Supplement

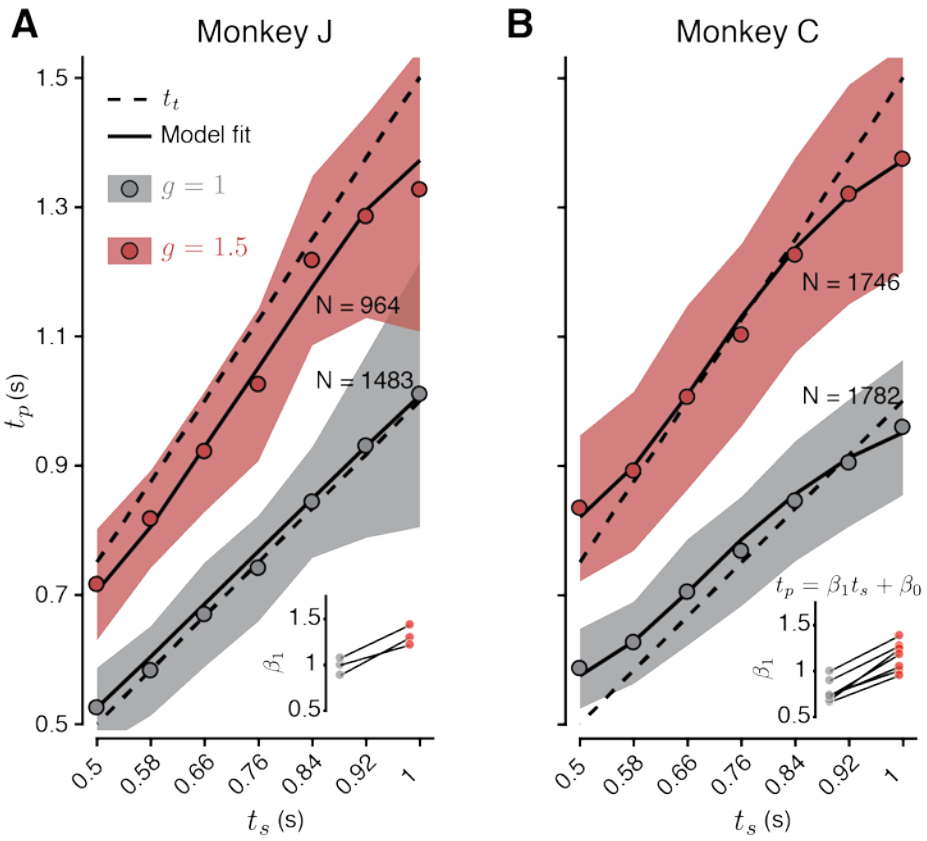

Figure S1. Behavior in the RSG task for individual animals. Related to **Figure 1**, format like **Figure 1E**. In Monkey J, we tested additional conditions (e.g. different target locations) in the $g = 1$ context leading to a larger number of trials in that context.

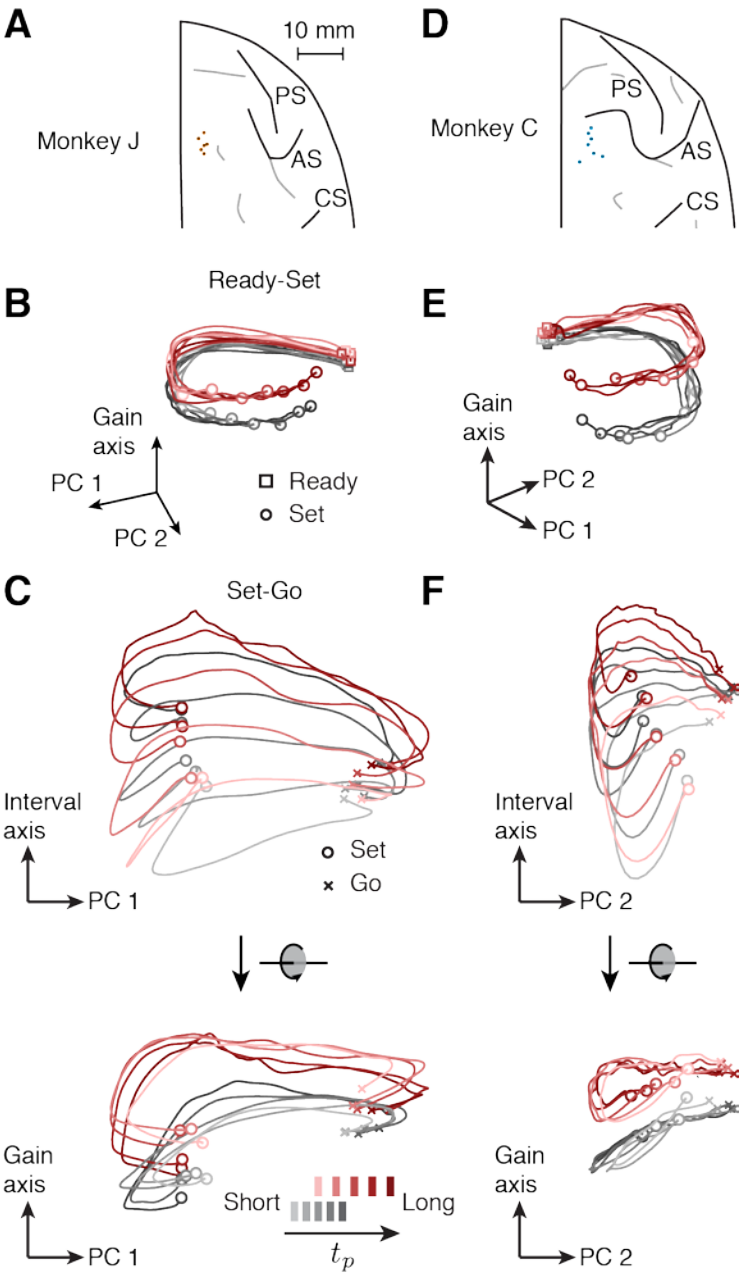

Figure 2. DMFC responses of individual animals. Related to **Figure 2**. (**A,D**) Recording locations for the two monkeys reconstructed based on individual structural MRI scans (PS, principal sulcus; AS, arcuate sulcus; CS, central sulcus). (**B,E**) Population activity in the Ready-Set epoch for individual animals formatted like **Figure 2B**. (**C,F**) Population activity in the Set-Go epoch for individual animals formatted like **Figure 2C**.

