The mean activities at early stage (the smallest median RT of the six conditions, i.e., 410 ms after the onset of stimulus, indicated by arrows **a** and **b**) and at the onset of model choice (indicated by arrows **c** and **d**) were examined in the following panels. **E***.* Quantification of the best-fit-to-behavior model prediction (dots and lines) to the empirical recordings (crosses). Upper panel: the early-stage activities at the median RT indicated by arrows **a** (chosen side) and **b** (unchosen side). Lower panel: the late-stage activities aligned to the onset of model choice (30 ms before saccade) indicated by arrows **c** (chosen side) and **d** (unchosen side). The model activities were rescaled to the threshold of the empirical activities, i.e., the mean activity across coherences indicated at arrow **c**. The root-mean-square error (RMSE) between the data and the model at the median RT and at the choice onset were calculated and indicated on the panels. **F**. The model predicted *G* dynamics show faster decreasing on the chosen units than the unchosen units, indicating that the chosen units are more strongly disinhibited. **G**. WTA competition is terminated earlier in the trials with higher input coherence, so that the aggregated activities in the chosen *G* units increase with coherence (**a** and **c**) as an outcome of a shorter period of inhibition from *D* units. The unchosen *G* activities decreases with coherence in the early stage (**b**) because of the impact of more rapid WTA in higher coherence condition but increases the coherence-dependency pattern in the late stage (**d**) because of a shorter period of competition when the coherence is higher. **H**. The model predicted *D* activities ramp up fast in the early stage (dynamics on the left, sorted to the stimulus onset); the aggregating strategy makes the coherence-dependency on the chosen side hard to see. In the late stage (dynamics on the right, sorted to the choice onset), the chosen *D* units do not reach a common threshold since the trials with lower coherence take longer time of competition and drive the *D* activities higher. **I**. In the early stage, the aggregated activities of the unchosen *D* show dependency on input coherence (**d**) but the chosen *D* show weaker dependency on coherence (**a**). In the late stage, the chosen *D* exhibit higher activity than the unchosen *D* as the outcome of WTA competition; *D* units with lower input coherence (i.e., longer RT) reaches higher activities since longer period of competition.