Bayesian Inference of Dynamical Cognitive and Oculomotor Processes in the SWIFT Model of Reading

Maximilian M. Rabe, Johan Chandra, André Krügel, Stefan A. Seelig, Shravan Vasishth, & Ralf Engbert

Research Focus Cognitive Sciences, University of Potsdam, Germany

Universität Potsdam

SFB 1287 Limits of Variability in Language
Cognitive, Grammatical, and Social Aspects

Gefördert durch DFG Deutsche Forschungsgemeinschaft

https://doi.org/10.17605/osf.io/mxq4n

Introduction

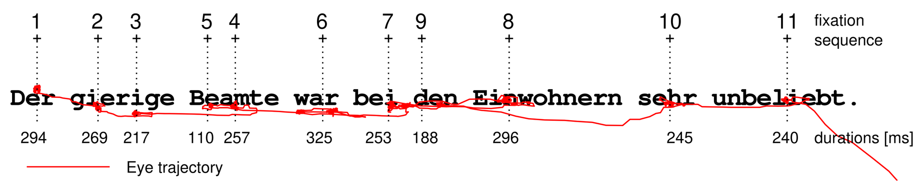

Figure 1. In a typical eye trajectory in reading, about half of all saccades move the foveal part of the visual field from word n to the next word $n+1$. Other saccade types generate refixations (e.g. 2–3), word skipping (e.g. 7–8), and regressions (8–9).

SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) is a dynamical cognitive model of eye-movement control in reading. Words within the processing span centered around the current fixation position are processed in parallel via a temporally evolving activation field (see Figure 2).

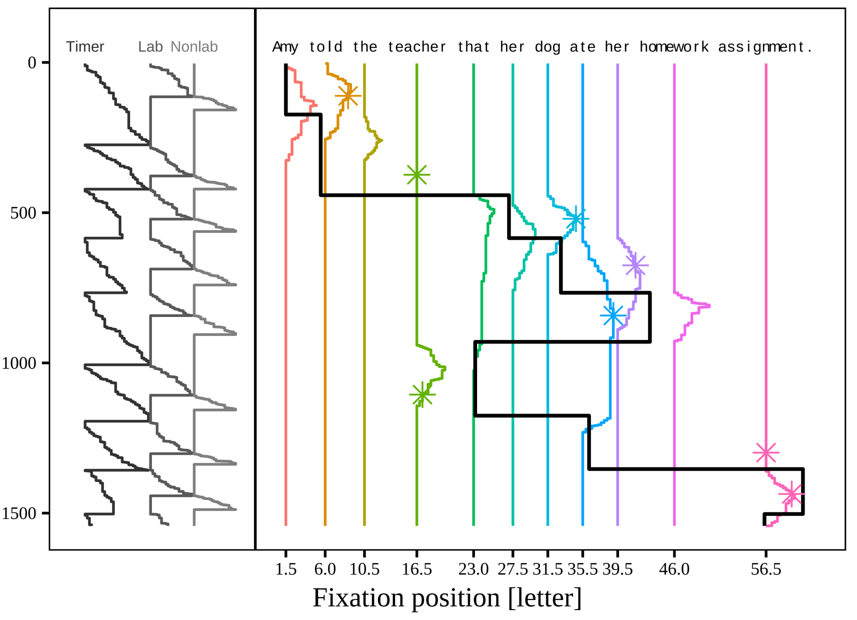

Figure 2. Simulated eye trajectory (solid black line) in the SWIFT model. Thin lines are word activations (colored) and timers (gray) as a function of time. Asterisks mark points in time when a saccade program is executed.

Recently, we implemented SWIFT for Bayesian parameter estimation (Seelig et al., 2019). We fitted the model to a diverse reading dataset in order to evaluate the goodness of fit with various temporal and spatial summary statistics.