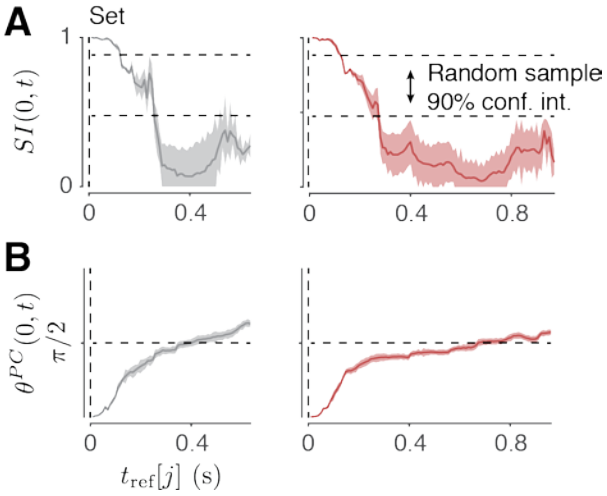

Figure S3. Rotation of trajectories through time, related to **Figure 4**. We estimated the degree to which the principal axes (PC directions) associated with nearest states along the five trajectories , $\{s_i[j]\}_i$, changed with time relative to $t=0$ using two metrics: a similarity index ($SI(0,t)$) that measures the variance explained by PCs at time t and $t=0$ (see below for full description), and a rotation index ($\theta_{PC_1}(0,t)$) measuring the angle of the first PC (*PC1*) in the state space at time t compared to $t=0$. (**A**) $SI(0,t)$. This index varies between 0 and 1 with 1 signifying matching PCs and 0 signifying orthogonal PCs. The gradual change in $SI(0,t)$ away from 1 and toward 0 indicated that $\{\Omega_i\}_i$ gradually changed orientation with time. Shaded area represents 90% bootstrap confidence intervals (n = 100). Dashed lines represent the 90% confidence intervals for the similarity of two sets of $\{s_i[j]\}_i$ drawn randomly from a multivariate Gaussian distribution with covariance matched to the data. $SI(0,t)$ captures the extent to which the orientation of $\{s_i[j]\}_i$ in state space changes with time and is therefore sensitive to both rotations and scaling transformations. (**B**) $\theta_{PC_1}(0,t)$. The gradual change in $\theta_{PC_1}(0,t)$ away from 0 toward 90 deg indicates that trajectories underwent rotations through state space from Set to Go. Unlike $SI(0,t)$ that is sensitive to both rotations and scaling transformations, $\theta_{PC_1}(0,t)$ is only sensitive to rotations. These data-driven observations motivated the use of KiNeT for analyzing neural trajectories throughout the paper.

Similarity Index: The similarity index, adapted from (Garcia, 2012), was calculated using the following procedure: 1) Select two datasets, one for neural activity patterns at the time of Set ($t=0$), denoted by $\mathbf{r_0}$, and one at time t after Set, denoted by $\mathbf{r_t}$. 2) Calculate the principal component coefficients for each dataset. 3) Project the points of each dataset onto their own and the others' principal coefficients, creating four sets of principal component scores. 4) Calculate the fraction of variance explained by each principal component in each of the four sets of scores. $\sigma_{0,0}^{2,i}$ is the fraction of variance in $\mathbf{r_0}$ explained by principal component i of $\mathbf{r_0}$,

$\sigma_{t,0}^{2,i}$ is the fraction of variance of $\mathbf{r_t}$ explained by principal component i of $\mathbf{r_0}$, $\sigma_{t,t}^{2,i}$ is the fraction of variance in $\mathbf{r_t}$ explained by principal component i of $\mathbf{r_t}$, and $\sigma_{0,t}^{2,i}$ is the fraction of variance in $\mathbf{r_0}$ explained by principal component i of $\mathbf{r_t}$. 5) For each component of each dataset, calculate the difference between (1) the fraction of variance explained by that component for its own dataset (e.g. $\sigma_{0,0}^{2,i}$) and (2) the fraction explained by that same component for the other dataset (e.g. $\sigma_{t,0}^{2,i}$). 6). Sum and normalize the calculated differences. This can be written as follows:

$$S(0,t) = 1 - 1/4 \sum_i (|\sigma_{0,0}^{2,i} - \sigma_{t,0}^{2,i}| + |\sigma_{t,t}^{2,t} - \sigma_{0,t}^{2,i}|)$$

The similarity index is 0 when the associated covariance matrix of one dataset lies in the nullspace of the other, and 1 when the covariance matrices are identical.

In order to interpret the values of similarity index in the DMFC dataset, we compared similarity index for two surrogate datasets that matched the statistics of DMFC activity. Each dataset was constructed by drawing five samples (the number of t_p bins) from a ten-dimensional Gaussian distribution (the number of principal components) with a diagonal covariance matrix constructed using the eigenvalues of the covariance matrix of the DMFC data. We calculated the similarity index for 1000 pairs of surrogate data (i.e., null distribution), and used the 5th and 95th percentiles to generate 90% confidence intervals. With this procedure, a similarity index above the 90% confidence interval was considered more “similar” than expected by chance, whereas a similarity index below the 90% confidence interval was considered dissimilar.

