent experimental condition, as well as for error non-decision times. Results are very similar to those obtained for the incongruent condition. However, for boundary separation, the second change point was estimated to lie already at age 40, with the subsequent age-related increase being less pronounced compared to the incongruent condition.

Change points were estimated at: Mean RT change point 1 posterior mean = 17.5, 95% HDI = [17.1 – 18.0]); Mean RT change point 2 posterior mean = 62.1, 95% HDI = [60.1 – 64.2]); Drift rate change point 1 posterior mean = 22.0, 95% HDI = [20.9 – 23.3]); Drift rate change point 2 posterior mean = 56.8, 95% HDI = [53.8 – 59.8]); Boundary separation change point 1 posterior mean = 17.5, 95% HDI = [16.5 – 18.4]); Boundary separation change point 2 posterior mean = 40.2, 95% HDI = [37.4 – 43.0]); Error non-decision times change point posterior mean = 14.6, 95% HDI = [12.0 – 17.0].

Change per year as predicted by Bayesian ridge regression was: Mean RTs before first change point $b = -0.027$, 95% HDI = $[-0.028 - -0.026]$; Mean RTs before second change point $b = 0.005$, 95% HDI = $[0.005 - 0.005]$; Mean RTs after second change point $b = 0.010$, 95% HDI = $[0.009 - -0.011]$; Drift rates before first change point $b = 0.065$, 95% HDI = $[0.064 - 0.066]$; Drift rates before second change point $b = 0.005$, 95% HDI = $[0.005 - 0.006]$; Drift rates after second change point $b = -0.015$, 95% HDI = $[-0.017 - -0.014]$; Boundary separations before first change point $b = -0.026$, 95% HDI = $[-0.028 - -0.024]$; Boundary separations before second change point $b = 0.016$, 95% HDI = $[0.015-0.016]$; Boundary separations after second change point $b = 0.005$, 95% HDI = $[0.004-0.005]$; Non-decision times before first change point $b = -0.020$, 95% HDI = $[-0.030 - -0.010]$; Non-decision times after first change point $b = 0.020$, 95% HDI = $[0.020 - 0.020]$.

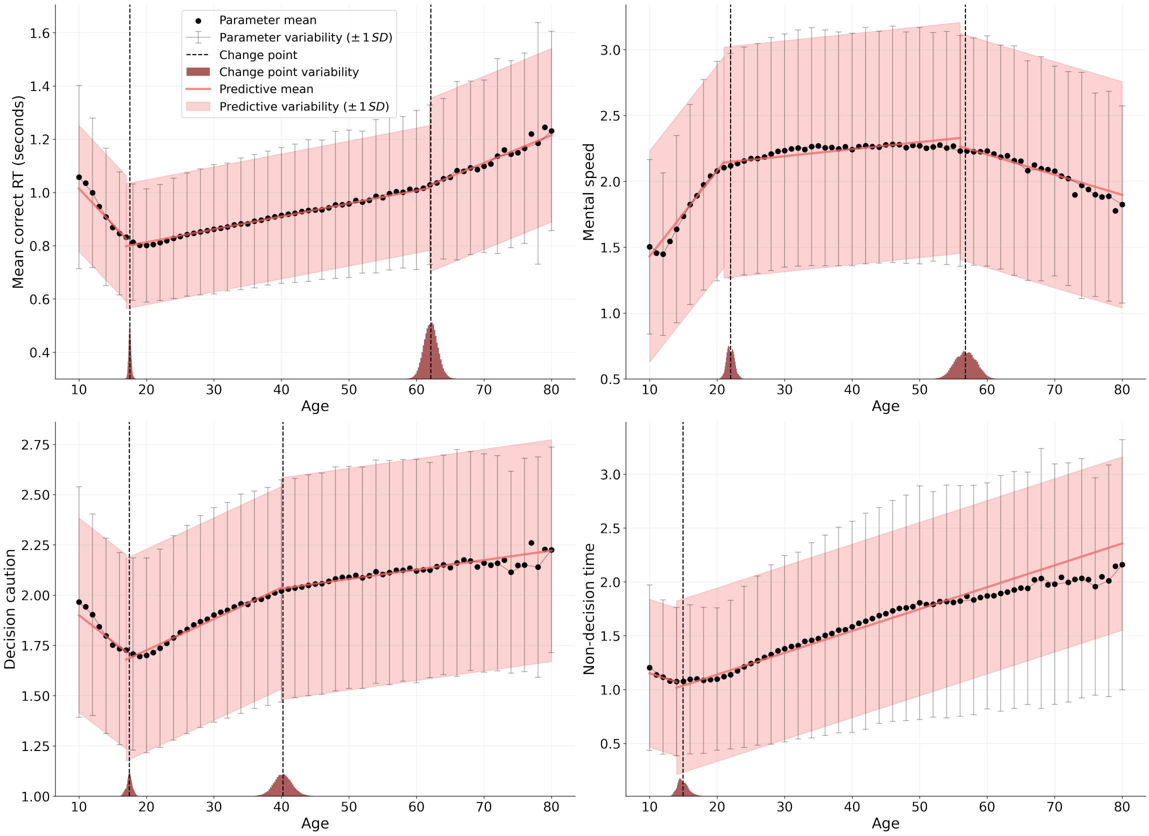
Supplementary Figures 13, 14, 15, and 16 show different robustness checks. Our main results on age trends in drift rates were consistent across sub-samples (13, 14), participant countries of origin (15), and different experimental stimuli (16), all in regard to both experimental conditions. Across all robustness checks, the age-related decline in drifts rates after age 60 was less pronounced in the congruent compared to the incongruent experimental condition.

