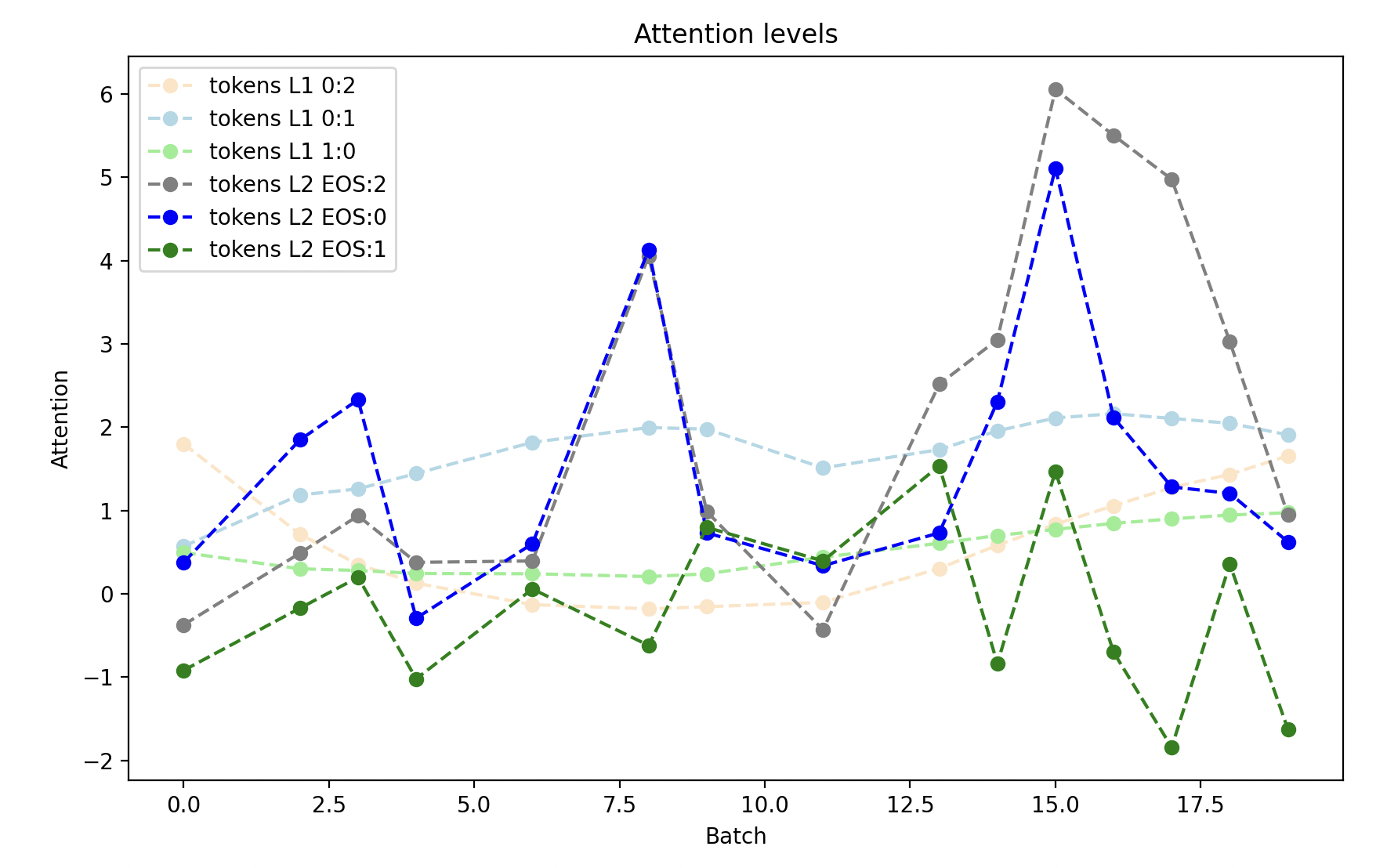
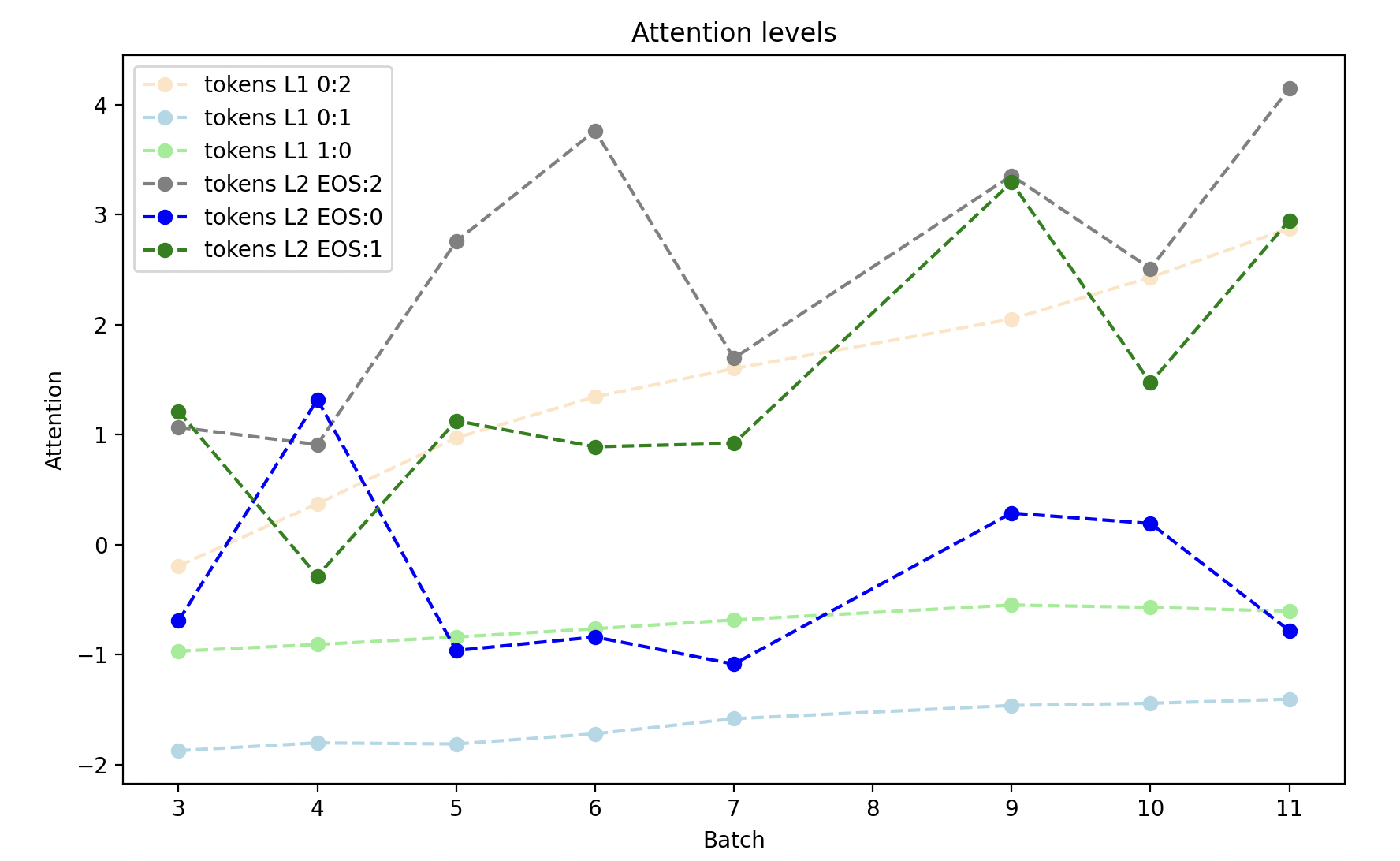
# Attention Levels and Individual Layer Learning rates notes

## Single Learning Rate 0.01

### Typical convergence

Smooth momentum L1, overshooting L2



### Occasional failure to converge