Supplementary Information

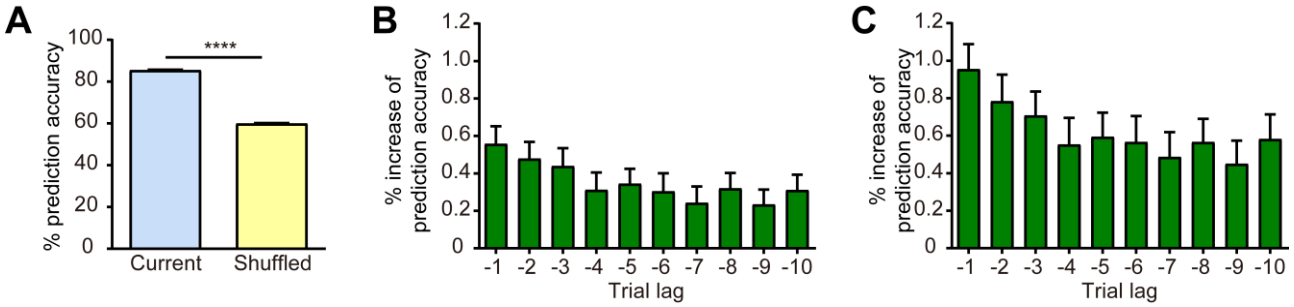

Fig. S1 Contribution of current and historical trial information to behavioral prediction. **A** Prediction accuracy of a model using only current-trial stimuli, compared to a shuffled control (paired *t*-test, ****$P < 0.0001$). **B** Improvement in prediction accuracy gained by incorporating stimuli from a single historical trial (lags 1–10), relative to shuffling each lag-specific history. **C** Improvement in prediction accuracy gained by incorporating both stimuli and action-outcome associations from a single historical trial (lags 1–10), relative to the corresponding shuffled control.

Beyond the influence of current stimuli, both stimuli and action-outcome associations from previous trials contributed to the prediction of rat behavior. We first computed the accuracy of predicting rat choices based solely on current stimuli to evaluate the relative contribution of the variables in behavioral prediction. The mean accuracy across 132 sessions was 85.44% ± 0.32% (mean ± SEM), which was significantly higher—by 25.51% ± 0.44%—than the accuracy obtained after shuffling current stimuli within each session (59.93% ± 0.28%; Fig. S1A; $P < 0.0001$). Subsequently, we separately incorporated data from each of the preceding 1–10 trials to assess the impact of historical trial information. Specifically, for each model, we added the stimuli (or stimuli and action-outcome associations) from only a single historical trial (e.g., trial −1, −2, up to −10) to Eq. 2 and compared the resulting prediction accuracy against that derived from shuffling that specific historical data, while preserving current trial stimuli (Eq. 3 and 4). We observed that the most recent trial (i.e., trial −1) exerted the strongest influence on behavioral prediction: including its stimuli improved accuracy by 0.56% ± 0.09% (from shuffled: 86.54% ± 0.30% to real: 87.10% ± 0.28%; Fig. S1B), while incorporating both stimuli and action-outcome associations from the most recent trial enhanced accuracy by 0.95% ± 0.13% (from shuffled: 86.97% ± 0.31% to real: 87.92% ± 0.27%; Fig. S1C). Although these improvements were statistically significant ($P < 0.0001$), they were markedly smaller than the gain attributable to current trial stimuli (25.50% ± 0.44%).

