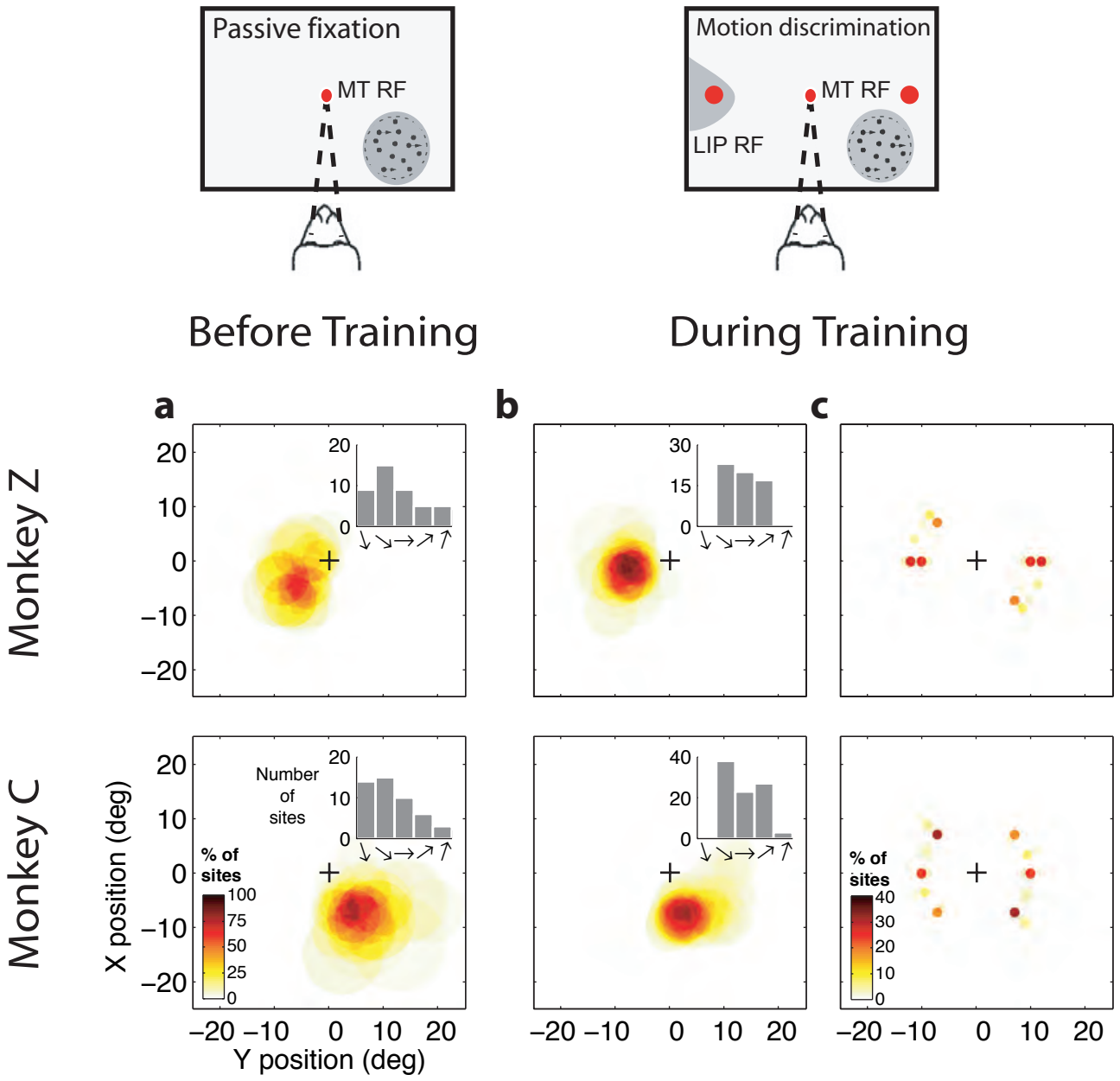

Figure S1. Task geometry. The parameters of the visual stimulus used in a given session matched the properties of the MT and/or LIP neurons recorded during that session. To maintain consistent visual stimuli across sessions, we searched for neurons with similar tuning properties. “Before training” (**a**) refers to sessions in which the monkeys were rewarded for simply fixating a central spot while the motion stimulus was shown. “During training” (**b and c**) refers to sessions in which the monkeys were trained to perform the direction-discrimination task. **a and b**. MT receptive field locations (modal locations = 2.3° horizontal, -8.7° vertical relative to fixation for monkey C and -7.5°, -2.7° for monkey Z) and directional tuning (inset). **c.** LIP response field locations. Note that the full spatial extent of each response field was not mapped; rather, the locations shown here correspond to the angular positions at 10±0.9° (mean±s.d., monkey C) or 11±1.3° (monkey Z) eccentricity that elicited the strongest, spatially tuned responses during the delay period of a memory-saccade task.