For the simulations reported here, we replaced the standard Gaussian saccade error model (McConkie, Kerr, Reddix, & Zola, 1988) with Gamma-distributed saccade amplitudes:

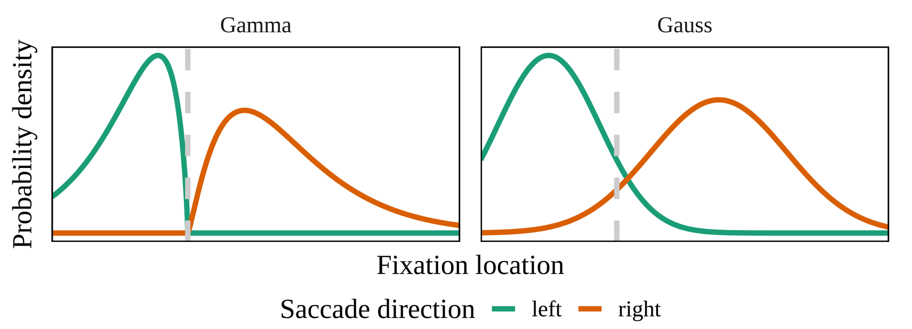

Figure 3. Comparison of saccade amplitude distributions for saccade targets left or right to the current fixation location (dashed gray line).

Bayesian parameter estimation

Bayesian model fitting allows us to infer rigorous credibility intervals for model parameters. Additionally, we can determine model parameters *for individual participants*, which was often precluded in previous methods.

In SWIFT, the likelihood of a fixation $f_i = (k_i, l_i, T_i, s_i)$ on letter l_i of word k_i for T_i, is given as the combined spatial and temporal likelihoods, both conditional on all preceding fixations $F_{i-1} = \{f_1, ..., f_{i-1}\}$, model parameters θ and internal degrees of freedom ξ due to model stochasticity:

$$P_M(f_i|F_{i-1},\theta,\xi) = P_{temp}(T_i, s_i|k_i, l_i, F_{i-1}, \theta, \xi) \cdot P_{spat}(k_i, l_i|F_{i-1}, \theta, \xi)$$

While P_{spat} is exact and only depends on the initial state of the system (ξ), P_{temp} must be approximated by simulation.

Computational modelling

- SWIFT has many parameters with possibly multimodal distributions and stochastic likelihood
- Differential evolution Metropolis algorithm DREAM$_{ZS}$ (Vrugt et al., 2009) has been shown to work very reliably with complex models
- Implementation based on *PyDREAM* (Shockley, 2019)
- Customized to enable reevaluation of stochastic likelihood of previously accepted proposal for each MCMC iteration
- Model is fitted on 70% of each subject's data
- Predictive checks are carried out on the unseen 30%

Computational faithfulness

Likelihood profiles for selected parameters show that for simulated data, (1) the true parameter values are most likely and (2) model parameters have different selective influences on spatial and temporal likelihood components:

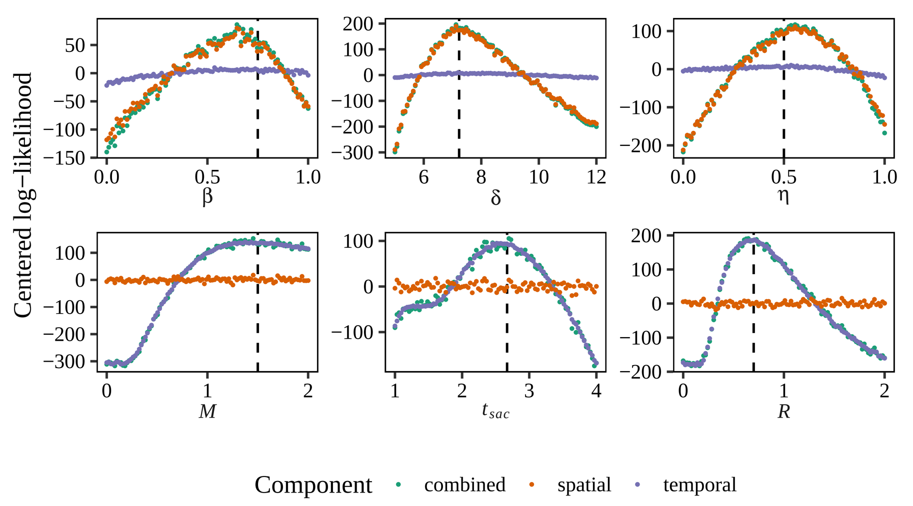

Figure 4. Centered log-likelihood profiles of a subset of model parameters with temporal and spatial components. Vertical dashed lines are true parameter values.