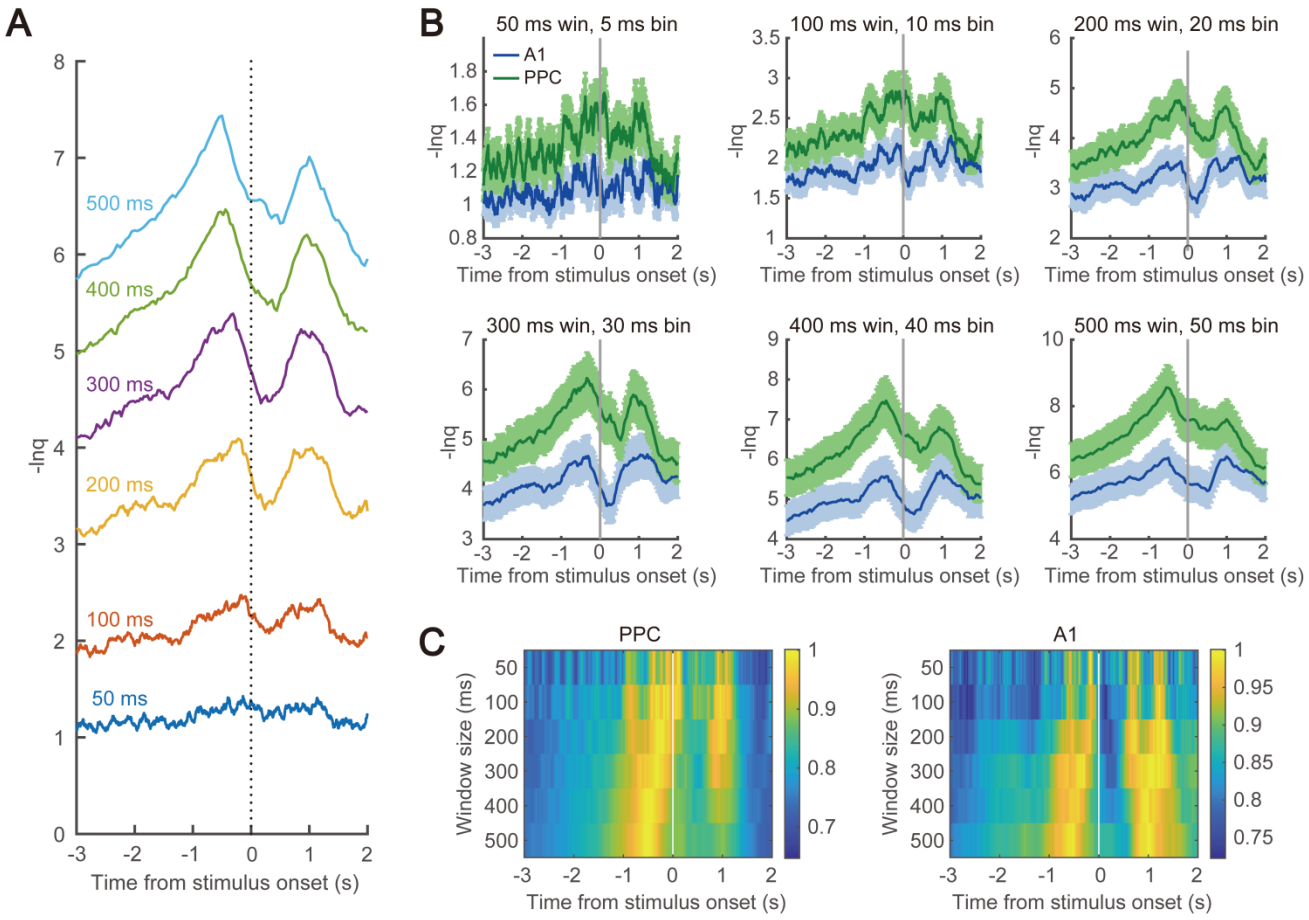

Fig. S2 Temporal dynamics of rule encoding strength across different analysis parameters. Rule encoding was quantified using six sliding-window sizes: 50 ms (5 ms bins), 100 ms (10 ms bins), 200 ms (20 ms bins), 300 ms (30 ms bins), 400 ms (40 ms bins), and 500 ms (50 ms bins). Encoding strength is expressed as the negative natural logarithm of the FDR-corrected *q*-value from the regression *F*-test (mean across sessions, $n = 132$). **A** Rule encoding was consistently enhanced during the pre-stimulus and post-choice epochs across all window sizes. Absolute statistical values increased with larger windows (due to greater smoothing and noise reduction), but the direction and relative timing of effects remained unchanged. **B** PPC consistently showed stronger rule encoding than A1 for all parameter sets. **C** Each curve was normalized to its maximum value (range 0–1) to facilitate comparison of temporal profiles. The pre-stimulus peak shifted closer to stimulus onset as the window size decreased.

We repeated the entire analysis using six different window sizes (50, 100, 200, 300, 400, and 500 ms) to examine whether the observed rule-encoding depends on the choice of sliding-window size. Across all window sizes, the same qualitative pattern emerged: rule encoding was consistently enhanced during the pre-stimulus and post-choice epochs (Fig. S2A), and PPC encoding remained stronger than A1 (Fig. S2B). As expected, larger windows increased absolute statistical values due to greater smoothing and noise reduction, but the direction and relative timing of effects were unchanged. In addition, the pre-stimulus peak shifted closer to stimulus onset with smaller windows (Fig. S2C), a

direct consequence of temporal smoothing: larger windows integrate activity over longer intervals, shifting the estimated peak away from true event times. Importantly, despite this systematic shift in peak timing, the qualitative pattern—enhanced rule encoding in both pre-stimulus and post-choice epochs, with PPC consistently stronger than A1—was present across the entire range of window sizes (50–500 ms).