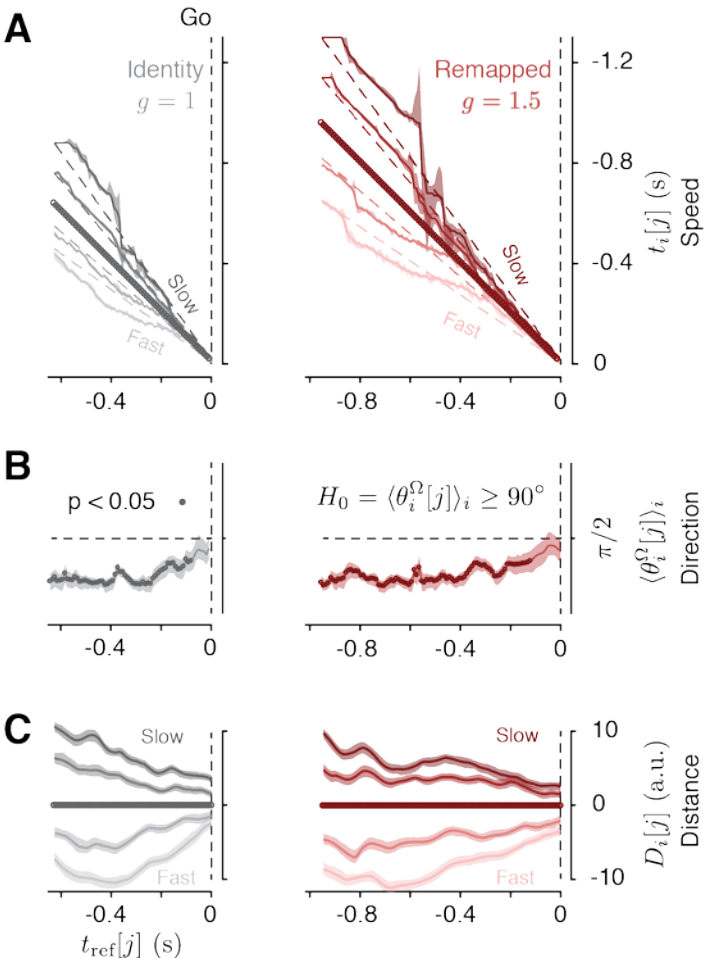

Figure S4. "Go"-aligned KiNeT, related to **Figure 4**. Applying KiNeT to neural trajectories aligned to the Set cue resulted in $t_i[j]$ which diverged from $t_{\mathrm{ref}}[j]$ to scale with trajectory length in a manner consistent with neural speed control as a means to produce different t_p. To rule out the possibility that this temporal scaling of trajectories was an artifact of temporal smearing of PSTHs near the time of Go caused by averaging trials of different lengths, we applied KiNeT to data aligned to Go (saccade). (**A**). Aligned times (speed) across both contexts. As in the Set-aligned analysis, $t_i[j]$ for shorter Ω_i diverged to shorter values, while $t_i[j]$ for longer Ω_i diverged towards longer values as $t_{\mathrm{ref}}[j]$ (here time before Go) increased. In contrast to the lack of temporal scaling proximal to the Set cue, $t_i[j]$ were ordered according to t_p leading all the way up to the Go cue. Circles on the $t_{\mathrm{ref}}[j]$ line indicate j for which the ordering of $t_i[j]$ was significantly correlated with the t_p bin (bootstrap test, r > 0.1, p < 0.05, n = 100). (**B,C**) t_p-related structure of $\{\Omega\}_i$ (**B**). Analysis of direction. As in the Set-aligned KiNeT, $\left\langle \theta_i^{\Omega}[j] \right\rangle_i$ was significantly smaller than 90 degrees for the majority of the Set-Go interval (bootstrap test, $\left\langle \theta_i^{\Omega}[j] \right\rangle_i$ < 90, p < 0.05, n = 100) indicating that $\Delta_i^{\Omega}[j]$ were similar, across Ω_i. (**C**) Analysis of distance. Euclidean distance to Ω_{ref}. Trajectories were ordered in neural space according to $D_i[j]$, where Ω_i with t_p more similar to the middle t_p bin to being located closer to Ω_{ref}. Significance tested by

counting the number of times in which $D_i[j]$ was not ordered according to t_p bin in bootstrap samples for each j ($p < 0.05$, $n = 100$).

