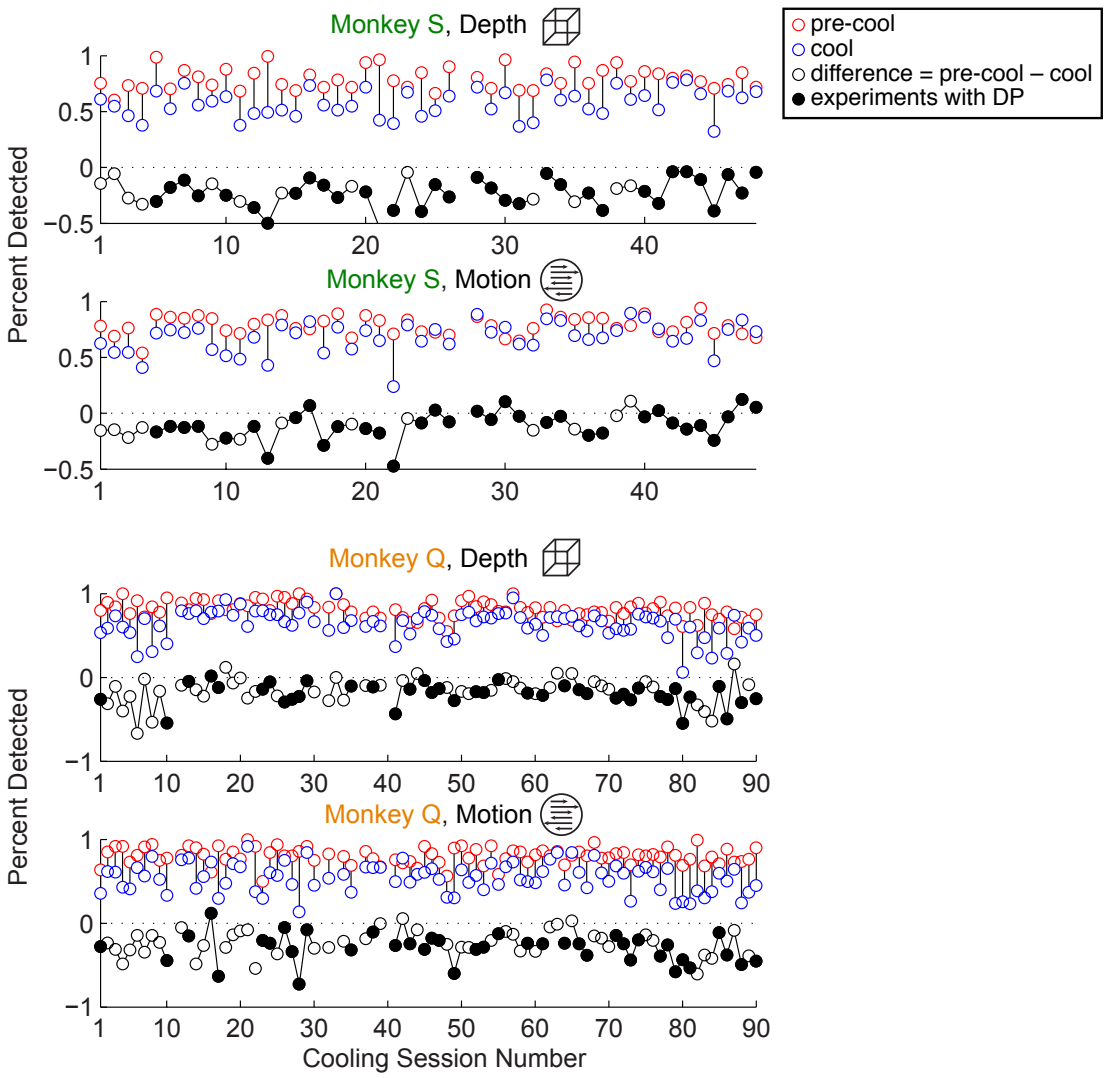

Figure S1: Behavioral performance as a function of session number (related to Figure 2)

Each panel shows the average percent correct during the pre-cool (red) and cool (blue) conditions. We used average percent correct rather than behavioral thresholds to include sessions that were not well fit by psychometric functions. The change in behavioral performance (cool-control) is shown in black, with filled markers indicating sessions from which DP data was collected. The days with unfilled black markers were not included in any of the analyses described in this manuscript.