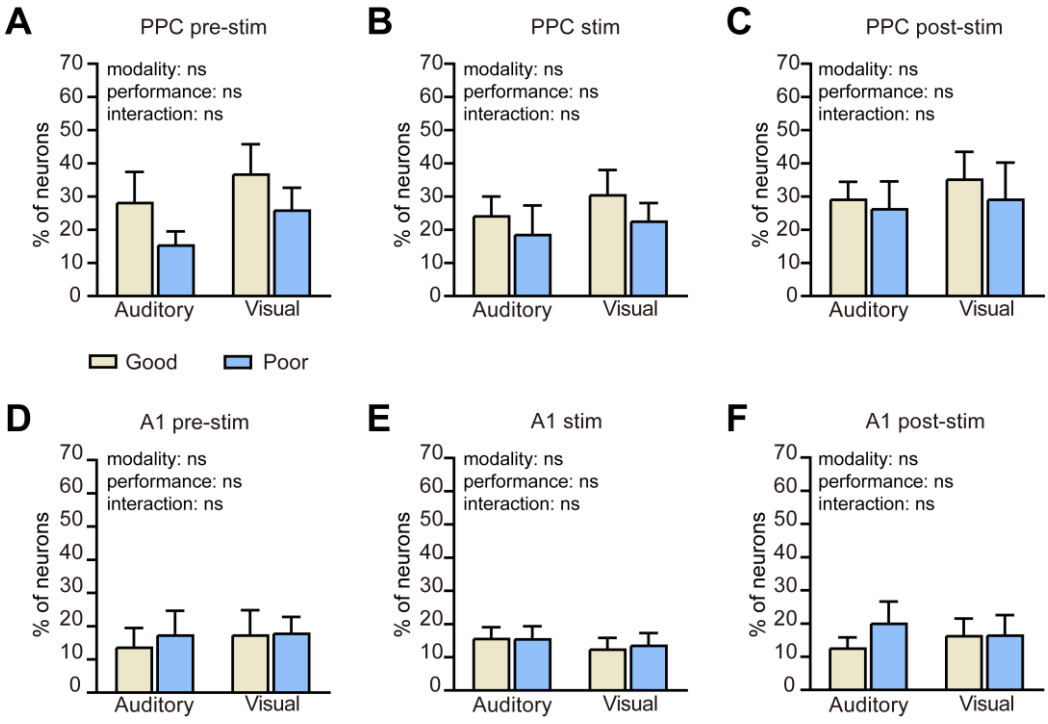

Fig. S3 Neural encoding of task rules in PPC and A1 based on individual rats. **A** Proportion of rule-encoding neurons in PPC during the pre-stimulus epoch, compared between rats with good (yellow) and poor (blue) performance on auditory and visual attention tasks. Two-way ANOVA with performance level and attention modality as factors; n = 8 rats. **B** Same comparison as in **A** for PPC during the stimulus epoch. **C** Same comparison as in **A** for PPC during the post-stimulus epoch. **D–F** Same comparisons as in **A–C** for A1 neurons in the pre-stimulus, stimulus, and post-stimulus epochs, respectively.