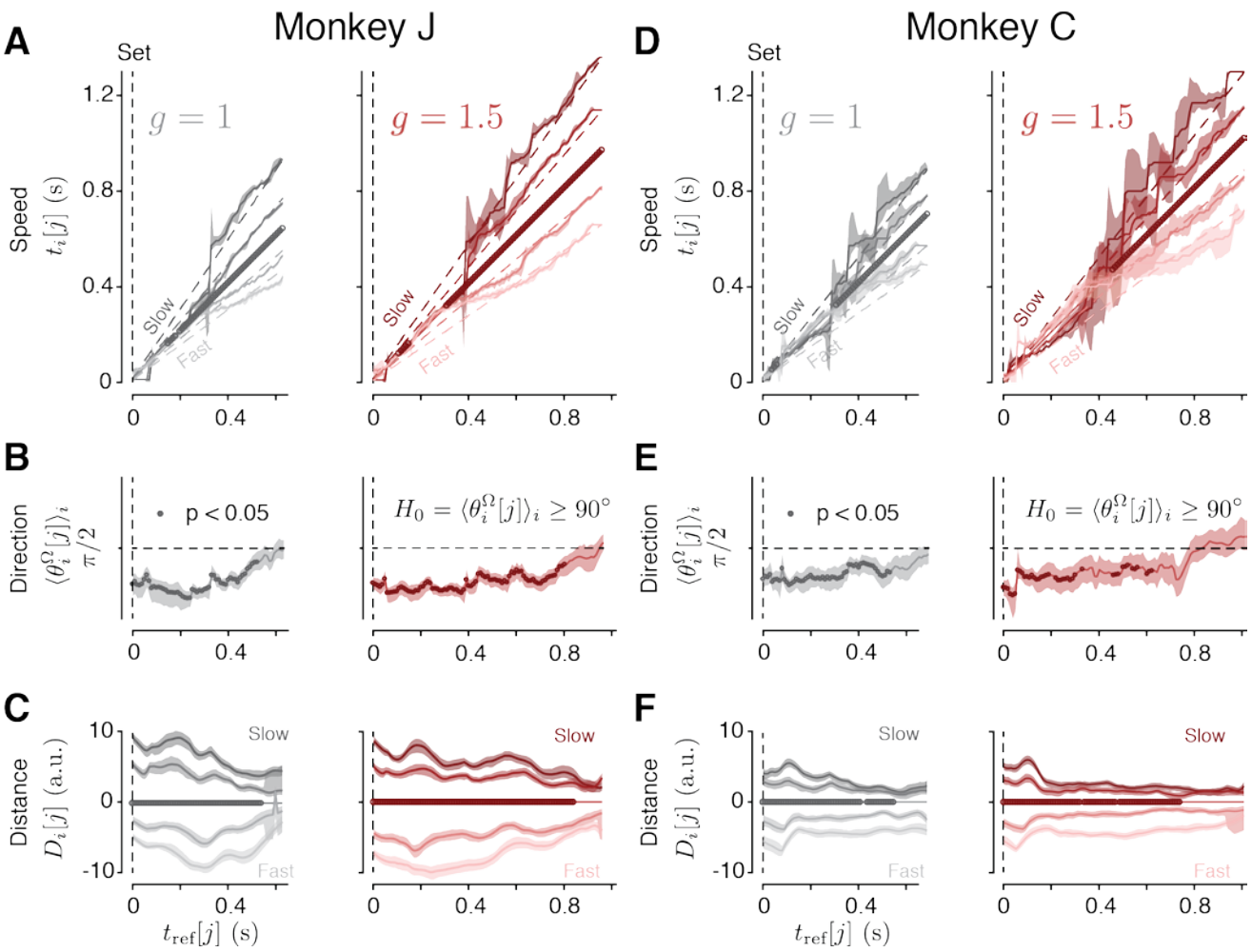

Figure S5. Kinematic analysis of neural trajectories (KiNeT) for individual animals. Related to **Figure 4**; Format like **Figure 4B-D**. (**A-C**). Monkey J, N = 195 units. (**D-F**). Monkey C, N = 129.

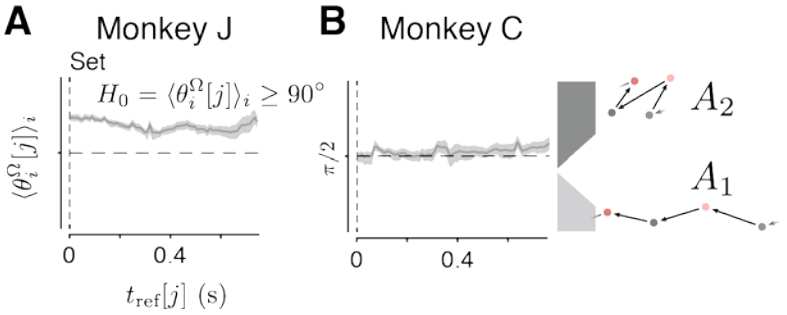

Figure S6: KiNeT across contexts for individual subjects. Related to **Figure 5**; format like **Figure 5B**. (**A**). Monkey J. (**B**). Monkey C.

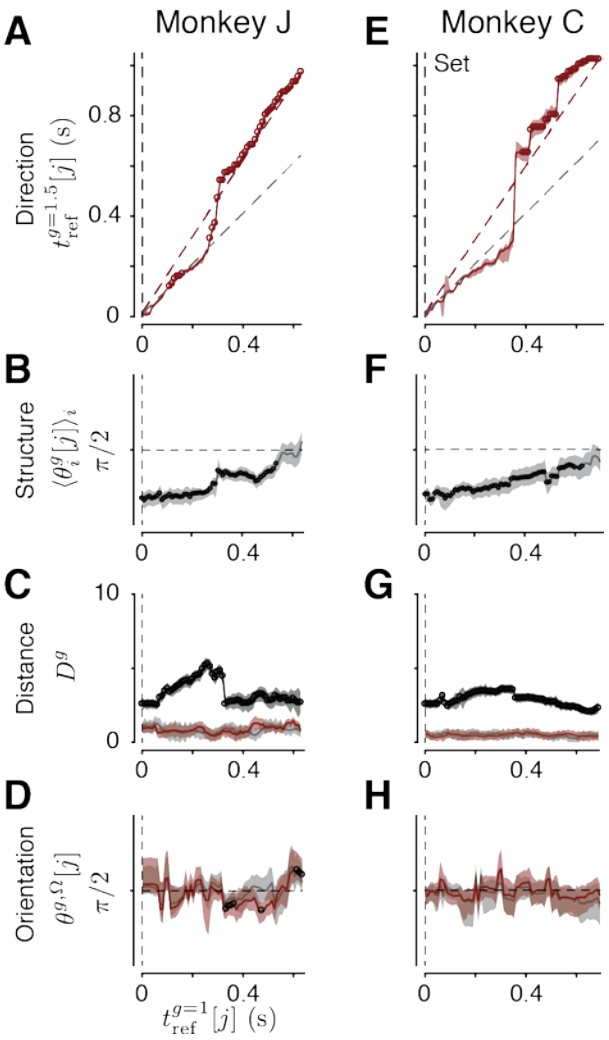

Figure S7. KiNeT applied across gains, individual animals. Related to **Figure 6**; format like **Figure 6C-F**. (**A-D**) Monkey J. (**E-H**). Monkey C.