Supplementary Figures 17, 18, and 19 show mean accuracy rates and error response times in the two experimental conditions plotted against years of age.

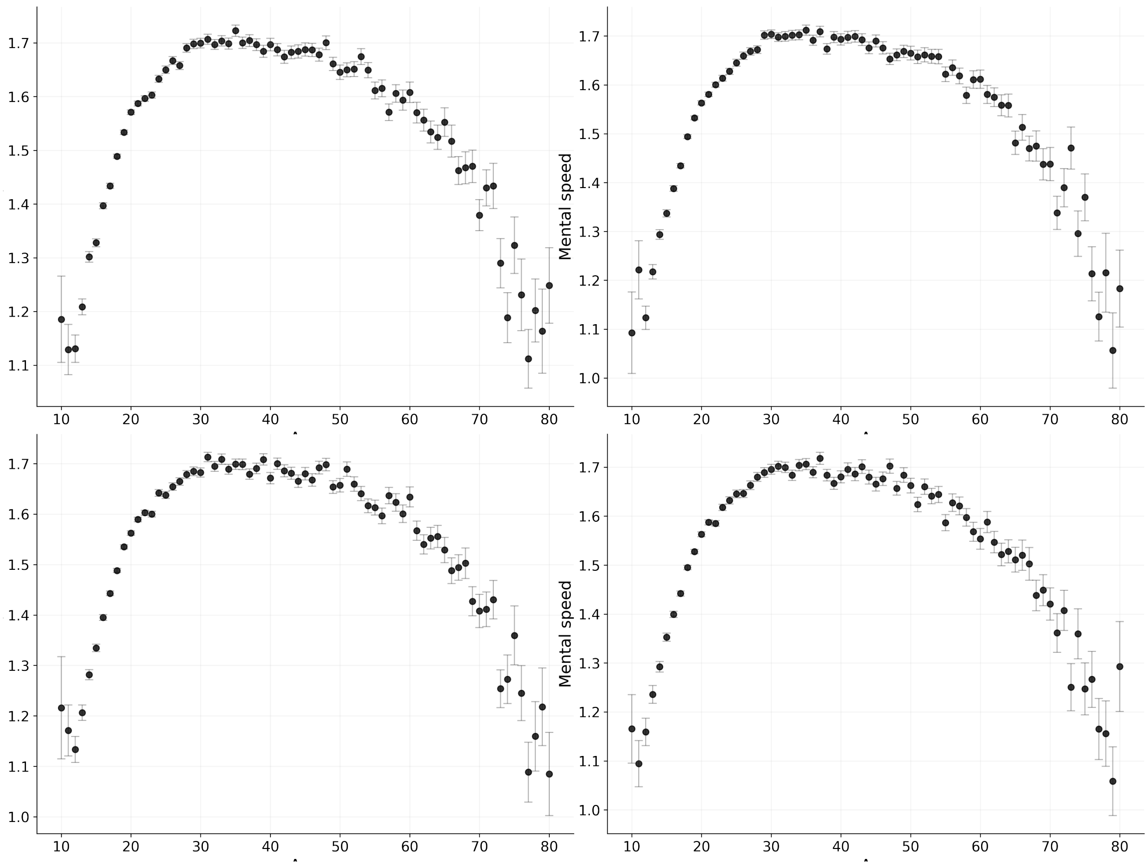
Supplementary Figure 20 once more shows the main diffusion model parameters plotted against years of age. For this analysis, we used a diffusion model integrating so-called collapsing boundaries [6] [7] to test for the robustness of our results under a different model specification. Age-related patterns of the main model parameters obtained on the basis of the collapsing bounds model are essentially the same as those obtained from the more parsimonious basic diffusion model. The full Bayesian workflow and code for the collapsing boundary model can be accessed on the project's GitHub page.

In Supplementary Table 1 we report across-person correlations of the diffusion model parameters, while in Supplementary Table 2 we present the summarized within-person correlations based on the individual joined posterior distributions.