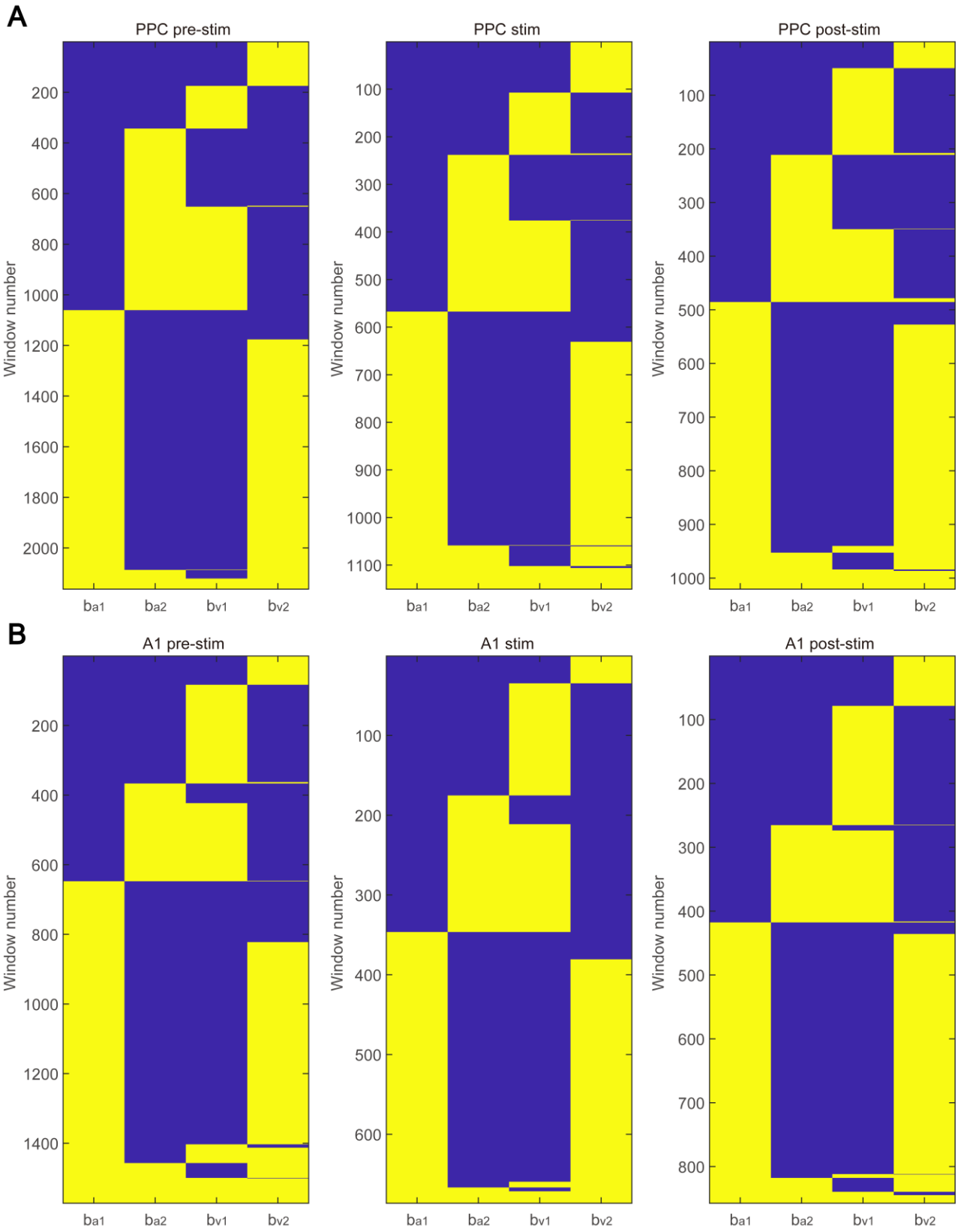

Fig. S4 Binary encoding of rule-factor combinations across task epochs. **A** Encoding profiles of PPC rule-selective neurons during pre-stimulus (left), stimulus (middle), and post-stimulus (right) epochs. **B** Encoding profiles of A1 rule-selective neurons across the same three epochs. Colors indicate whether each rule factor (b_{a1}, b_{a2}, b_{v1}, b_{v2}) was significantly encoded (yellow) or not (blue) in a given

time window. Profiles are categorized into 15 classes of rule-factor combinations. No temporal differences in encoding profiles were observed across epochs (see Methods).

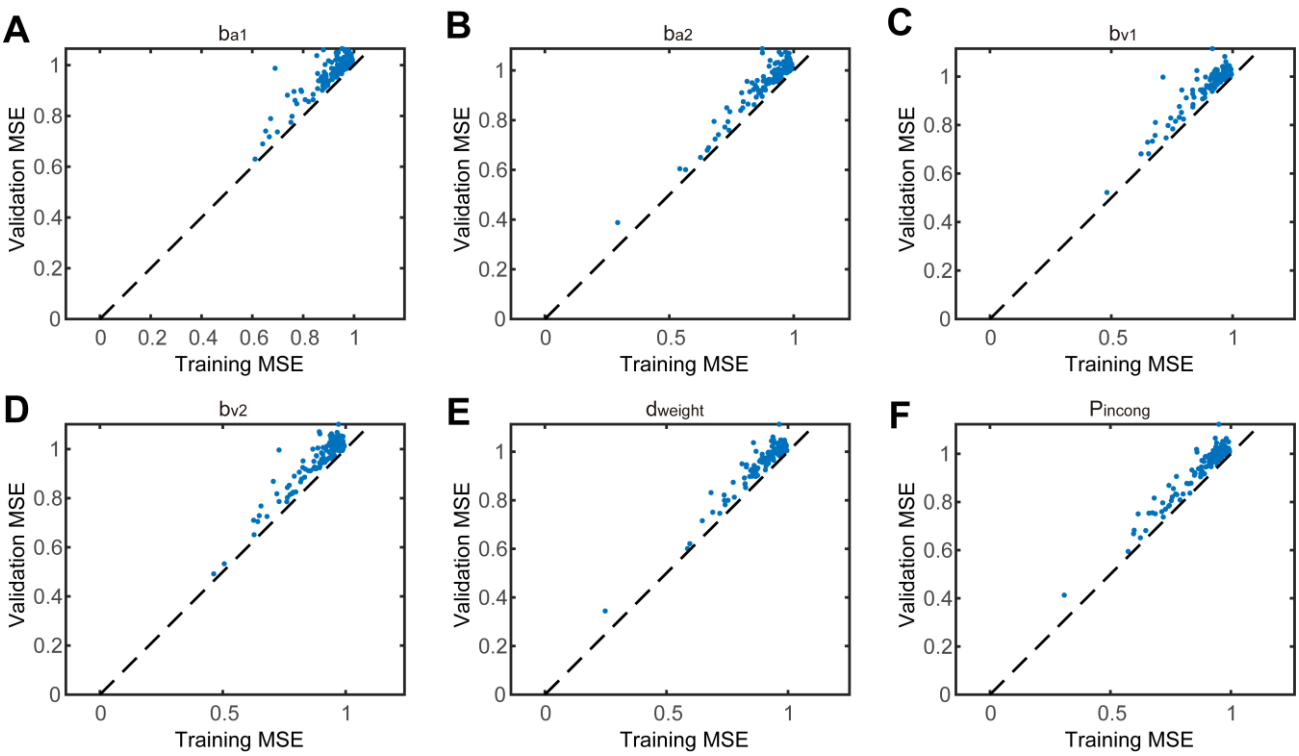

Fig. S5 Comparison between training MSE and validation MSE in cross-validation for population decoding. Scatter plots of the mean training MSE versus mean validation MSE for each session (n = 132) are shown for the four rule variables (**A**, b_{a1}; **B**, b_{a2}; **C**, b_{v1}; **D**, b_{v2}), d_{weight} (**E**), and P_{incong} (**F**). For all conditions, Pearson's correlation between training MSE and validation MSE was significant (P < 0.0001).