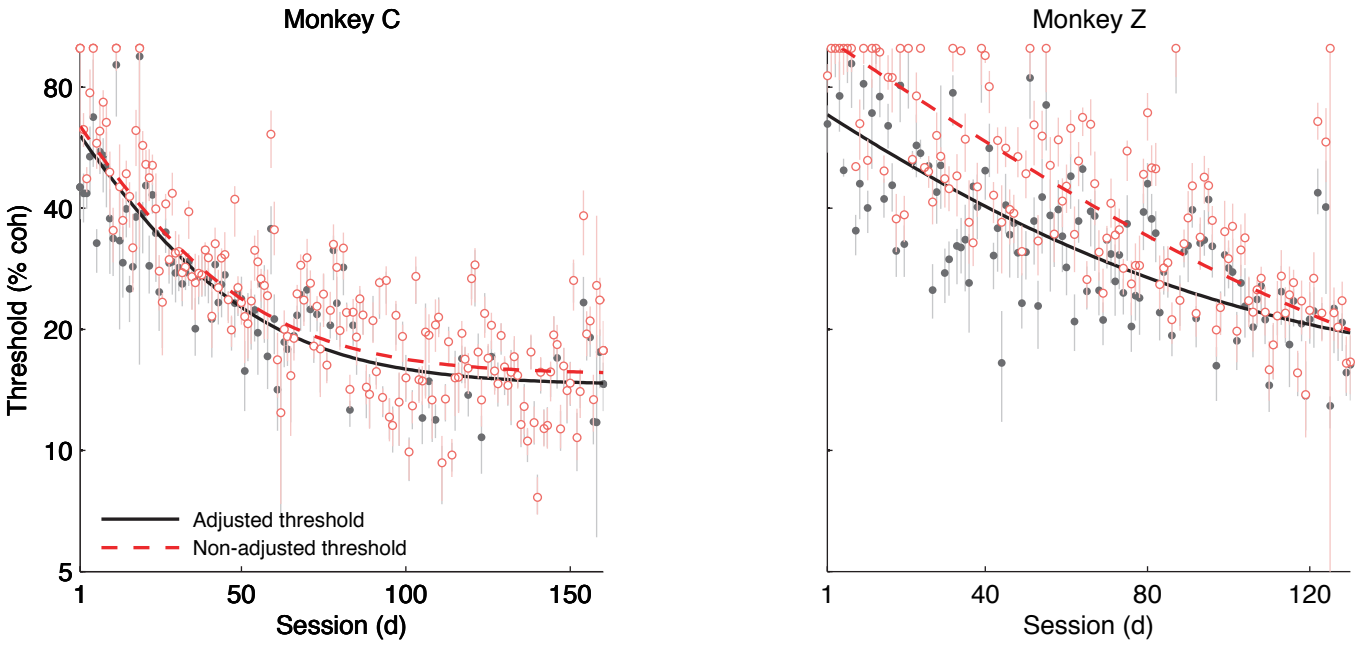

Figure S2. Behavioural thresholds computed with (black) and without (red) adjustments for associative (high-coherence) errors. Behavioural threhsolds were computed using a time-dependent cumulative Weibull function (Eq. 1 and Fig. 2). The adjustment for high-coherence errors (λ in Eq. 1) assumes that these errors do not reflect limitations in perceptual processing but rather other problems in performing the task (these errors are often referred to as "lapses" because in trained, human subjects they are thought to be attributed to lapses in attention; here we think of them as a more general phenomenon that also incoporates incomplete knowledge of the sensory-motor association). Most importantly, the errors at high coherence are treated as an estimate of the overall percentage of errors that can be attributed to non-perceptual causes and thus are used to scale the entire psychometric function. The red curves show that thresholds computed without this adjustment tend to be slightly higher and, in the case of monkey Z, change more dramatically with training than the adjusted thresholds because they conflate associative and perceptual changes. The black curves show that even after these associative errors are taken into account, discrimination thresholds change substantially with training.