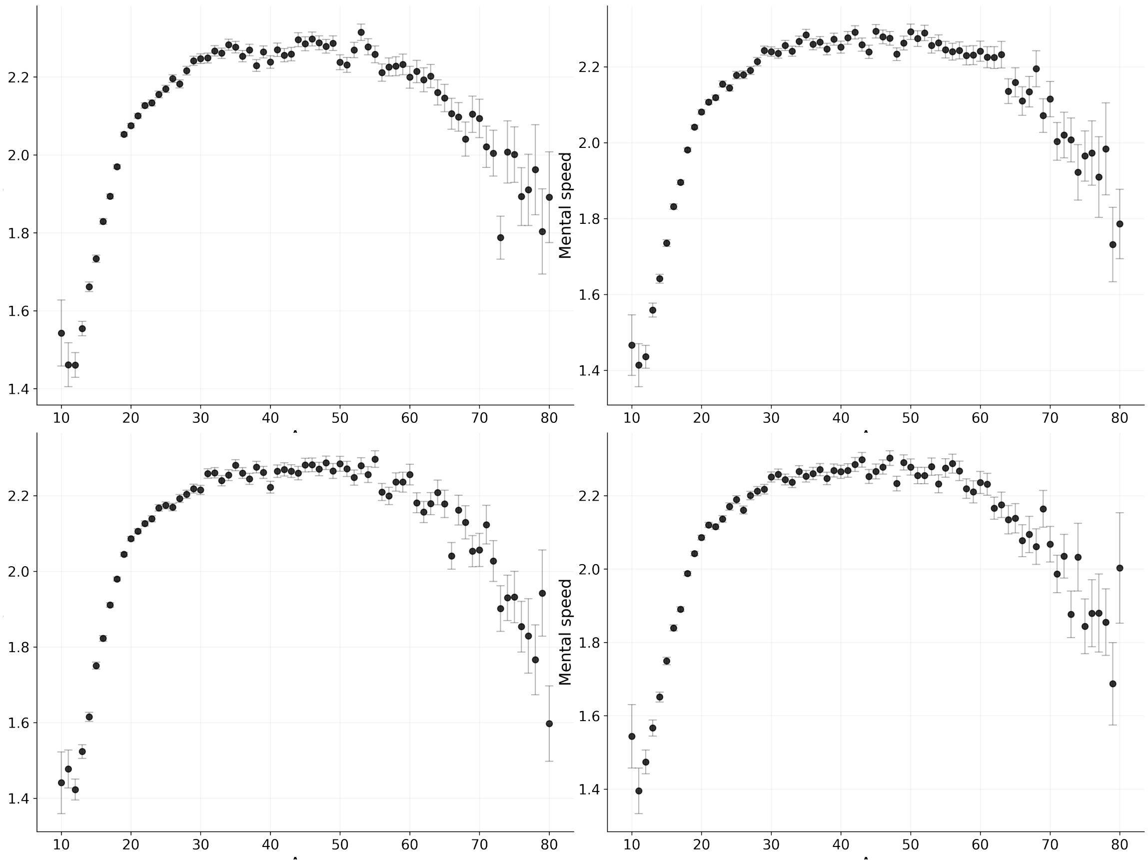
Supplementary Table 3 shows the correlations of the diffusion model parameter estimates obtained via BayesFlow with parameters obtained via standard MCMC sampling. We applied HMC-MCMC (as implemented in the Stan programming language [9]; 1000 posterior samples + 1000 burn-in samples, 4 chains) to fit a corresponding diffusion model to a random sub-group of 6400 participants (200 from each of the 32 time-ordered original data files). We provide the complete code for reproducing these results on our GitHub page. Correlations across estimation methods were very high (>= .90 for all parameters), once more emphasizing the robustness of our Bayesian workflow.