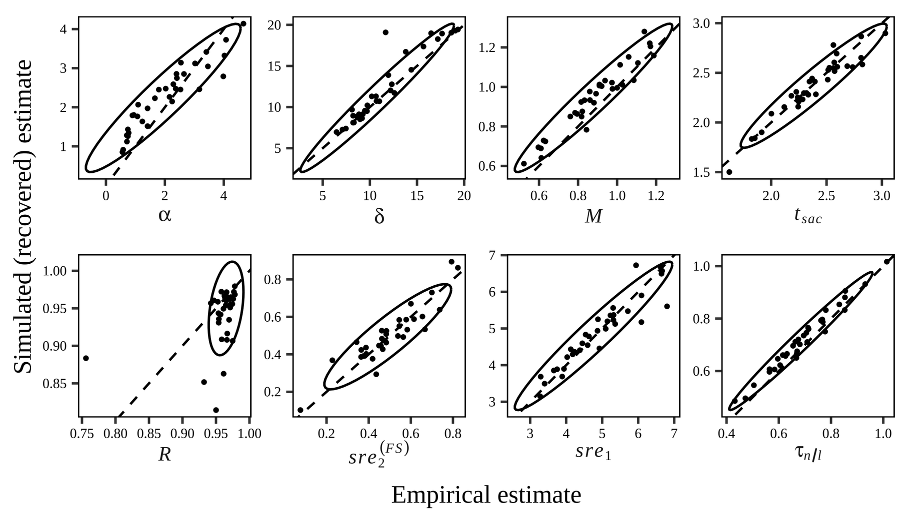

Figure 5. Parameter recovery study for 72 simulated datasets and a subset of model parameters. Recovered estimates are MAP estimators.

Experimental method

Table 1. Reading conditions

N	Jede	Sprache	der	Welt	besitzt	eine	Grammatik
mL	lɘbɘ	Ƨqɿɒɔʜɘ	bɘɿ	Wɘlʇ	dɘƨiʇƨʇ	ɘiuɘ	Ꭾɿɒmmɒʇiʞ
sL	Jdee	Scrahpe	der	Wlet	bsizett	enie	Gmartimak
iW	edeJ	ehcarpS	red	tleW	tztiseb	enie	kitammarG
mW	ɘbɘႱ	ɘʜɔɒɿqƧ	ɿɘb	ʇlɘW	ʇƨʇiƨɘd	ɘiuɘ	ʞiʇɒmmɒɿᎮ

- Chandra, Krügel, and Engbert (2019) tested 36 native German speakers with normal or corrected-to-normal vision
- First session: Reading normal text (condition N)
- Second session: One of four manipulated reading conditions (conditions mL, iW, mW, and sL; see Table 1)

