A) RNN prediction + Match / Non-match (M/NM) task B) H during training prediction $\hat{\mathbf{y}}_{t}$ neural activity \mathbf{r}_{t} input xt H_{Pred} (Pred-only) 1.0 Cross entropy H 0 weights V $H_{M/NM}$ (MNM-only) ō A_{α} , · H_{Pred} (Pred & MNM) 0 Ō 0.5 $H_{M/NM}$ (Pred & MNM) 0 B_{α} Ō O (o) M $(\alpha = \beta)$ weights ${\bf U}$ D NM (α≠β) 10 20 30 0 40 weights T weights W **Epoch** t+1C) Distribution of cross-temporal α decoding accuracy D) Histogram of red regime Average α accuracy Variance α accuracy 0.10 - 1.0 5 15 8 0.8 0.08 Fraining time τ Frequency Fraining time 0.06 0.6 10 -0.04 0.4 5 0.2 0.02 0.0 0.00 0.00 0.50 ВВ CC DD ВВ CC DD 0.25 0.75 1.00 AΑ Testing time $t \rightarrow$ Testing time $t \rightarrow$ Average accuracy E) Number of switch neurons F) Number of stable neurons G) Match vs Non-match # projecting neurons # 0.00 Pred-only Fraction of runs Fraction of runs Pred-only Pred & Stable P = 0.29M/NM Pred & M/NM Switch 0.1 P = 0.019P = 0.0040.0 0.0 0.0 2.5 5.0 7.5 10 15 5 # Switch neurons # Stable neurons Match Non-match