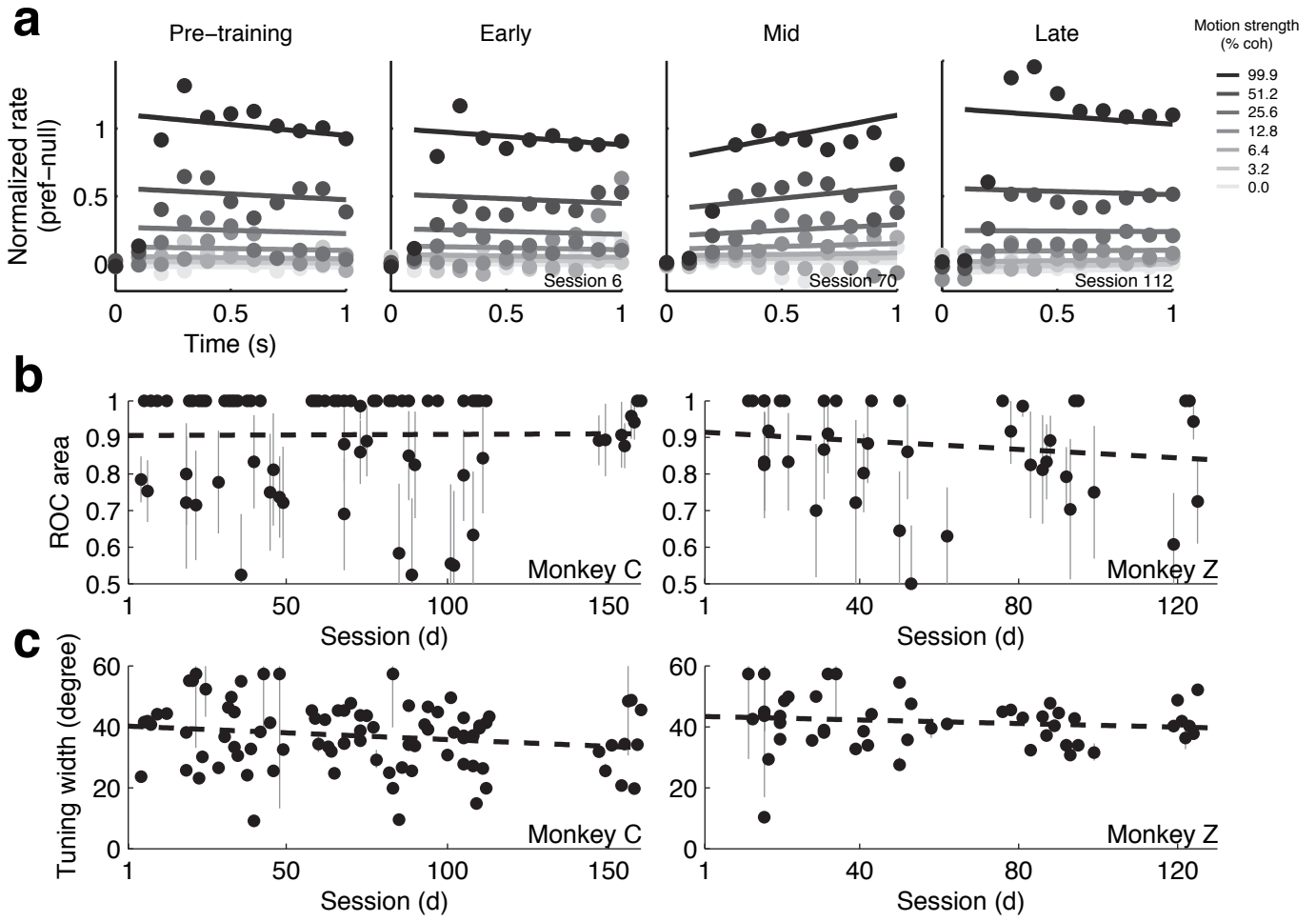

Figure S3. a, Normalized responses of MT neurons (Eq. 2) as a function of viewing time (0.2-s-wide bins in 0.1s intervals) for different motion strengths (see legend) during different training periods for monkey C. "Pre-training" refers to responses to the motion stimulus measured while the monkey was rewarded for simply fixating a central spot, before being trained on the discrimination task. Solid lines are the fitted values given by Eq. 3. **b,c.** MT directional tuning as a function of training session for Monkey C (left) and Z (right) measured using a 99.9% coherence motion stimulus during the passive-vewing task. **b.