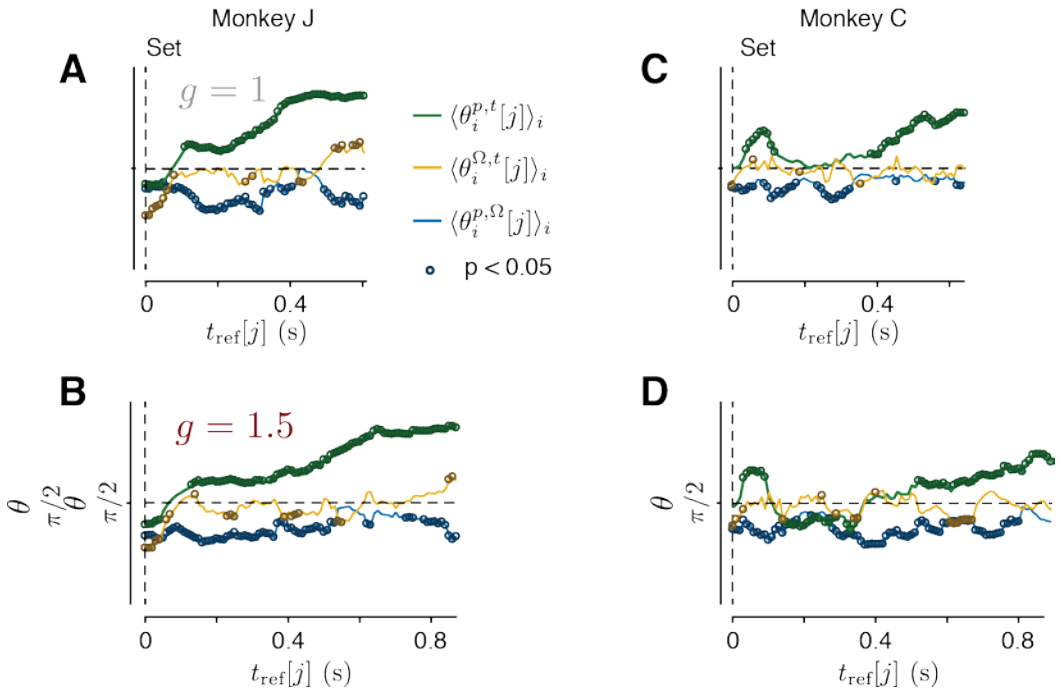

Figure S8. Relating neural variability to behavioral variability in individual animals. Related to **Figure 7**; format like **Figure 7D,E**. (**A**,**B**) Results of Monkey J for $g = 1$ (**A**) and $g = 1.5$ (**B**), N = 178 Units. (**C**,**D**) Results of Monkey C for $g = 1$ (**C**) and $g = 1.5$ (**D**), N = 95.

