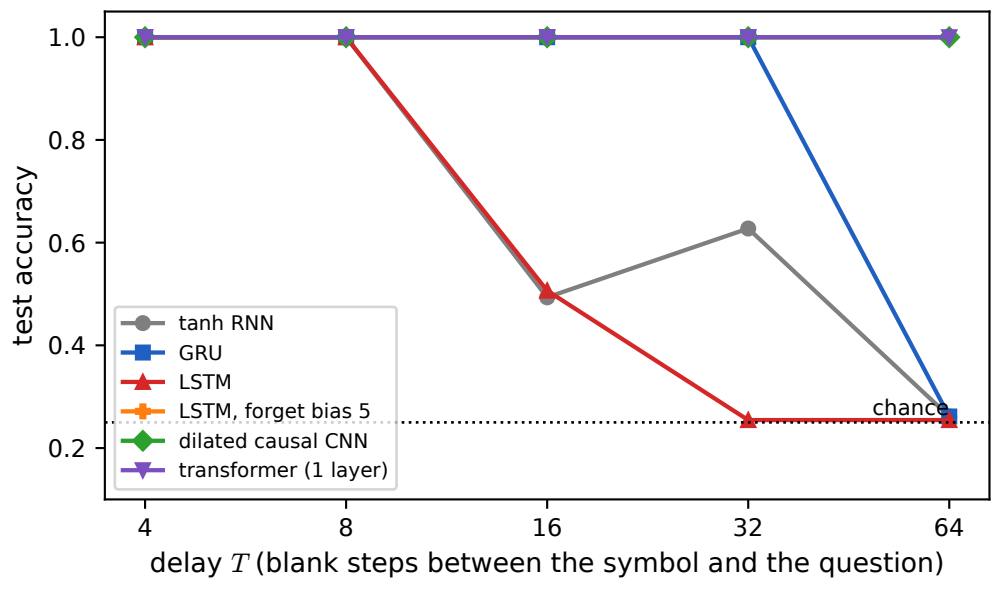

1.0
0.8
0.6
0.4
0.2
test accuracy
tanh RNN
GRU
LSTM
LSTM, forget bias 5
dilated causal CNN
transformer (1 layer)
chance
4
8
16
32
64
delay T (blank steps between the symbol and the question)