** The discriminability of MT responses for preferred versus null motion computed using an ROC analysis. **c**. Directional tuning width computed by fitting a von Mises function to the MT responses to 8 different directions of motion. In all four panels, the data did not depend significantly on session number (linear regression, H_0: slope=0, **b,** p=0.9334 for monkey C, p=0.2842 for monkey Z; **c,** p=0.0833 for monkey C, p=0.3798 for monkey Z).

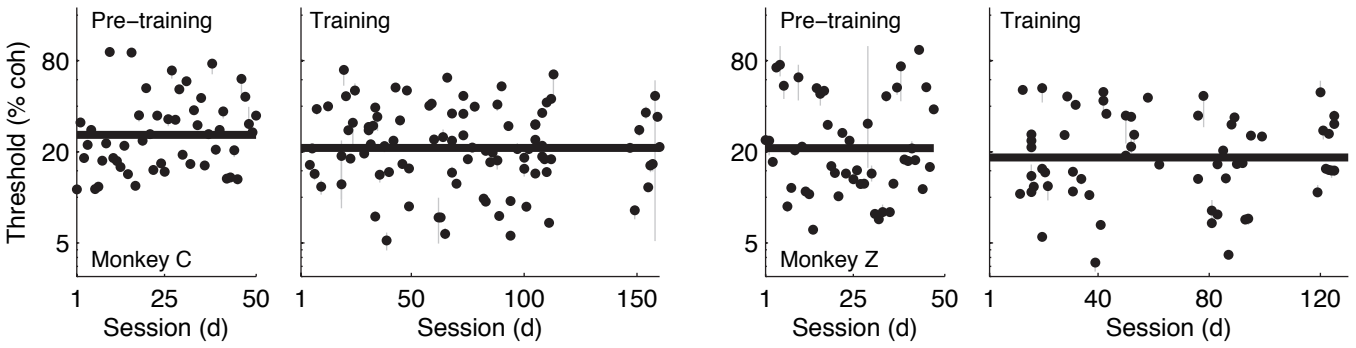

Figure S4. Neurometric thresholds of individual MT neurons measured before and during training for monkeys C (left) and Z (right). Error bars are 68% CIs. Solid lines indicate the geometric mean threshold coherence before (left) and during (right) training. The neurometric thresholds were computed using a time-dependent cumulative Weibull function (Eq 1, see Methods).

