Appendix: Comparison of learning curves for the GRUs and LSTMs in each language (English, German, Spanish). To augment Figure 2 in the main paper, we also show the learning curves for all models experimented with in this paper beyond the $\Delta$-RNN. Validation learning curves are provided for the GRU and LSTM language models, both multimodal and unimodal variations.