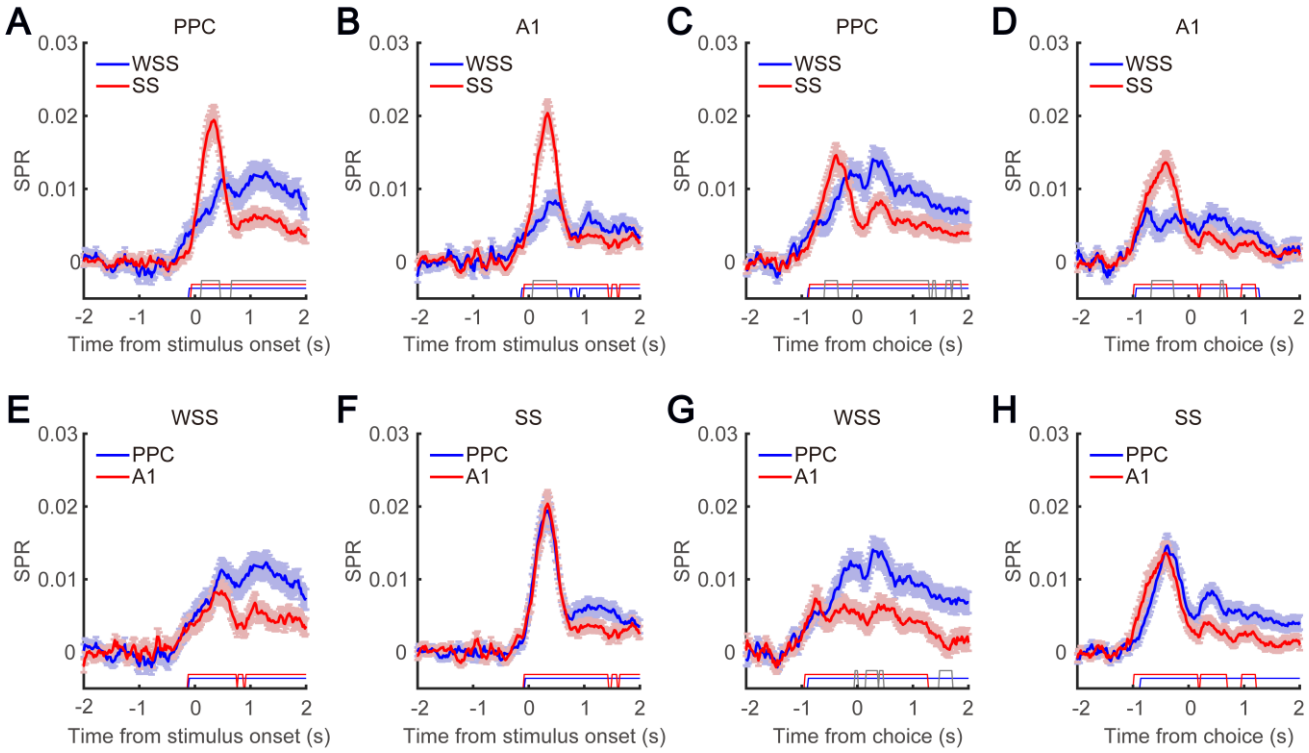

Fig. S6 Temporal hierarchy of stimulus-driven and rule-modulated neural representations. **A–D** Population time course of semi-partial R^2 (SPR) values for SS (stimulus-sensitive) and WSS (weighted stimulus-sensitive) neurons. **A, B** SPR for SS and WSS neurons in PPC (**A**) and A1 (**B**), aligned to stimulus onset. **C, D** Same analysis aligned to choice time in PPC (**C**) and A1 (**D**). Colored horizontal bars above the x-axis mark time bins with significant encoding relative to baseline (paired *t*-test, FDR-corrected $q < 0.01$): blue for WSS, red for SS. Gray bars indicate time bins with significant differences between WSS and SS (*t*-test, FDR-corrected $q < 0.01$). **E–H** Cross-regional comparison of WSS and SS representations. **E, F** Direct comparison of SPR values between PPC and A1 over time for WSS (**E**) and SS (**F**), aligned to stimulus onset. **G, H** Same comparison aligned to choice time. Colored bars indicate time bins with significant difference: blue for PPC versus baseline, red for A1 versus baseline (paired *t*-test, FDR-corrected $q < 0.01$); gray for PPC versus A1 (*t*-test, FDR-corrected $q < 0.01$).