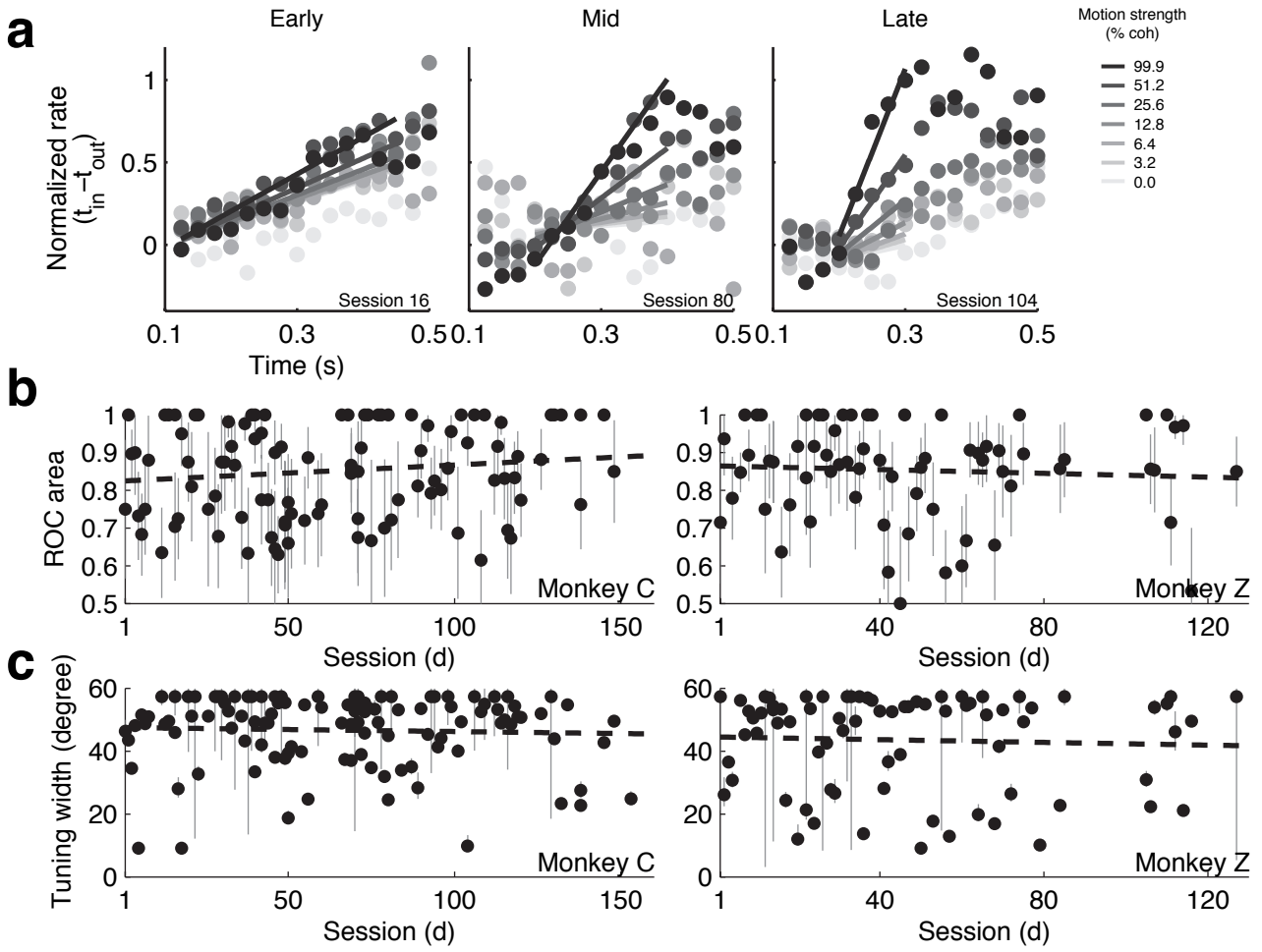

Figure S5. a. Normalized responses of LIP neurons (Eq. 2) as a function of viewing time (0.5-s-wide time bins with 0.025-s increments) for different motion strengths (see legend) during different training periods for monkey C. Only correct trials were included. Solid lines are fits to Eq. 3. **b,c.** LIP spatial tuning as a function of training session for monkey C (left) and Z (right). **b.** Predictive index of LIP responses during the delay period of the direction-discrimination task. This measures how well an ideal obserer could identify the monkey's subsequent choice of saccadic target based only on the responses of the LIP neuron measured between motion offset and fixation offset. **c.** Spatial tuning width computed by fitting a von Mises function to the LIP responses measured during the delay period of a memory-saccade task using 8 different saccade target directions. In all four panels, the data did not depend significantly on session number (linear regression, H_0: slope=0, **b,** p=0.1743 for monkey C, p=0.6019 for monkey Z; **c,** p=0.6500 for monkey C, p=0.6900 for monkey Z).