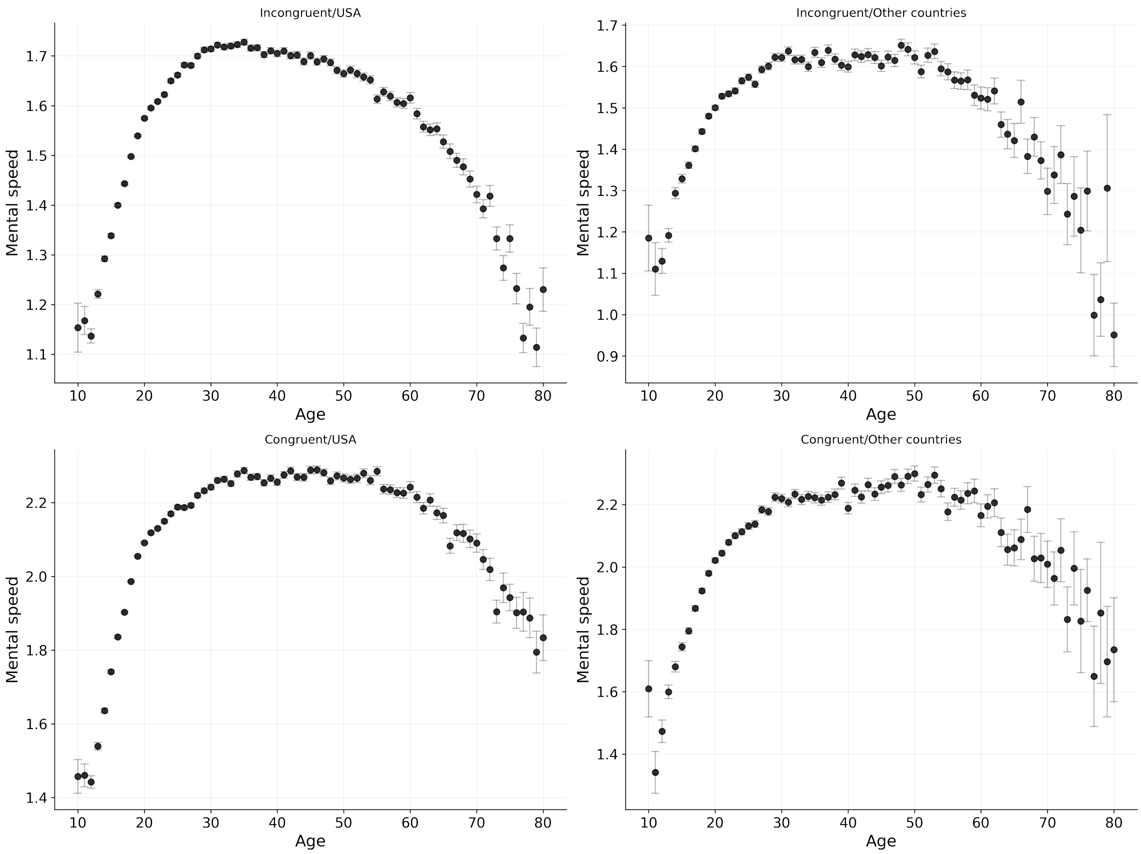
Supplementary Figure 12: Results for the congruent experimental condition and error non-decision times. Mean response times (RTs) and diffusion model parameters as a function of age. Black points indicate means computed separately for each year of age. Bars indicate standard deviations (only shown for every second year for better clarity). Red lines denote the Bayesian piece-wise ridge regression model's mean predictions, which describe the observed means fairly well. The shaded red region denotes the uncertainty (standard deviation) of the piece-wise model's predictions. The dashed lines indicate the mean change points estimated from the per-age-group averaged data, with the full posterior distributions (scaled for readability) of the change points shown at the bottom of each plot. Trends are very similar to the ones found for the incongruent condition for mean correct RTs and drift rates; the same holds true for error non-decision times in relation to correct non-decision times. Boundary separation values show an earlier second change point and less pronounced increasing trend in old age than is found for the incongruent condition.



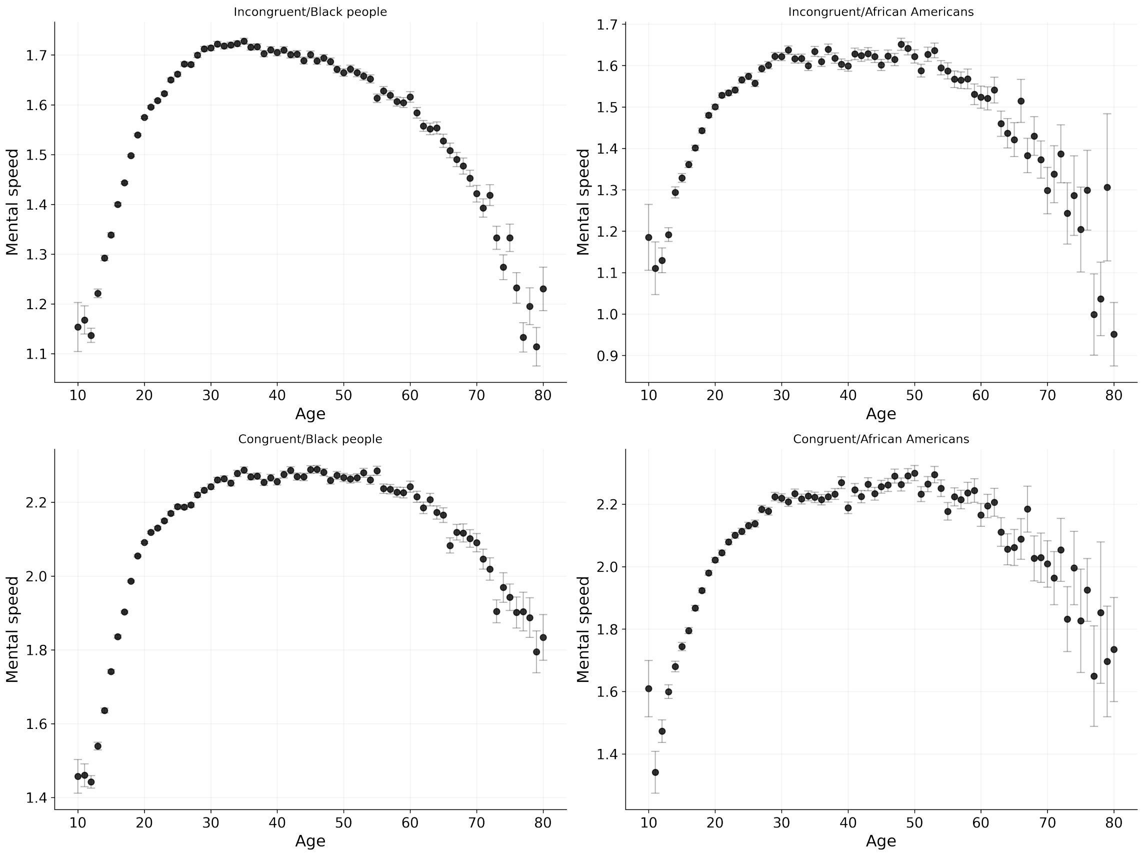
Supplementary Figure 13: Robustness check I. We randomly divided our sample in four sub-groups, with three groups containing 300,000 participants each, and one containing the rest. The figure shows the trends in drift rates for the incongruent condition. Trends are very similar across the sub-groups. Dots indicate means per year of age, while the bars indicate the standard errors of the means.