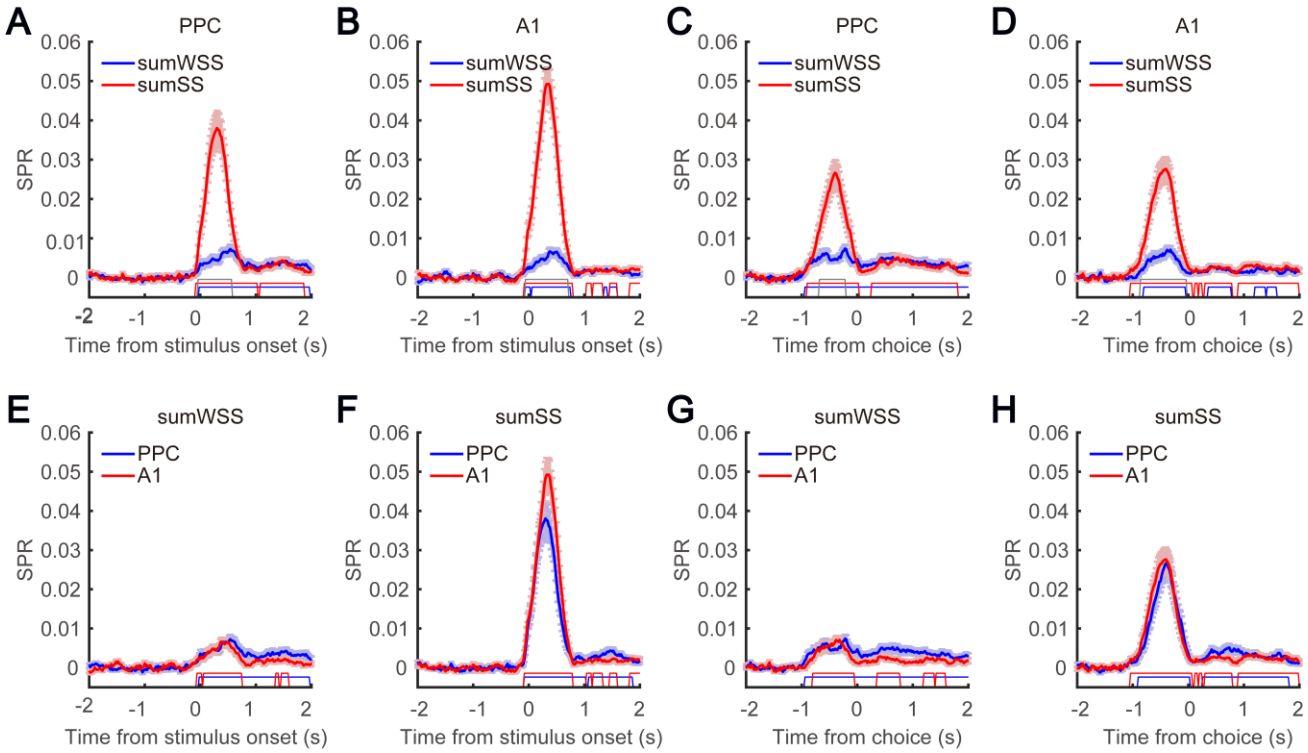

Fig. S7 Temporal hierarchy of integrated sensory representations. **A–D** Population time course of semi-partial R^2 (SPR) values for sumSS (summed stimulus-sensitive) and sumWSS (summed weighted stimulus-sensitive) neurons. **A, B** S*PR* for sumSS and sumWSS neurons in PPC (**A**) and A1 (**B**), aligned to stimulus onset. **C, D** Same analysis aligned to choice time in PPC (**C**) and A1 (**D**). Colored horizontal bars above the x-axis mark time bins with significant modulation relative to baseline (paired *t*-test, FDR-corrected $q < 0.01$): blue for sumWSS, red for sumSS. Gray bars indicate time bins with significant differences between sumWSS and sumSS (*t*-test, FDR-corrected $q < 0.01$). **E–H** Cross-regional comparison of sumWSS and sumSS representations. **E, F** Direct comparison of SPR values between PPC and A1 over time for sumWSS (**E**) and sumSS (**F**), aligned to stimulus onset. **G, H** Same comparison aligned to choice time. Colored bars indicate time bins with significant differences: blue for PPC relative to baseline, red for A1 relative to baseline (paired *t*-test, FDR-corrected $q < 0.01$); gray for PPC versus A1 (*t*-test, FDR-corrected $q < 0.01$).

