Decomposing word learning trajectories: a Bayesian hierarchical Gompertz model for fixation patterns in cross-situational word learning

Jens Schmidtke
Leipzig University
Jens.schmidtke@me.com

Accuracy is the most widely used measure in word learning research, but it conflates distinct aspects of the learning process: two learners who reach ceiling performance may have arrived there via different trajectories. The visual world paradigm offers a way into this problem, as eye movements to visual referents are closely time-locked to the speech signal and reflect the unfolding of lexical activation in real time. Here, we combine this method with the cross-situational word learning paradigm in which participants see two images of known objects, the target and a foil, while hearing a word in an unknown language. No single exposure disambiguates the mapping, but across exposures the correct word-referent pairing can be inferred. Under the assumption that eye movements to the target reflect mapping strength, time course data over trials and exposures reveal how the learning trajectory evolves moment-by-moment. Data come from a study in which participants were exposed to novel words under high (12 repetitions) or low (5 repetitions) exposure conditions. In the low exposure condition, the foil was always from the high-exposure condition, thereby reducing referential uncertainty. Additionally, participants were assigned to hear words from either a single or from multiple talkers.

To capture the learning trajectory, we present a Bayesian hierarchical three-parameter Gompertz model fitted to eye-tracking data. The Gompertz function was chosen over the more common logistic function because it accommodates asymmetric slopes. The advantage of this generative model over a descriptive model such as regression is that its parameters are directly interpretable: A (upper asymptote: strength of the final word-referent mapping), k (growth rate: steepness of the activation rise), and τ (displacement: the point at which activation begins to rise; see Fig. 1). Critically, rather than estimating these parameters at a single time point, we model their rate of change across successive exposures, yielding a dynamic picture of how each aspect of learning evolves.

Model fit was assessed via leave-one-out cross-validation, indicating adequate fit for both exposure conditions (99.8% of observations had Pareto $k < 0.5$, indicating reliable estimates). The model captured clear learning trajectories across exposures: A increased substantially above chance, τ shifted from late to early in the trial window, and k steepened, reflecting a sharpening of the activation rise (Fig. 2). When referential uncertainty was reduced, word-referent mappings strengthened more rapidly, reaching asymptotic levels after just 5 exposures comparable to those achieved after 12 exposures in the high exposure condition (Fig. 2). Talker condition had an effect on τ, suggesting that from exposure 6 onwards participants in the 5-talker condition shifted their gaze to the target earlier relative to the single talker condition (see Fig. 3). A subsequently administered four-alternative forced-choice test provided an independent measure of accuracy, which was associated with A ($r = 0.65$ for high exposure, $r = 0.71$ for low exposure) but not with τ or k, suggesting that decomposing learning trajectories into distinct parameters captures aspects of individual variation that accuracy alone cannot recover.

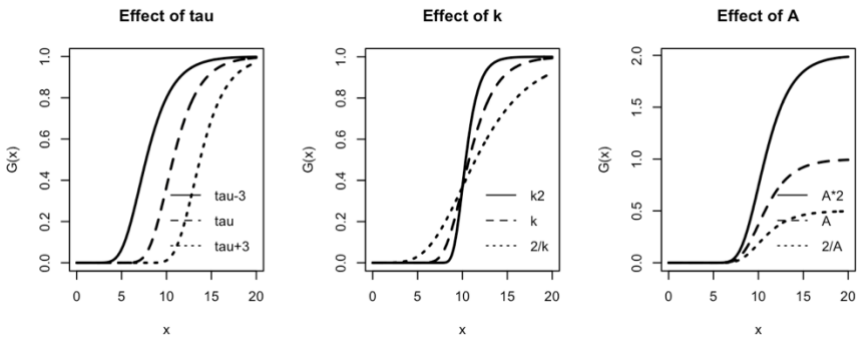

Figure 1. Effect of each Gompertz parameter on the shape of the learning trajectory. Each panel varies one parameter while holding the others constant. Left: shifting τ translates the curve horizontally, moving the onset of activation earlier or later without affecting its shape. Center: increasing k steepens the rise, while decreasing k produces a more gradual trajectory. Right: A scales the upper asymptote, determining the maximum level of target preference reached.

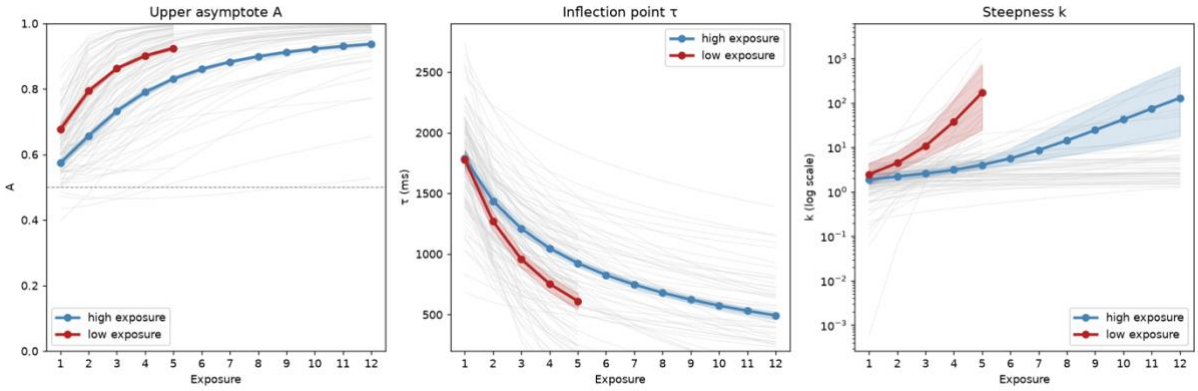

Figure 2. Change in the three model parameters over exposure. Each grey line shows the mean posterior estimate for one participant, and the bold lines show the group mean in each exposure condition. High exposure = 12 target repetitions. Low exposure = 5 target repetitions.

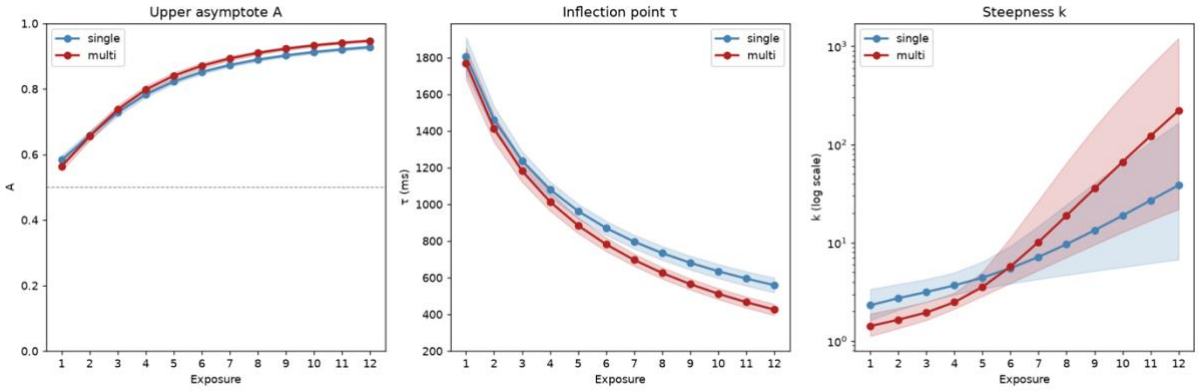

Figure 3. Posterior mean trajectories of the upper asymptote (A), inflection point (τ), and steepness (k) across exposures for single-talker (blue) and multi-talker (red) conditions. Shaded regions indicate 95% posterior credible intervals. Parameters are averaged across participants within each group.