Supplementary Table 1

Monkey C		Pre-training	Sessions 1-53	Sessions 54-105	Sessions 106-160	Slope (95% CI)[1]
Behaviour	Eye drift during fixation (deg visual angle)	0.09 (0.40)	0.16 (0.03)	0.21 (0.11)	0.17 (0.06)	-0.0001 (-0.0002, 0.0000)
	SD of eye position along the axis of motion (deg visual angle)	0.21 (0.22)	0.10 (0.08)	0.14 (0.07)	0.10 (0.04)	0.0000 (0.0000, 0.0000)
	SD of eye position orthogonal to the axis of motion (deg visual angle)	0.21 (0.17)	0.12 (0.03)	0.15 (0.09)	0.13 (0.05)	-0.0001 (-0.0003, 0.0001)
	Peak saccade velocity (deg visual angle/s)	N/A	252 (17)	253 (12)	273 (14)	0.1937 (0.1391, 0.2483) *
	Average saccade velocity (deg visual angle/s)	N/A	163 (14)	161 (14)	174 (11)	0.1147 (0.0695, 0.1599) *
	Median saccade latency (ms)	N/A	173 (16)	186 (7)	184 (10)	0.0144 (-0.0022, 0.0012)
	Saccade-to-target accuracy (deg visual angle)[3]	N/A	-0.03 (0.45)	-0.12 (0.49)	-0.07 (0.56)	-0.0005 (-0.0234, 0.0522)
	Fitted parameter n (time exponent) from Eq. 1	N/A	-0.08 (0.50)	-0.07 (0.26)	-0.22 (0.17)	-0.0007 (-0.0028, 0.0004)
	Fitted parameter β (shape exponent) from Eq. 1	N/A	1.66 (0.45)	1.71 (0.48)	1.63 (0.39)	-0.0002 (-0.0018, 0.0014)
	Broken fixation (percent)	6.52 (4.39)	14.00 (7.70)	14.80 (5.18)	9.90 (6.27)	-0.0002 (-0.0004, 0.0000) *
	Time to attain fixation (s)	0.60 (0.02)	0.86 (0.11)	0.86 (0.12)	0.82 (0.09)	-0.0003 (-0.0005, -0.0001) *
	Reward per trial (drops of juice)	1.80 (0.06)	2.12 (0.39)	2.28 (0.37)	2.30 (0.36)	0.0012 (0.0001, 0.0023) *
MT	Width of tuning for motion direction (deg)[4,5]	39.77 (14.30)	39.10 (17.65)	37.00 (12.15)	32.98 (12.52)	-0.0465 (-0.0993, 0.0063)
	Discriminabilty of responses to preferred vs null motion[4,6]	0.70 (0.15)	0.90 (0.14)	0.91 (0.16)	0.93 (0.10)	0.0000 (-0.0070, 0.0070)
	Slope of the response to motion vs coh (spikes/s/coh)	29.04 (20.02)	33.65 (17.20)	33.44 (17.60)	23.62 (12.04)	-0.0659 (-0.1484, 0.0166)
	Fano factor during motion viewing (slope of variance vs mean)	2.59 (4.42)	2.46 (1.33)	2.07 (0.98)	3.38 (3.88)	0.0095 (-0.0006, 0.0196)
	Fitted parameter n (time exponent) from Eq. 1	-0.43 (0.28)	-0.39 (0.33)	-0.44 (0.29)	-0.59 (0.36)	-0.0013 (-0.0028, 0.0002)
	Fitted parameter β (shape exponent) from Eq. 1	2.83 (4.45)	1.90 (1.90)	1.96 (1.38)	1.76 (1.04)	-0.0017 (-0.0093, 0.0059)
LIP	Width of tuning for saccade direction (deg)[5,7]	N/A	46.90 (10.35)	46.50 (13.66)	46.52 (11.26)	-0.0127 (-0.0682, 0.0428)
	Discriminability of responses into vs away from the response field[6,7]	N/A	0.83 (0.13)	0.87 (0.12)	0.87 (0.12)	0.0005 (-0.0001, 0.0011)
	Fitted parameter τ (accumulation begin time) from Eq. 6	N/A	0.30 (0.27)	0.21 (0.12)	0.17 (0.05)	-0.0014 (-0.0023, -0.0005) *
	Discriminability of responses into vs away from the response field[6,8]	N/A	0.76 (0.17)	0.91 (0.10)	0.92 (0.13)	0.0002 (-0.0001, 0.0005)