Group-level results

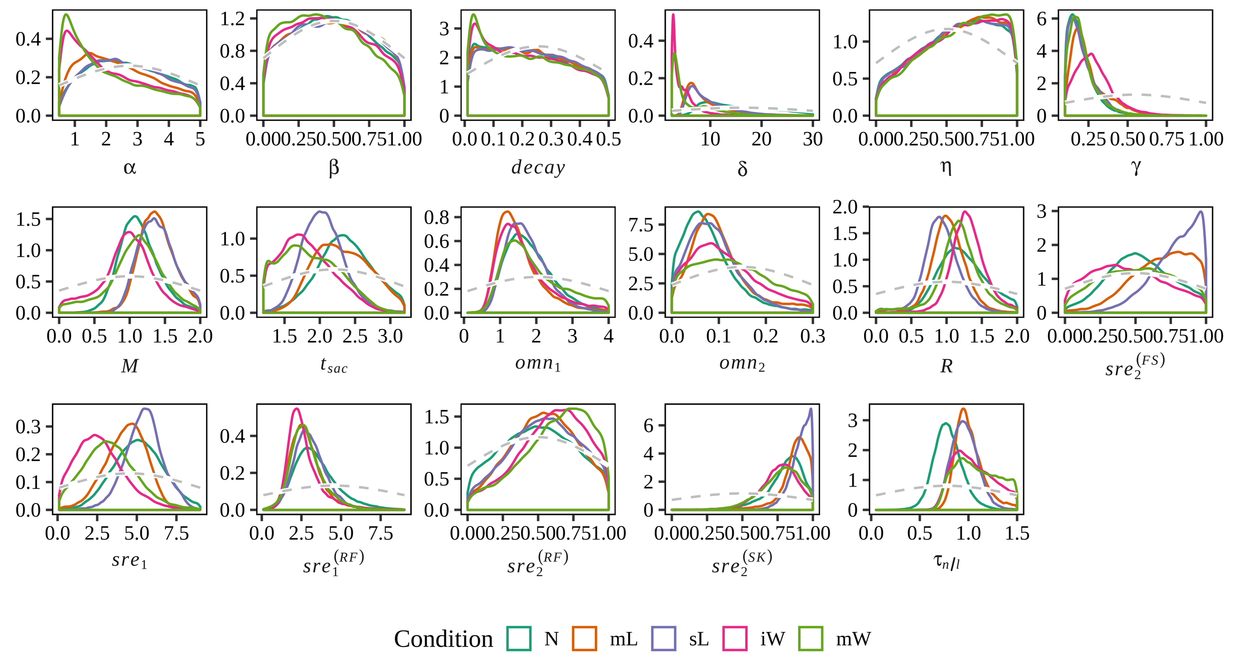

Figure 6. Posterior parameter distributions. Each color represents the aggregated sampled posteriors across all subjects in that condition. Priors (gray dashed lines) are truncated normal with support on 1 SD around the mean and were identical across subjects and conditions.

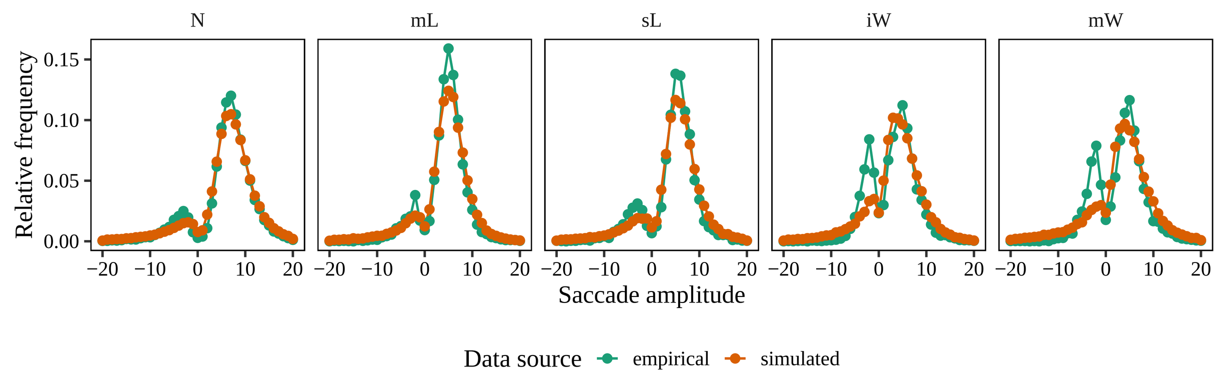

Figure 7. Empirical and simulated saccade amplitudes aggregated across all participants in each experimental condition, including the baseline condition (N).

