## B) H during training A) RNN prediction + Match / Non-match (M/NM) task prediction $\hat{\mathbf{v}}_{\scriptscriptstyle{+}}$ input x<sub>t</sub> neural activity r<sub>t</sub> 00 M/NM 1,00 Cross entropy *H* 0.75 0.50 0.25 0 weights V 00000 $A_{\alpha}$ 0 0 $B_{\alpha}$ ŏ $C_{\beta 1}$ 0 ŏ o' 0.00 ιŏ weights **U** D Ō NM 5 10 15 20 weights T weights W Epoch t+1 C) Distribution of cross-temporal α decoding accuracy D) Histogram of red regime Average $\alpha$ accuracy Variance $\alpha$ accuracy 1.0 0.10 0.8 0.08 Training time Training time 15 Frequency 0.6 0.06 10 0.04 0.4 5 0.2 0.02 00 Δ 0.0 0.2 0.4 0.6 0.8 1.0 Average accuracy 0.0 0.00 A A ВВ C C D D A A ВВ C C D D Testing time $t \rightarrow$ Testing time $t \rightarrow$ E) Number of switch neurons F) Number of stable neurons G) Match vs Non-match Stable 0.75 cnm. fraction of runs 0.50 cnm. 0.25 0.00 1.00 0.75 0.50 0.25 0.00 Switch # projecting neurons Prediction only P = 8e-085 Prediction + M/NM 3 Prediction only P = 0.009Prediction + M/NM 0.0 2.5 5.0 7.5 10.0 10 15 Match Non-match # Switch neurons # Stable neurons