Monkey Z		Pre-training	Sessions 1-43	Sessions 44-85	Sessions 86-130	Slope (95% CI)
Behaviour	Eye drift during fixation (deg visual angle)[2]	0.10 (0.25)	-0.25 (0.06)	-0.24 (0.07)	0.26 (0.11)	-0.0003 (-0.0006, 0.0000)
	SD of eye position along the axis of motion (deg visual angle)	0.15 (0.06)	0.18 (0.06)	0.15 (0.04)	0.18 (0.10)	0.0000 (-0.0001, 0.0001)
	SD of eye position orthogonal to the axis of motion (deg visual angle)	0.16 (0.07)	0.16 (0.04)	0.18 (0.07)	0.18 (0.06)	0.0001 (-0.0001, 0.0003)
	Peak saccade velocity (deg visual angle/s)	N/A	357 (70)	378 (82)	362 (42)	0.0777 (-0.2421, 0.3974)
	Average saccade velocity (deg visual angle/s)	N/A	198 (24)	204 (24)	205 (17)	0.0839 (-0.0195, 0.1873)
	Median saccade latency (ms)	N/A	206 (7)	202 (6)	193 (13)	-0.1463 (-0.1873, -0.1053) *
	Saccade-to-target accuracy (deg visual angle)[3]	N/A	-0.33 (0.70)	-0.18 (0.62)	-0.26 (0.62)	0.0009 (-0.0025, 0.0042)
	Fitted parameter n (time exponent) from Eq. 1	N/A	-0.45 (1.91)	-0.30 (0.39)	-0.04 (1.57)	0.0055 (-0.0011, 0.0121)
	Fitted parameter β (shape exponent) from Eq. 1	N/A	1.19 (0.43)	1.29 (0.38)	1.30 (0.41)	0.0008 (-0.0014, 0.0003)
	Broken fixation (percent)	6.63 (4.14)	17.03 (7.61)	11.81 (7.71)	5.54 (2.37)	-0.1162 (-0.1396, -0.0927) *
	Time to attain fixation (s)	0.90 (0.23)	1.08 (0.18)	1.00 (0.12)	0.90 (0.10)	-0.0009 (-0.0013, -0.0005) *
	Reward per trial (drops of juice)	1.38 (0.37)	1.94 (0.33)	2.30 (0.32)	2.31 (0.37)	0.0034 (0.0019, 0.0050) *
MT	Width of tuning for motion direction (deg)[4,5]	42.27 (13.18)	42.91 (12.36)	41.57 (9.97)	40.02 (12.21)	-0.0293 (-0.0957, 0.0371)
	Discriminabilty of responses to preferred vs null motion[4,6]	0.69 (0.14)	0.92 (0.10)	0.83 (0.16)	0.87 (0.14)	-0.0006 (-0.0017, 0.0005)
	Slope of the response to motion vs coh (spikes/s/coh)	33.73 (25.64)	40.53 (28.94)	33.41 (23.90)	43.24 (24.84)	0.0432 (-0.1394, 0.2536)
	Fano factor during motion viewing (slope of variance vs mean)	2.85 (1.92)	2.53 (2.27)	2.45 (1.30)	2.97 (1.30)	0.0116 (-0.0011, 0.0243)
	Fitted parameter n (time exponent) from Eq. 1	-0.47 (0.37)	-0.49 (0.24)	-0.43 (0.21)	-0.33 (0.26)	0.0016 (-0.0001, 0.0033)
	Fitted parameter β (shape exponent) from Eq. 1	2.63 (4.28)	1.51 (0.73)	2.06 (3.26)	2.38 (4.29)	0.0088 (-0.0119, 0.0295)
LIP	Width of tuning for saccade direction (deg)[5,7]	N/A	44.21(14.62)	42.38 (19.37)	43.77 (18.58)	-0.0217 (-0.1297, 0.0863)
	Discriminability of responses into vs away from the response field[6,7]	N/A	0.87 (0.11)	0.82 (0.14)	0.87 (0.15)	-0.0002 (-0.0011, 0.0007)
	Fitted parameter τ (accumulation begin time) from Eq. 6	N/A	0.64 (0.28)	0.54 (0.12)	0.48 (0.12)	-0.0011 (-0.0027, 0.0005)
	Discriminability of responses into vs away from the response field[6,8]	N/A	0.71 (0.13)	0.79 (0.10)	0.79 (0.11)	0.0014 (0.0006, 0.0022) *

All values are mean (SD) unless otherwise noted

[1] Slope of linear regression versus training session; * indicates H_0: slope=0, $p<0.05$

[2] Velocity of eye position in the direction of coherent motion during motion viewing.

[3] Average displacement of saccade endpoints to the choice target within a session

[4] Measured during the fixation-only task using a 99.9% coherence stimulus.

[5] The tuning width is defined as the angular deviation of the von Mises function, which is analogous to the standard deviation of a normal distribution (Ref: Batschelet, E. Circular statistics in biology, Academic Press, London, 1981).

[6] Computed using the ROC area to quantify the ability of an ideal oyserver to discriminate between the two conditions given only the spike rate data.

[7] Measured during the delay period of the delayed-saccade task.

[8] Measured during the delay period, 0.2 s before saccade initiation, of the discrimination task.