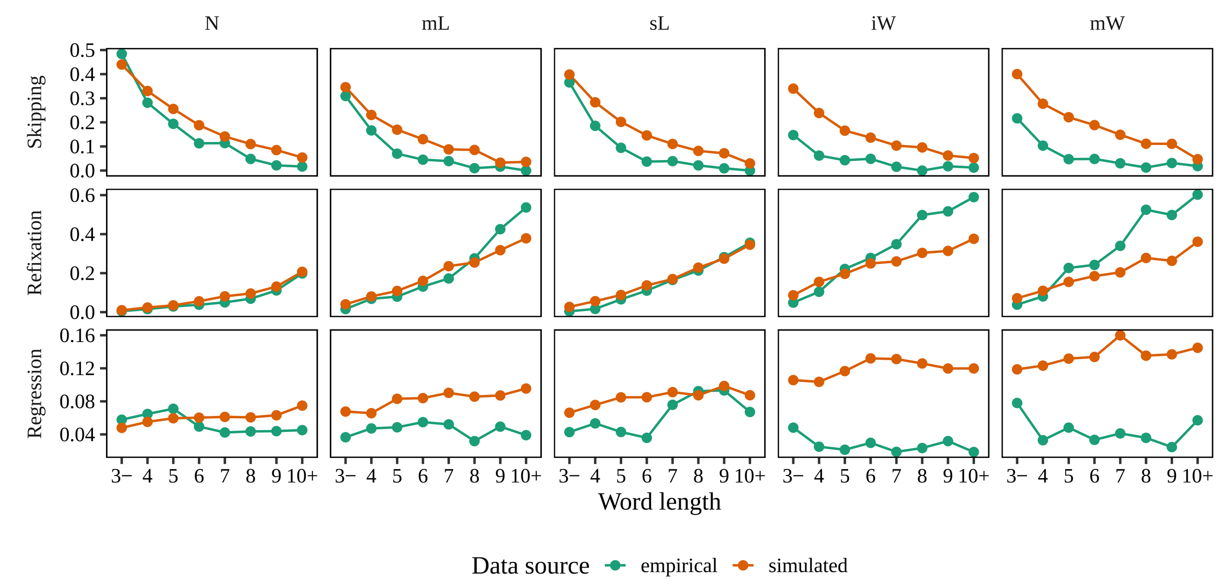

Figure 8. Empirical and simulated spatial summary statistics for different experimental conditions, aggregated across subjects, as a function of word length.

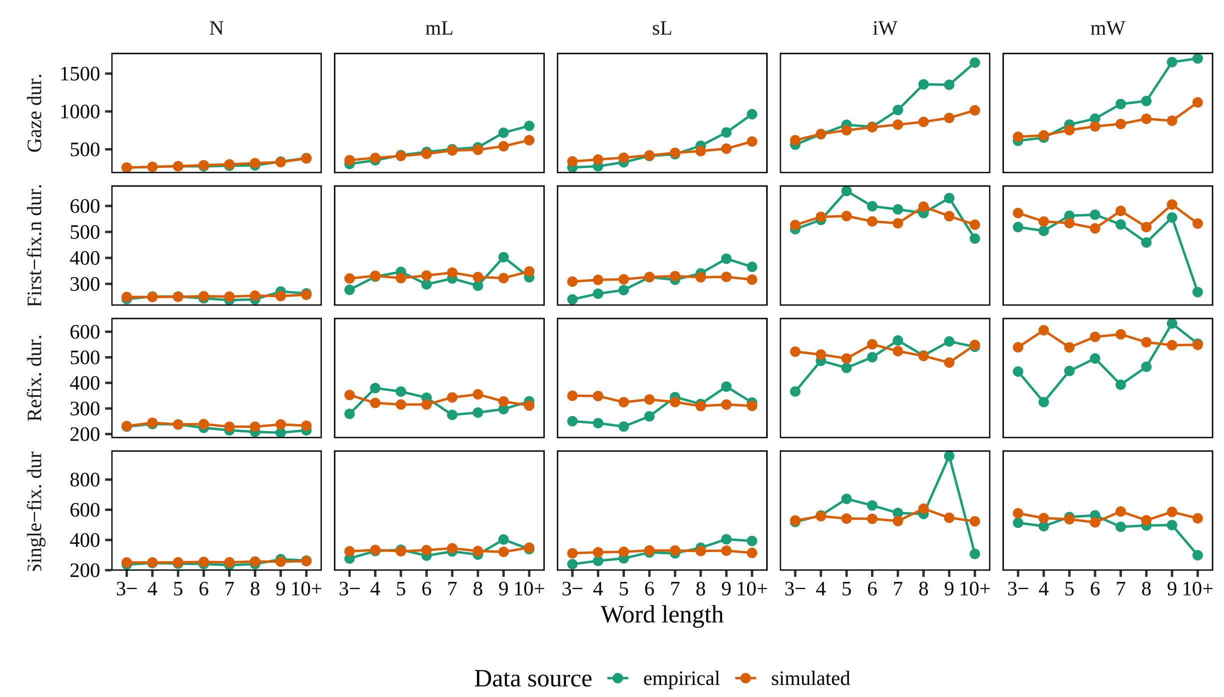

Figure 9. Empirical and simulated temporal summary statistics for different experimental conditions, aggregated across subjects, as a function of word length.

Interindividual variability

In addition to variability at the group level (i.e., effects between experimental conditions), the method also reliably predicts the between-subject variability for most spatial and temporal summary statistics.