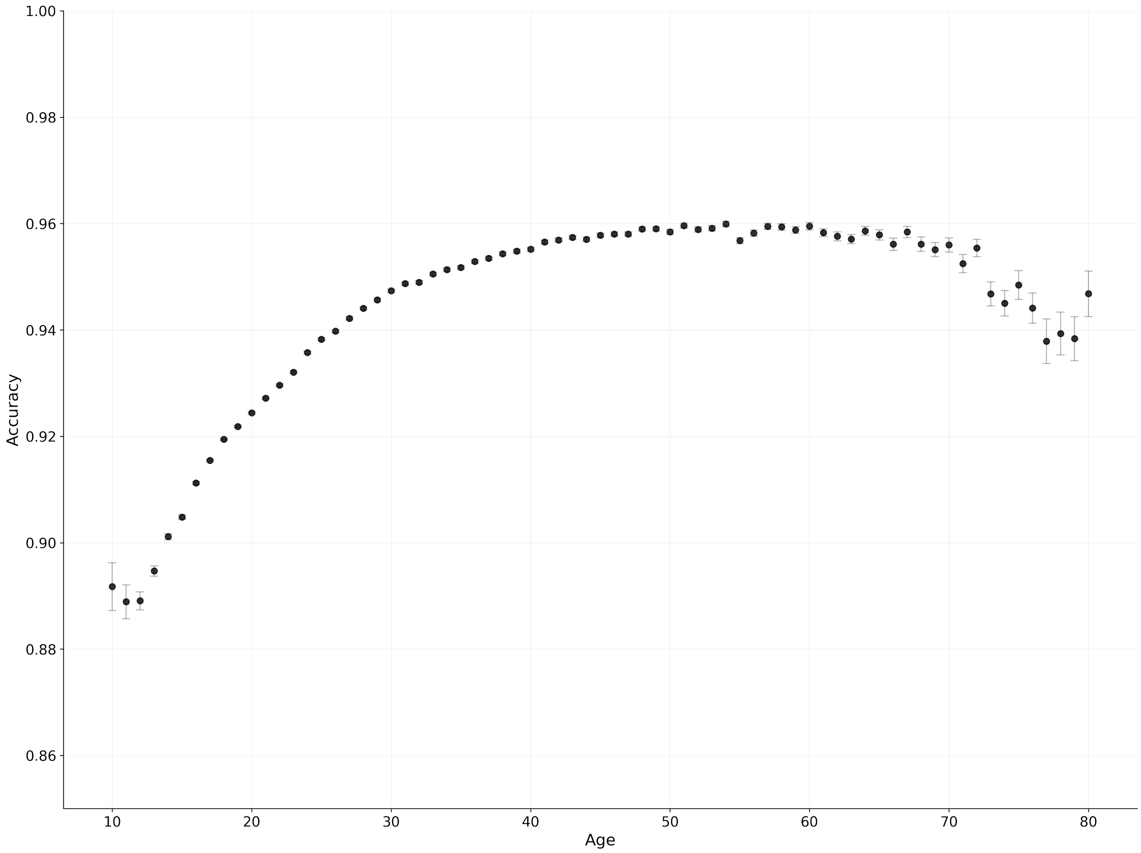
Supplementary Figure 14: Robustness check II. We randomly divided our sample in four sub-groups, with three groups containing 300,000 participants each, and one containing the rest. The figure shows the trends in drift rates for the congruent condition. Trends are very similar across the sub-groups. Notably, the decrease in drift rates after age 60 is less pronounced than for the incongruent condition. Dots indicate means per year of age, while the bars indicate the standard errors of the means.