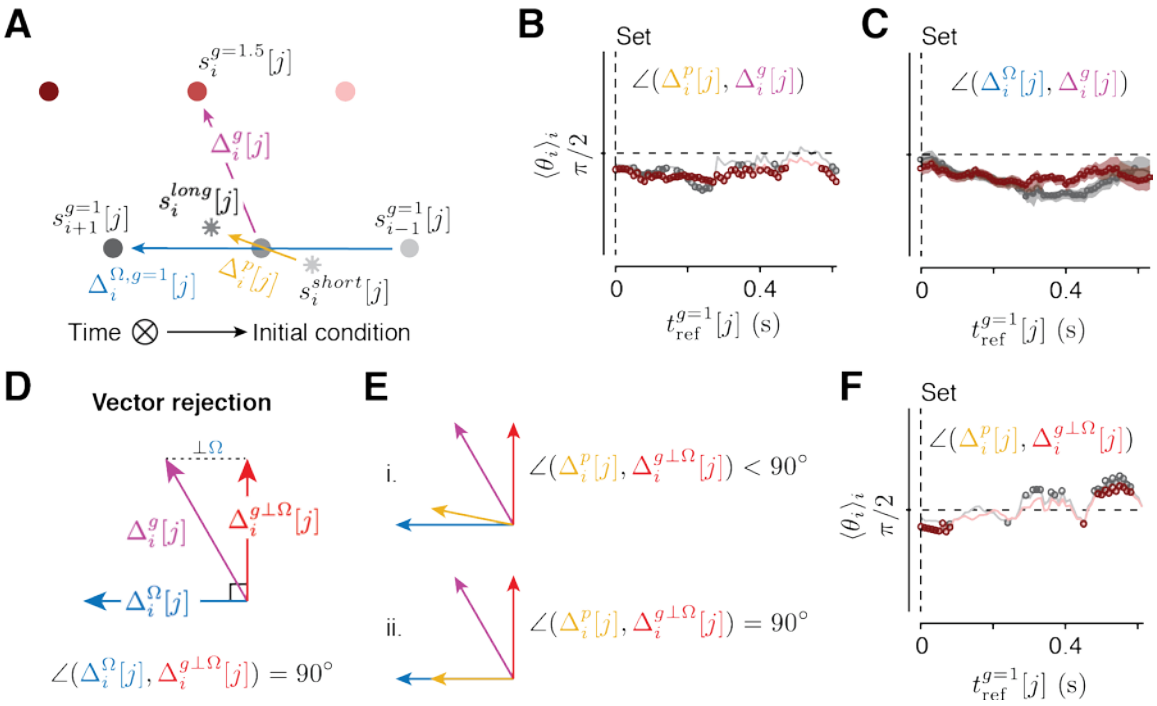

Figure S9. Alignment between behavioral- and context-related neural variability. Related to **Figure 8**. **A**. Schematic illustrating neural trajectories for $g=1$ (gray filled circles) and $g=1.5$ (red filled circles) with the time axis normal to the plane of illustration. Blue arrow illustrates $\Delta_i^{\Omega}[j]$ for $g=1$, and the yellow arrow illustrates $\Delta_i^{p}[j]$. Both are calculated as in **Figure 8**. The purple arrow illustrates $\Delta_i^{g}[j]$, which is defined here as the vector connecting trajectories for matching intervals across contexts. We asked whether neural variability associated with longer t_p ($\Delta_i^{p}[j]$) was aligned to the gain vector ($\Delta_i^{g}[j]$). **B**. Angle between $\Delta_i^{g}[j]$ and $\Delta_i^{p}[j]$ (gray and red for $g=1$ and $g=1.5$, respectively). Average angles were less than 90 deg for the majority of the Set-Go epoch suggesting that some behavioral variability might be explained by variability along the gain vector. **C**. Alignment between $\Delta_i^{g}[j]$ and $\Delta_i^{\Omega}[j]$. Angles between vectors $\Delta_i^{\Omega}[j]$ and $\Delta_i^{g}[j]$ (i.e., displacements of trajectories with interval and context) were also less than 90 deg for most of the Set-Go epoch. This makes it difficult to interpret whether the alignment measured in **B** was spurious. In other words, it could be the case that $\Delta_i^{p}[j]$ is primarily aligned with $\Delta_i^{\Omega}[j]$, and only aligned with $\Delta_i^{g}[j]$ through $\Delta_i^{\Omega}[j]$. **D**. Orthogonalization of $\Delta_i^{g}[j]$ with respect to $\Delta_i^{\Omega}[j]$. To test whether $\Delta_i^{p}[j]$ was aligned with $\Delta_i^{g}[j]$ independently from $\Delta_i^{\Omega}[j]$, we calculated a vector $\Delta_i^{g\perp\Omega}[j]$ (red arrow**)**, which was the vector rejection of $\Delta_i^{g}[j]$ from $\Delta_i^{\Omega}[j]$. **E**. Possible configurations. i. $\Delta_i^{p}[j]$ is aligned with $\Delta_i^{g}[j]$ independently from $\Delta_i^{\Omega}[j]$. ii. $\Delta_i^{p}[j]$ is aligned with $\Delta_i^{g}[j]$ only through $\Delta_i^{\Omega}[j]$. **F**. Alignment between $\Delta_i^{g\perp\Omega}[j]$ and $\Delta_i^{p}[j]$. Angles between these two vectors was close to 90 deg through the Set-Go epoch. We interpret this result as a lack of evidence that variability in the context-dependent input contributed to behavioral variability in the RSG task. However, this is not surprising given that the gain varied in blocks and was given through clear visual cues throughout each trial.

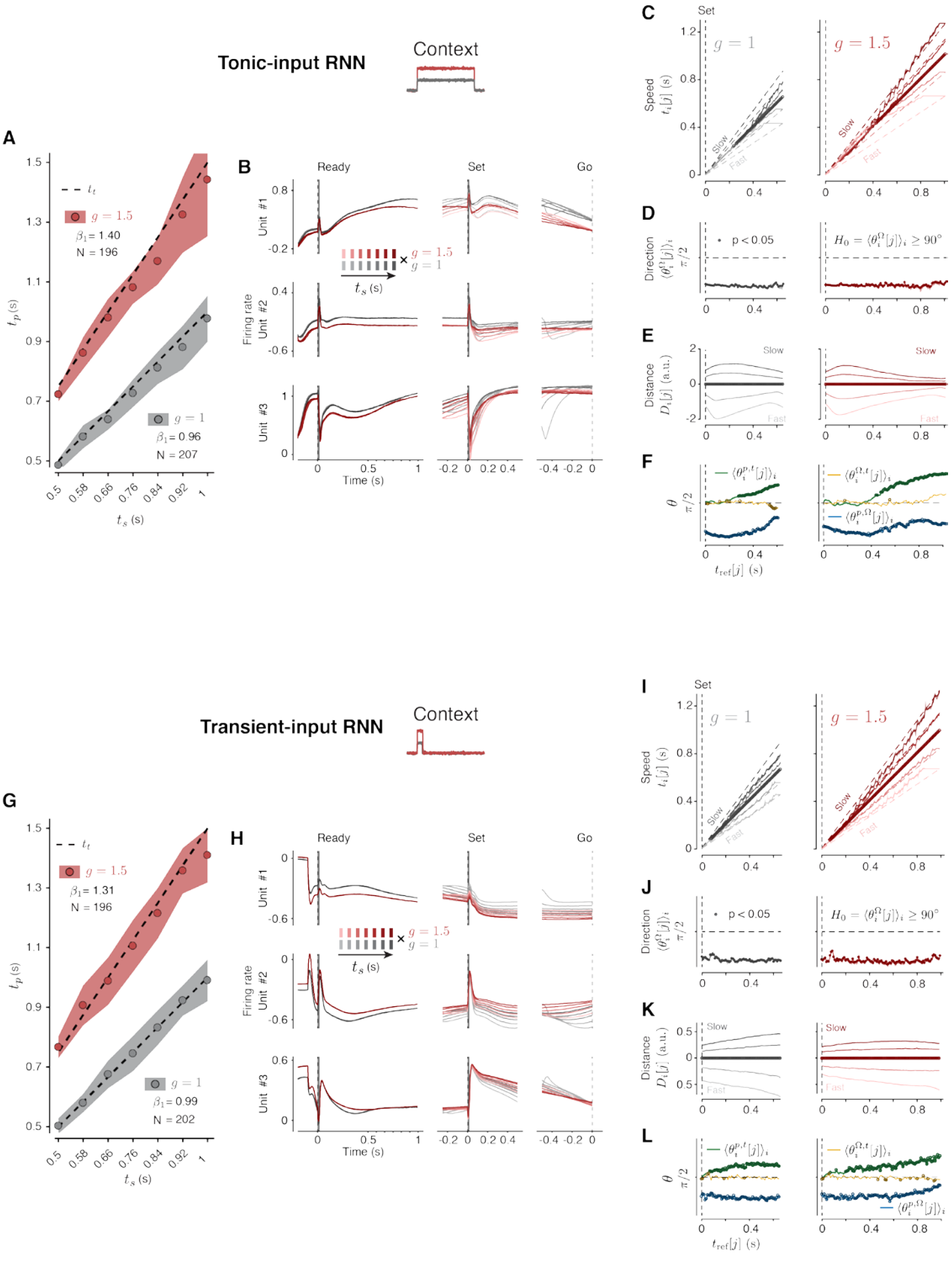
Tonic-input RNN
Context
A
B
C
D
E
F
Ready
Set
Go
Time (s)
Unit #1
Unit #2
Unit #3
Firing rate
Speed
Direction
Distance
Slow
Fast
$g = 1.5$
$g = 1$
$\beta_1 = 1.40$
N = 196
$\beta_1 = 0.96$
N = 207
t_s (s)
t_p (s)
t_t
p < 0.05
$H_0 = \langle \theta_i^{\Omega}[j] \rangle_i \geq 90°$
$t_{\text{ref}}[j]$ (s)
Transient-input RNN
G
H
I
J
K
L
$\beta_1 = 1.31$
N = 196
$\beta_1 = 0.99$
N = 202