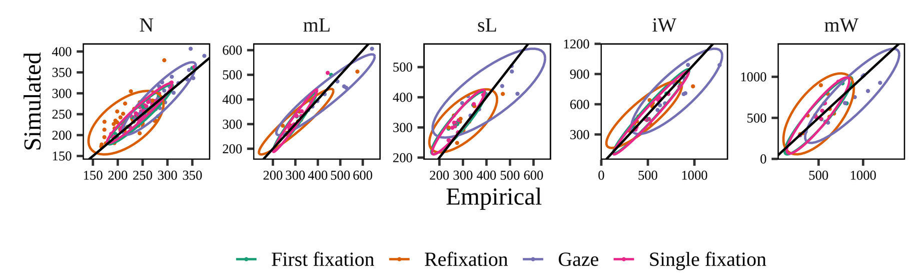

Figure 10. Correlation between empirical and simulated temporal summary statistics. Each participant is represented by one dot in each color in the respective experimental condition (panel).

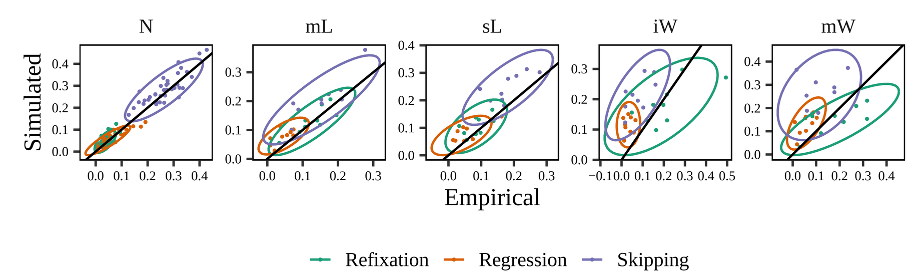

Figure 11. Correlation between empirical and simulated spatial summary statistics. Each participant is represented by one dot in each color in the respective experimental condition (panel).

Analysis of parameter estimates

Effects in model parameters between conditions (see Figure 12) may explain observable differences between experimental conditions. We therefore conducted linear regression analyses with the model parameters as dependent variables. p-values were corrected for multiple testing (Šidák, 1967) and assumed "significant" if $p_S < .05$. The contrast matrix was derived from these four null hypotheses using the *hypr* package in R (Rabe, Vasishth, Hohenstein, Kliegl, & Schad, 2019; Schad, Vasishth, Hohenstein, & Kliegl, 2020):

$$\begin{aligned} \textit{word inverted}: &\quad \mu_{iW} = \mu_N \\ \textit{letter flipped}: &\quad \mu_{mL} = \mu_N \\ \textit{both flipped}: &\quad \mu_{mW} = \mu_N + (\mu_{mL} - \mu_N) + (\mu_{iW} - \mu_N) \\ \textit{scrambled}: &\quad \mu_{sL} = \mu_{mL} \end{aligned}$$

Inverting words is associated with...

- narrower processing span δ
- higher baseline word difficulty α
- smaller optimal saccade amplitudes for forward fixations (sre_1)
- shorter global saccade timer t_{sac}
- longer labile and non-labile saccade programs $\tau_{n/l}$

Flipping letters horizontally is associated with...

- narrower processing span δ
- longer labile and non-labile saccade programs $\tau_{n/l}$
- delayed saccade timer M after misplaced fixations
- accelerated saccade timer R after well-placed refixations
- less targeted saccade execution onto word $n+1$ ($sre_2^{(FS)}$)

The effects of word inversion and letter flipping on model parameters are not significantly non-additive, as there is no significant effect for the interaction (*both flipped*).

In the *scrambled words* condition, model parameters are not significantly different from the effects of letter flipping, except for the optimal saccade target location (sre_1) being shifted to the right.