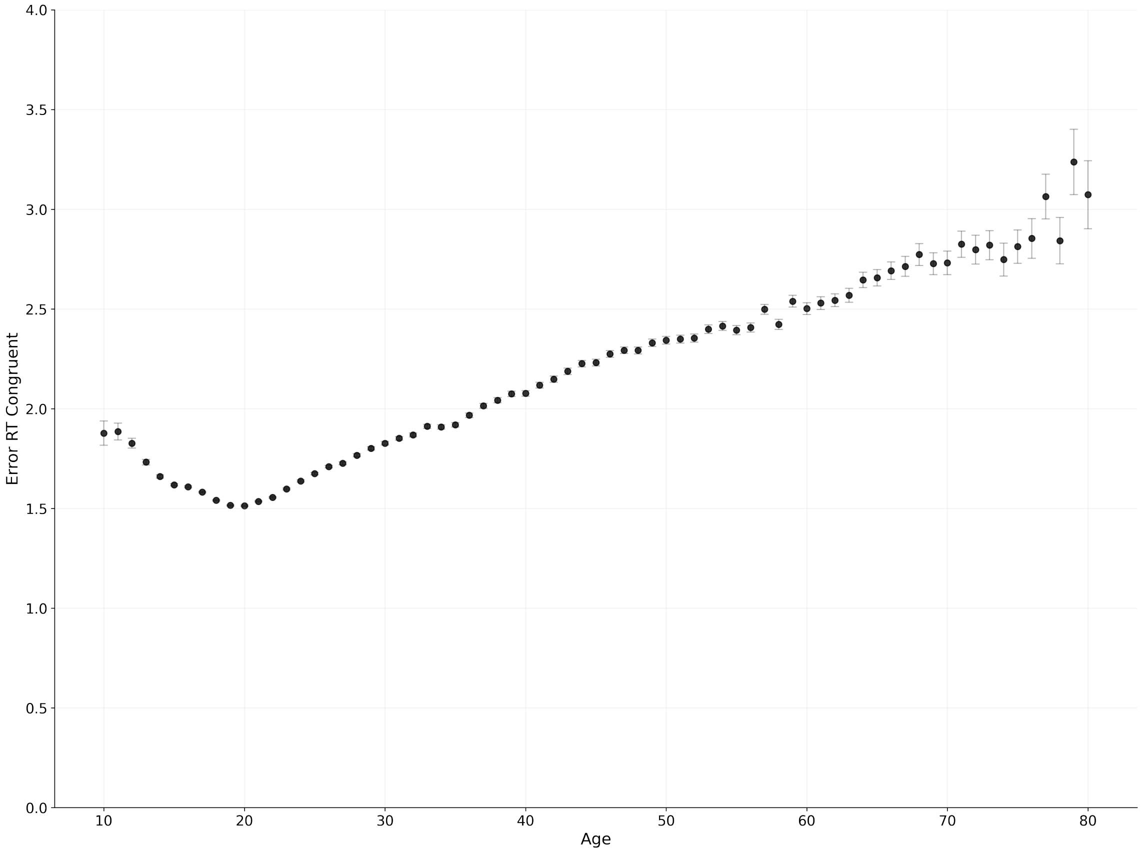
Supplementary Figure 15: Robustness check III. We compared the age trends in mental speed between participants indicating that they were born in the United States vs. all other countries. Trends were similar both in the congruent and the incongruent experimental conditions. Dots indicate means per year of age, while the bars indicate the standard errors of the means.



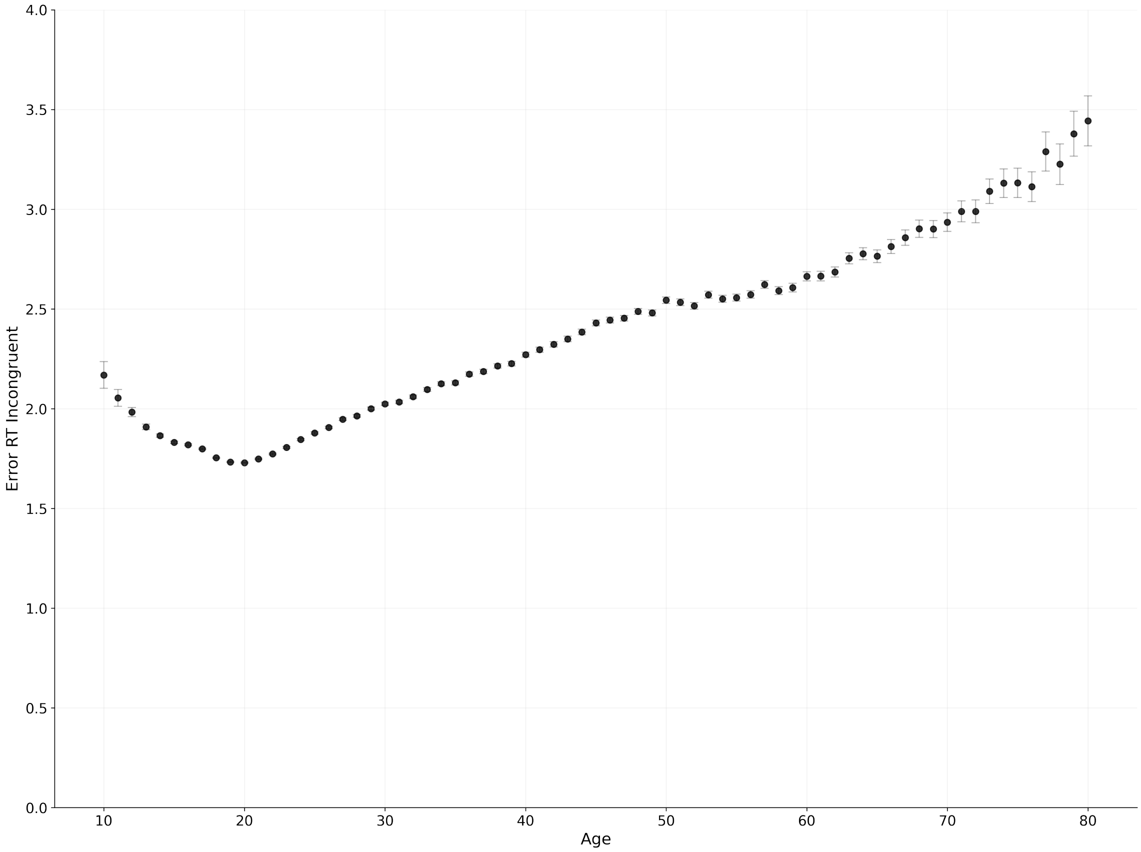
Supplementary Figure 16: Robustness check IV. We compared the age trends in mental speed between experimental sessions where "Black/White" were the stimuli and those that used "African American/European American". Trends were similar both in the congruent and the incongruent experimental conditions. Dots indicate means per year of age, while the bars indicate the standard errors of the means.