Figure S10. Analysis of the recurrent neural networks (RNNs), related to **Figure 4** and **Figure 8**. (**A,G**) "Behavior" of tonic (**A**) and transient (**G**) networks. Same format as in **Figure 1E**. The networks successfully learned the task as evidenced by positive regression slopes (β_1, larger for $g = 1.5$ context) when analyzing contexts independently, and a significant positive interaction between t_s and g ($p << 0.001$) when analyzing contexts together. For each network, we simulated 30 trials per t_s and g, removing outliers in which t_p was more than 3.5 times the median absolute deviation (MAD) away from the mean. (**B,H**) Single unit examples. Three of the 200 units from each network are plotted in the same format as **Figure 2A**. Gray shaded regions show the duration of the Ready and Set input pulses (10 ms duration). Population analyses were carried out on activity immediately following the offset of the Set pulse. (**C-E, I-K**) Organization of neural trajectories within each context; same format as **Figure 4B-D**. KiNeT analysis verified that the organization of neural trajectories within contexts in the RNNs qualitatively matched the key features observed in DMFC. Specifically, trajectories were organized in state space according to t_p and evolved at different speeds (compare to **Figure 4B-D**). Some features observed in the DMFC data were not consistently observed in all networks. As observed in the DMFC data (**Figure 4B**), KiNeT did not detect t_p-dependent speed changes in the early part of the set interval for the tonic network (**C**). In contrast, t_p-dependent speed changes were detected substantially earlier for the transient network (**I**). However, the onset as measured by KiNeT was variable across networks and did not appear to be correlated with input type (i.e. tonic vs. transient). Similarly, trajectories of the tonic network appeared to coalesce near Go (**E**), like the neural data (**Figure 4D**) but unlike the transient network (**K**). As with the onset of variable speed, this feature was variable across networks and was not clearly associated with input type. (**F,L**) Relating unit variability to behavioral variability; same format as in **Figure 7**.