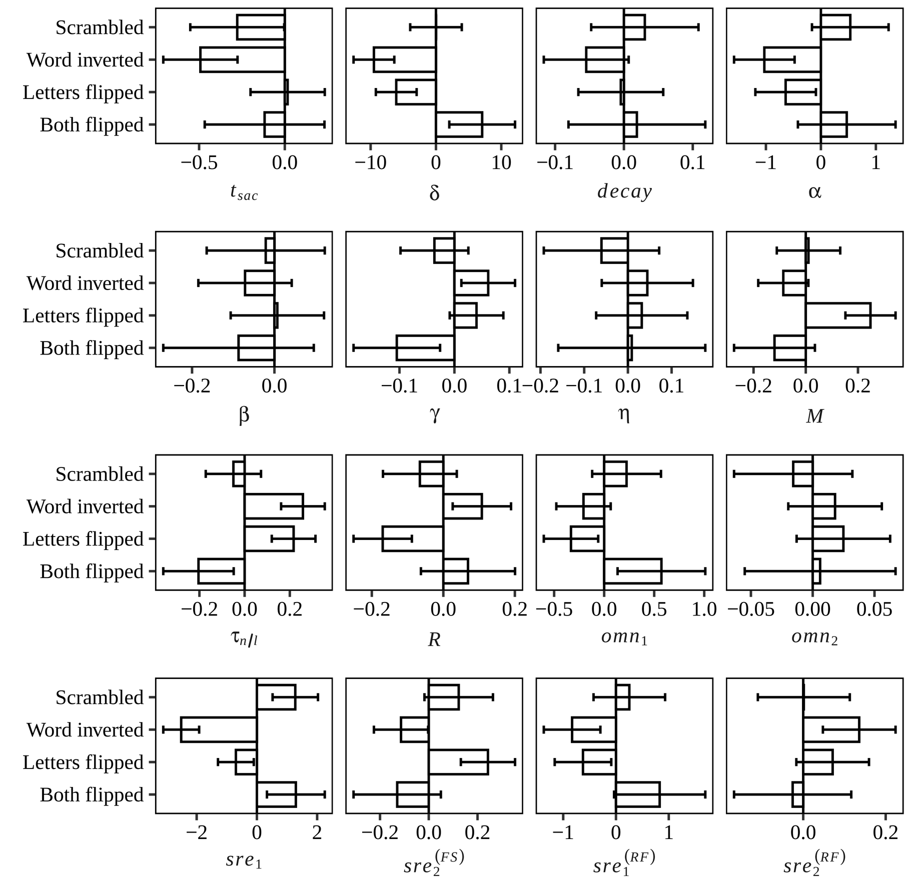

Figure 12. Linear regression coefficient estimates with uncorrected 95% confidence intervals around the estimated means. The baseline is tested against zero, while conditions are tested against the baseline.

Summary and Conclusion

- SWIFT was successfully fitted to empirical data collected under different reading conditions
- The modelling approach is not data-hungry and can operate on single-participant data
- Goodness of fit was evaluated by comparing empirical and simulated summary statistics
- Subject-levfel parameters can reliably predict characteristics of reading patterns in unseen trials
- Differences in subject-level parameters could explain why and how differences in reading behavior arise
- SWIFT could serve as a viable baseline model for the integration of linguistic parsing mechanisms

References

Chandra, J., Krügel, A., & Engbert, R. (2019). Modulation of oculomotor control during reading of mirrored and inverted texts. Retrieved from `https://arxiv.org/abs/1911.04585`

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813. doi: 10.1037/0033-295X.112.4.777

McConkie, G., Kerr, P., Reddix, M., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, *28*(10), 1107–1118. doi: 10.1016/0042-6989(88)90137-X

Rabe, M. M., Vasishth, S., Hohenstein, S., Kliegl, R., & Schad, D. J. (2019). hypr: Hypothesis matrix translation [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=hypr` (R package version 0.1.3)

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020, feb). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. doi: 10.1016/j.jml.2019.104038

Seelig, S. A., Rabe, M. M., Malem-Shinitski, N., Risse, S., Reich, S., & Engbert, R. (2019). Bayesian parameter estimation for the SWIFT model of eye-movement control during reading. Retrieved from `http://arxiv.org/abs/1901.11110`

Shockley, E. (2019). *PyDREAM*. Retrieved from `https://github.com/LoLab-VU/PyDREAM`

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*, 626–633. doi: 10.1080/01621459.1967.10482935

Vrugt, J. A., ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, *10*(3). doi: 10.1515/ijnsns.2009.10.3.273

Acknowledgment

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project no. 317633480 – SFB 1287, project B03. We acknowledge the North-German Supercomputing Alliance (HLRN) for providing high-performance computing resources that have contributed to the research results reported in this paper.