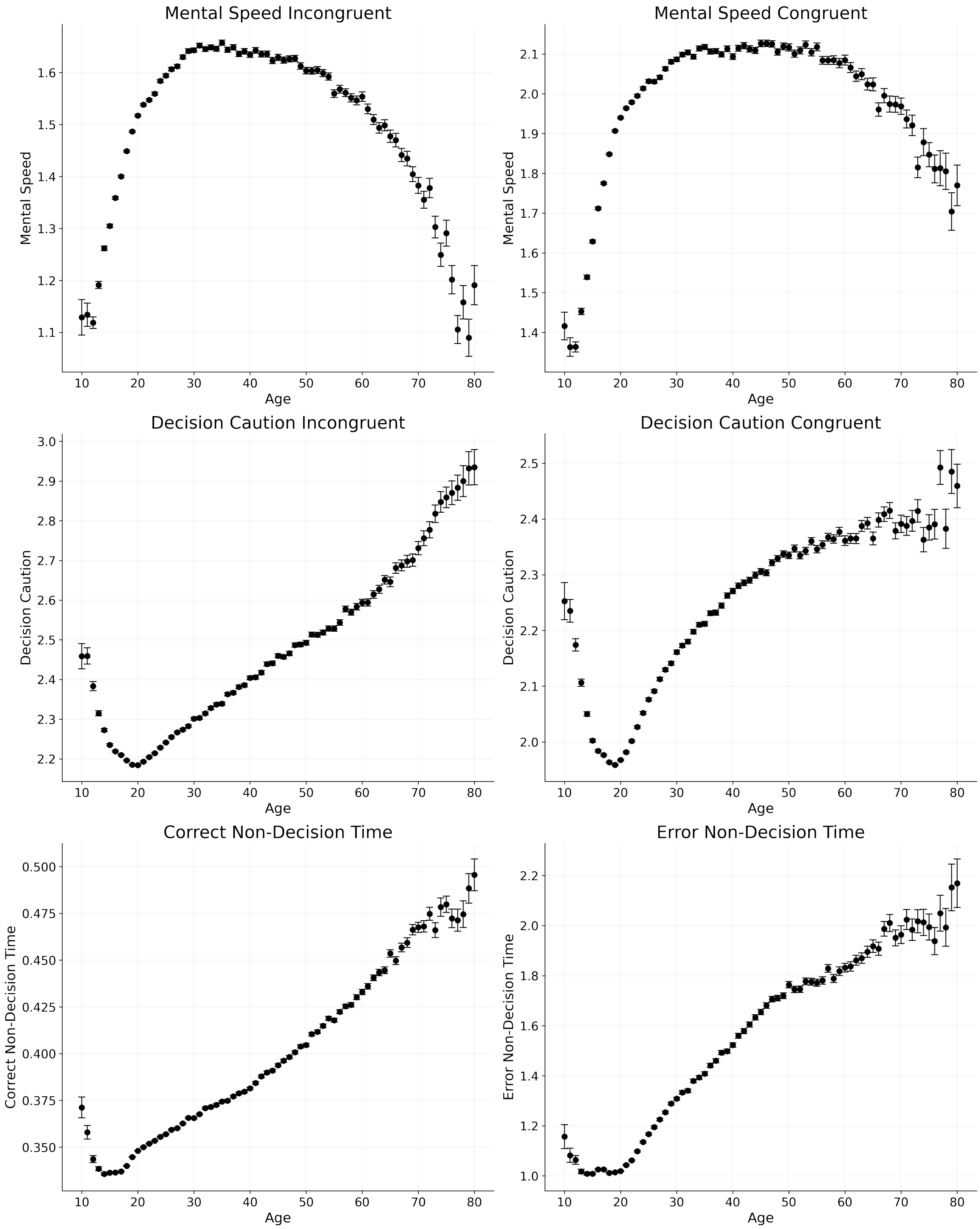
Supplementary Figure 17: Accuracy in the two experimental conditions in percent plotted against years of age. Dots indicate across-person means per year of age, while the bars indicate the standard errors of the means.