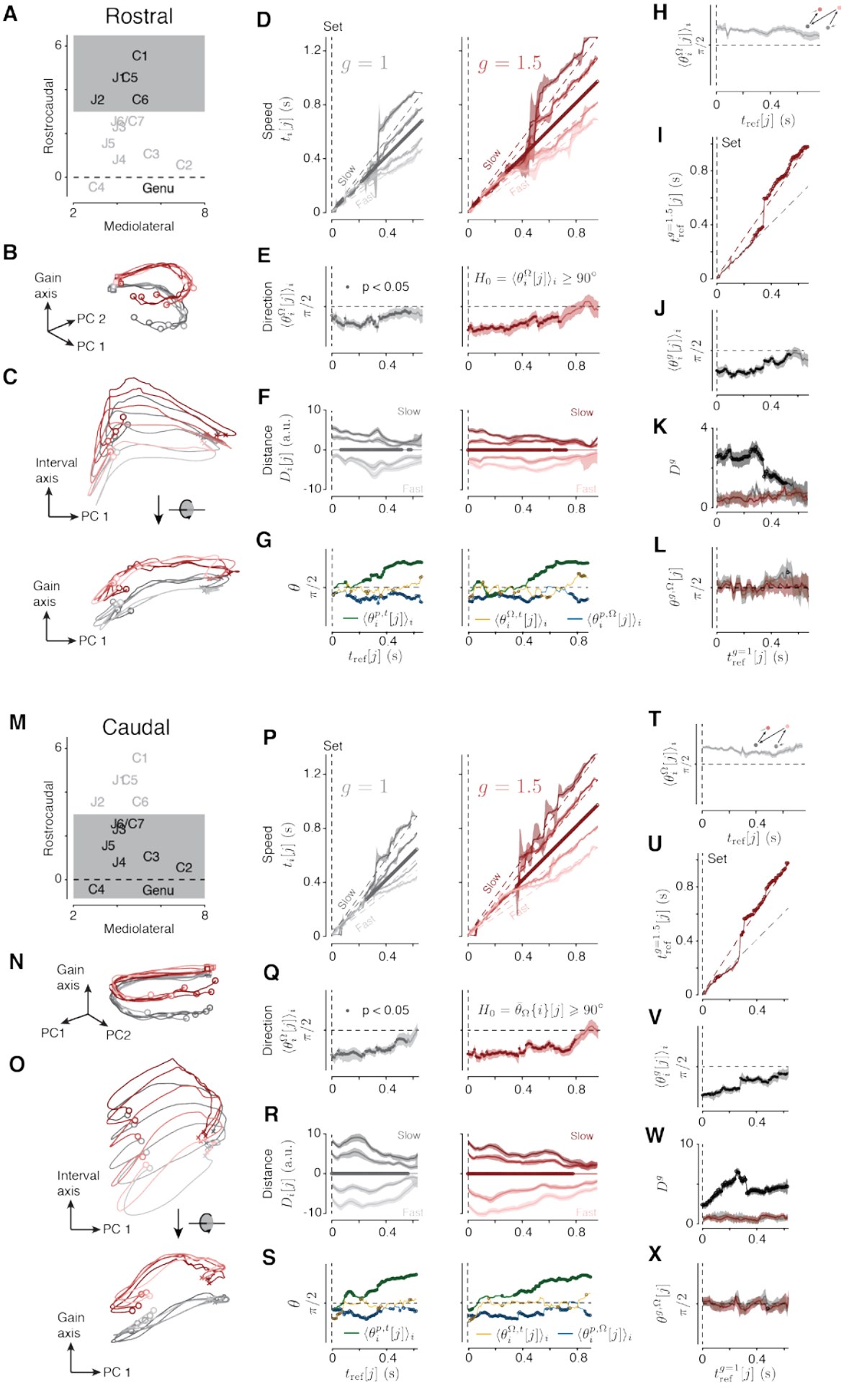
A
Rostral
Rostrocaudal
Mediolateral
C1
J1
C5
J2
C6
J6/C7
J3
J5
J4
C3
C2
C4
Genu
B
Gain axis
PC 2
PC 1
C
Interval axis
PC 1
Gain axis
PC 1
D
Set
$g = 1$
$g = 1.5$
Speed $t_i[j]$ (s)
Slow
Fast
E
Direction $\langle\theta_i^{\Omega}[j]\rangle_i$
$\pi/2$
p < 0.05
$H_0 = \langle\theta_i^{\Omega}[j]\rangle_i \geq 90°$
F
Distance $D_i[j]$ (a.u.)
G
θ
$\langle\theta_i^{p,t}[j]\rangle_i$
$\langle\theta_i^{\Omega,t}[j]\rangle_i$
$\langle\theta_i^{p,\Omega}[j]\rangle_i$
$t_{\mathrm{ref}}[j]$ (s)
H
$\langle\theta_i^{\Omega}[j]\rangle_i$
$t_{\mathrm{ref}}[j]$ (s)
I
Set
$t_{\mathrm{ref}}^{g=1.5}[j]$ (s)
J
$\langle\theta_i^{g}[j]\rangle_i$
K
D^g
L
$\theta^{g,\Omega}[j]$
$t_{\mathrm{ref}}^{g=1}[j]$ (s)
M
Caudal
N
Gain axis
PC1
PC2
O
Interval axis
PC 1
Gain axis
PC 1
P
Q
R
S
T
U
V
W
X

Figure S11. Analysis of Rostral vs. Caudal recording locations, related to **Figures 2,4-7**. (**A, M**) All recording locations, labelled by subject, with rostral (**A**) or caudal (**L**) locations highlighted in gray. (**B,C, N,O**) Neural population trajectories; format like **Figure 2B,C**. (**D-F, P-R**) KiNeT analysis of neural population trajectories within contexts; format like **Figure 4B-D**. (**G, S**). Comparison of neural and behavioral variability; format like **Figure 7D,E**. (**H, T**). KiNeT across contexts rejecting single structure; format like **Figure 5B**. (**I-K, U-X**) KiNeT applied across contexts showing similar, separate structures; format like **Figure 6C-F**. For rostral locations, panels **B-F** and **H-L**, N = 92 units (40 for monkey J, 52 for monkey C). For **G**, N = 86 (40 for monkey J, 46 for monkey C). (**M-X**) For caudal locations, panels **N-R** and **T-X**, N = 232 units (155 for monkey J, 77 for monkey C). For **S**, N = 187 (138 for monkey J, 49 for monkey C).

Symbol	Description
Ω_i	The i-th neural trajectory
$\{\Omega_i\}_i$	A collection of neural trajectories
Ω_{ref}	“Reference” neural trajectory
Ω_1	The trajectory of shortest duration
Ω_N	The trajectory of longest duration
t_{ref}	Elapsed time for j-th time bin on Ω_{ref}
$s_{\text{ref}}[j]$	Neural state on Ω_{ref} at t_{ref}
$s_i[j]$	Neural state on Ω_i with minimum distance to $s_{\text{ref}}[j]$
$\{s_i[j]\}_j$	$s_i[j]$ across all time bins
$\{s_i[j]\}_i$	$s_i[j]$ on all trajectories at j-th time bin
$t_i[j]$	Elapsed time on Ω_i at $s_i[j]$
$\{t_i[j]\}_j$	Elapsed time on Ω_i across all time bins
$\{t_i[j]\}_i$	Elapsed time on all trajectories at j-th time bin
$D_i[j]$	Euclidean distance between $s_i[j]$ and $s_{\text{ref}}[j]$
$\Delta_i^{\Omega}[j]$	Vector between nearby trajectories. For analyses of **Figure 4-6**, $\Delta_i^{\Omega}[j]$ connects $s_i[j]$ to $s_{i+1}[j]$, $1 \le i \le 4$. For analyses of **Figure 7**, see above section "Calculation of for comparing variability within and across neural trajectories"
$\theta_i^{\Omega}[j]$	Angle between $\Delta_i^{\Omega}[j]$ and $\Delta_{i+1}^{\Omega}[j]$, $1 \le i \le 3$
$\theta_i^{\Omega}[j]$	Average of $\theta_i^{\Omega}[j]$ across i for the j-th time bin

$\theta_i^g[j]$	Angle between $\Delta_i^{\Omega,g=1}[j]$ and $\Delta_i^{\Omega,g=1.5}[j]$, $1 \leq i \leq N-1$

Table S1. Abridged symbol key.