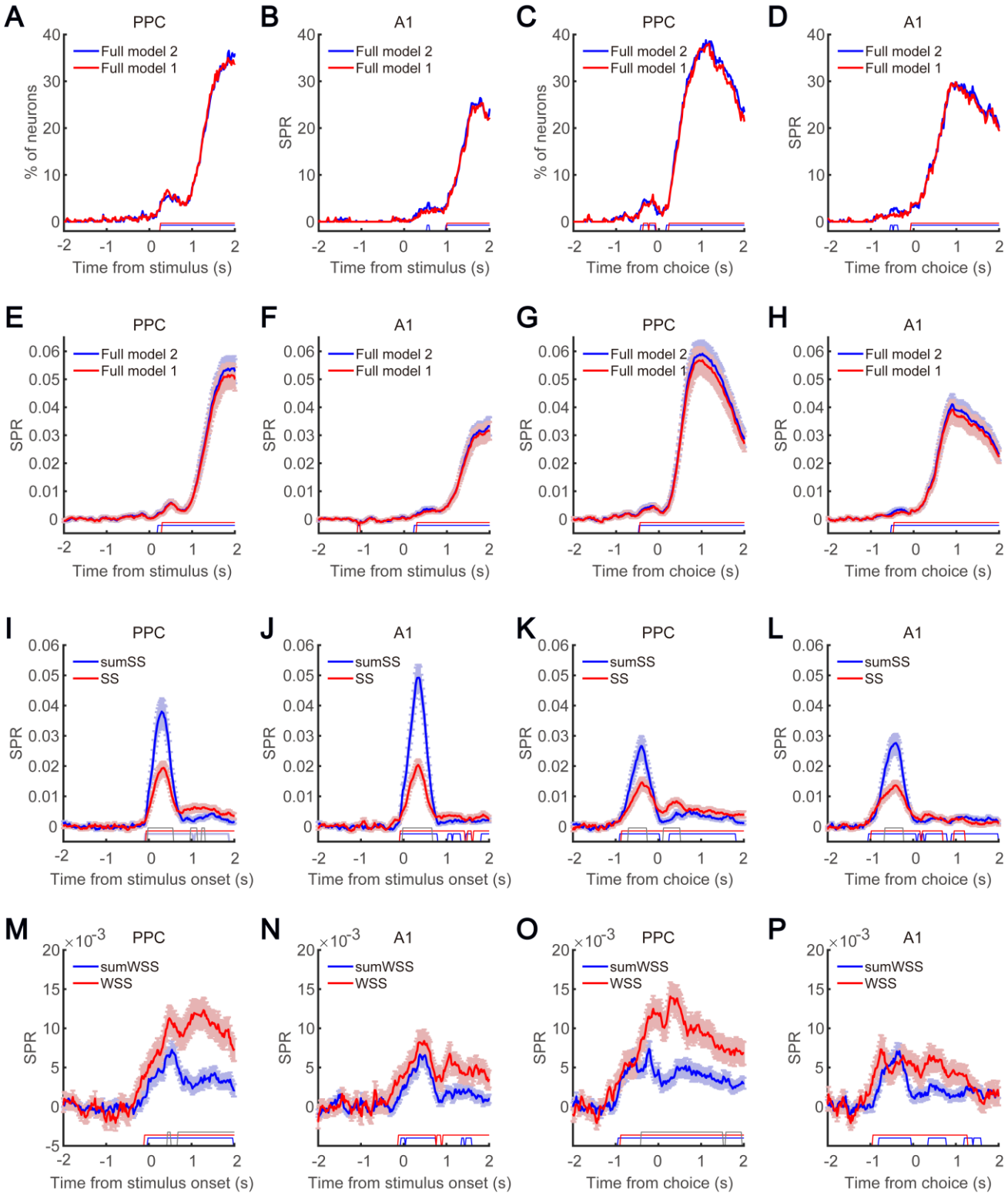

Fig. S8 Comparison of integrated and individual stimulus representations. **A–H** Control analysis validating cross-model comparability using reward encoding. **A–D** Proportions of neurons with significant reward encoding under Full Model 1 (red) and Full Model 2 (blue), aligned to stimulus onset (**A, B**) or choice (**C, D**). **E–H** Semi-partial R^2 (SPR) values for reward-related variance explained by Full Model 1 (red) and Full Model 2 (blue), aligned to stimulus onset (**E, F**) or choice (**G, H**).

Colored bars above the x-axis mark time bins with significant modulation: blue for Full Model 2 versus baseline (paired *t*-test, FDR-corrected $q < 0.01$), red for Full Model 1 versus baseline (paired *t*-test, FDR-corrected $q < 0.01$), gray for comparison between the two models (*t*-test, FDR-corrected $q < 0.01$). **I–P** Population dynamics of integrated versus individual stimulus encoding. **I–L** sumSS (summed stimulus-sensitive) versus SS (stimulus-sensitive) representations: temporal profiles of SPR values in PPC (**I, K**) and A1 (**J, L**), aligned to stimulus onset (**I, J**) or choice (**K, L**). **M–P** sumWSS (summed weighted stimulus-sensitive) versus WSS (weighted stimulus-sensitive) representations: temporal profiles of SPR values in PPC (**M, O**) and A1 (**N, P**), aligned to stimulus onset (**M, N**) or choice (**O, P**). Colored bars above the x-axis indicate time bins with significant modulation: blue for integrated encoding (sumSS/sumWSS) versus baseline (paired *t*-test, FDR-corrected $q < 0.01$); red for individual encoding (SS/WSS) versus baseline (paired *t*-test, FDR-corrected $q < 0.01$); gray for comparison between integrated and individual encoding (*t*-test, FDR-corrected $q < 0.01$).