Supplementary Figure 18: Means of error response times in the congruent experimental condition plotted against years of age. Note that error response times include the time needed to self-correct to the correct response. Dots indicate across-person means per year of age, while the bars indicate the standard errors of the means.



Supplementary Figure 19: Means of error response times in the incongruent experimental condition plotted against years of age. Note that error response times include the time needed to self-correct to the correct response. Dots indicate across-person means per year of age, while the bars indicate the standard errors of the means.



Supplementary Figure 20: Diffusion model parameter estimates originating from a diffusion model including collapsing boundaries, plotted against years of age. The observed age patterns are virtually identical to our main results, supporting their robustness also across different forms of model specifications. Dots indicate across-person means per year of age, while the bars indicate the standard errors of the means.

Supplementary Table 1: Across-person correlations of diffusion model parameters

	$\nu_{incongruent}$	$\nu_{incongruent}$	$a_{incongruent}$	$a_{congruent}$	$\tau_{correct}$
$\nu_{congruent}$	0.32				
$a_{incongruent}$	-0.19	-0.23			
$a_{congruent}$	-0.11	0.04	0.56		
$\tau_{correct}$	0.15	0.17	-0.03	-0.15	
τ_{error}	0.10	0.04	0.39	0.37	0.19

Note. ν = drift rate; a = boundary separation; τ = non-decision time. Correlation estimates based on the entire sample. While drift rates and boundary separations show the strongest correlations with the respective parameter from the other experimental condition, error non-decision times are related to boundary separations, highlighting the distinct interpretation of τ_{error} in relation to $\tau_{correct}$

Supplementary Table 2: Within-person correlations of diffusion model parameters

	$\nu_{incongruent}$	$\nu_{congruent}$	$a_{incongruent}$	$a_{congruent}$	$\tau_{correct}$
$\nu_{congruent}$	-0.02				
$a_{incongruent}$	0.37	0.07			
$a_{congruent}$	0.01	0.53	0.38		
$\tau_{correct}$	-0.05	-0.19	-0.62	-0.66	
τ_{error}	0.08	0.07	-0.20	-0.05	0.13

Note. ν = drift rate; a = boundary separation; τ = non-decision time. Person-specific correlations were computed based on the individual joined posterior distributions of diffusion model parameters. Individual correlations were then Fisher-z transformed, averaged, and transformed back to a correlation estimate.

Supplementary Table 3: Correlations of diffusion model parameters obtained via BayesFlow vs. Stan

Parameter	Correlation
Mental Speed (incongruent)	.97
Mental Speed (congruent)	.96
Decision Caution (incongruent)	.93
Decision Caution (congruent)	.96
Non-decision Time (correct)	.91
Non-Decision Time (error)	.90

Note. Mental speed = drift rate; decision caution = boundary separation. Corresponding models were fit using BayesFlow and standard HC-MCMC sampling as implemented in Stan (1000 samples, 1000 burn-in samples, 4 chains). All cases including divergent transitions in MCMC sampling were discarded.

References

[1] Schad, D. J., Betancourt, M. & Vasishth, S. Toward a principled bayesian workflow in cognitive science. *Psychological Methods* **26**, 103–126 (2020).

[2] Matzke, D. & Wagenmakers, E.-J. Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review* **16**, 798–817 (2009).

[3] Voss, A., Nagler, M. & Lerche, V. Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology* **60**, 385–402 (2013).

[4] Lerche, V., Voss, A. & Nagler, M. How many trials are required for parameter estimation in diffusion modeling? a comparison of different optimization criteria. *Behavior Research Methods* **49**, 513–537 (2017).

[5] Lerche, V. & Voss, A. Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research* **83**, 1194–1209 (2019).

[6] Voskuilen, C., Ratcliff, R. & Smith, P. L. Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of mathematical psychology* **73**, 59–79 (2016).

[7] Drugowitsch, J., Moreno-Bote, R., Churchland, A.K., Shadlen, M.N. & Pouget, A. The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience* **32**, 3612–3628 (2012).

[8] Talts, S., Betancourt, M., Simpson, D., Vehtari, A. & Gelman, A. Validating Bayesian inference algorithms with simulation-based calibration. *arXiv preprint* (2018).

[9] Stan Development Team. RStan: the R interface to Stan (2020). URL `http://mc-stan.org/`. R package version 2.21.2.