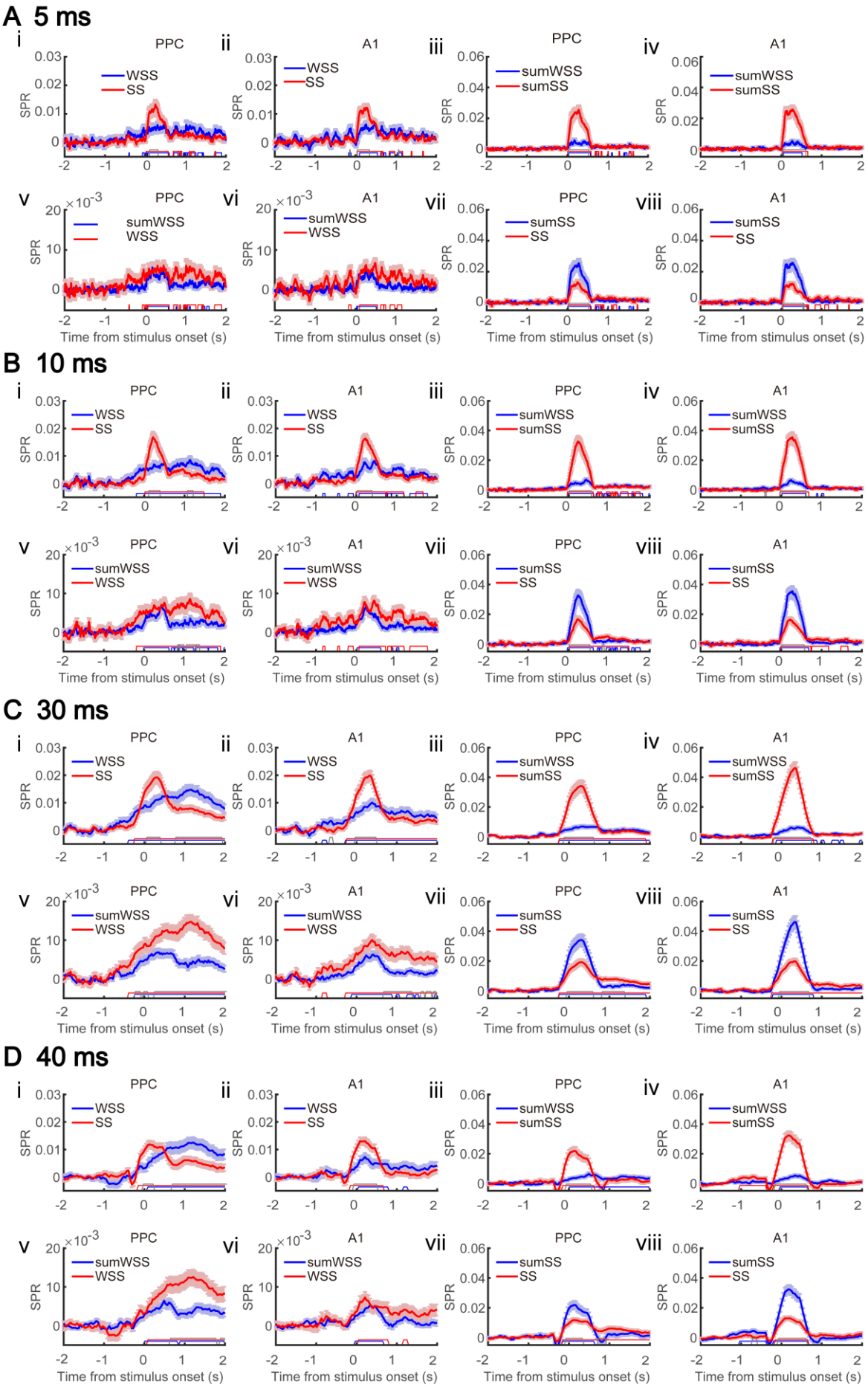
A 5 ms
B 10 ms
C 30 ms
D 40 ms
i
ii
iii
iv
v
vi
vii
viii
PPC
A1
WSS
SS
sumWSS
sumSS
SPR
Time from stimulus onset (s)

Fig. S9 Temporal hierarchy of stimulus-driven and rule-modulated neural representations across different analysis parameters. To test whether the main results (obtained with 20 ms bins and 200 ms window size) are dependent on parameter-specific overfitting, four additional parameter sets were used: **A** 50 ms window, 5 ms bins; **B** 100 ms window, 10 ms bins; **C** 300 ms window, 30 ms bins; **D** 400 ms window, 40 ms bins. For each parameter set, the following panels are shown (all aligned to stimulus onset): **i, ii** Population time course of semi-partial R^2 (SPR) values for SS (stimulus-sensitive) and WSS (weighted stimulus-sensitive) neurons in PPC (**i**) and A1 (**ii**). **iii, iv** Population time course of SPR values for sumSS (summed stimulus-sensitive) and sumWSS (summed weighted stimulus-sensitive) neurons in PPC (**iii**) and A1 (**iv**). **v, vi** SPR values for WSS and sumWSS neurons in PPC (**v**) and A1 (**vi**). **vii, viii** SPR values for SS and sumSS neurons in PPC (**vii**) and A1 (**viii**). Colored bars above the x-axis indicate time bins with significant differences: blue for PPC versus baseline (paired *t*-test, FDR-corrected $q < 0.01$); red for A1 versus baseline (paired *t*-test, FDR-corrected $q < 0.01$); gray for PPC versus A1 (*t*-test, FDR-corrected $q < 0.01$).