Note for monkey Q, session #1 is the 15th session in which we cooled this animal's lunate sulcus but behavioral performance could not be measured in the earlier sessions because we varied parameters such as cooling temperature.

Monkey S (top two panels) showed no significant relationship between session number and the change in behavioral performance (difference in average percent correct) during the depth task (multiple regression coefficient = 0.0012, p = 0.4, with session number, stimulus eccentricity, size, speed, direction, and disparity as regressors) and a small but significant relationship between session number and the behavioral effect on the motion task (regression coefficient = 0.0034, p = 0.008). Monkey Q (bottom two panels) exhibited no significant relationship between experiment number and the change in behavioral threshold on either task (depth task regression coefficient = -0.00028, p = 0.68; motion task regression coefficient = -0.00086, p = 0.23). The weak dependence on experiment number suggests to us that the animals did not gradually adopt a new strategy (i.e. read-out weights) in the course of this experiment.

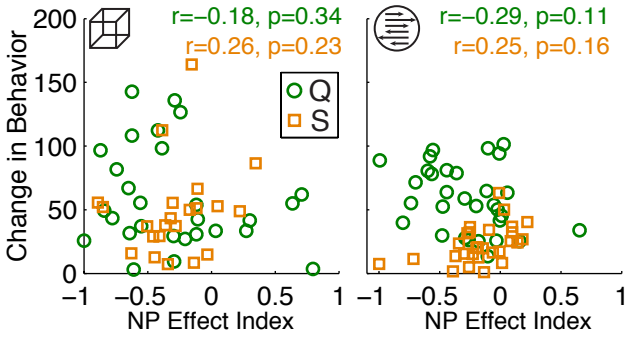

Figure S2: Relationship between changes in NP and behavior (related to figure 3)

NP effect index is plotted against changes in behavioral performance for the depth (left) and motion (right) tasks, color-coded by monkey.

There was no significant relationship between the magnitude of the NP change and the impairment in behavioral performance during the two tasks in either animal. Correlation values and concomitant p-values shown in the top right corner, color-coded by monkey.

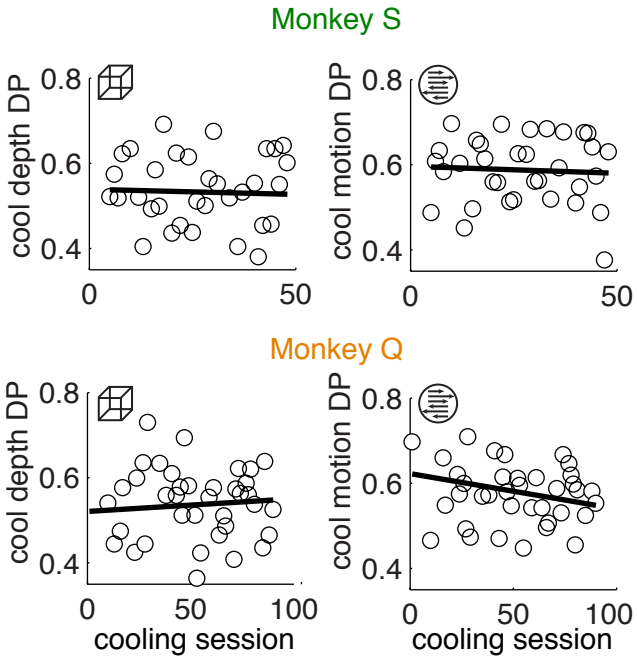

Figure S3: DP relationship with session number (related to Figure 4)

DP during cooling for the depth (left) and motion (right) tasks, shown separately by monkey (rows). Across animals and tasks the slope of this relationship was small and not significantly significant from zero (monkey S: depth slope = -0.0002, p = 0.8, motion slope = -0.0003, p = 0.7; monkey Q: depth slope = 0.0003, p = 0.61, motion slope = -0.0008, p = 0.1).

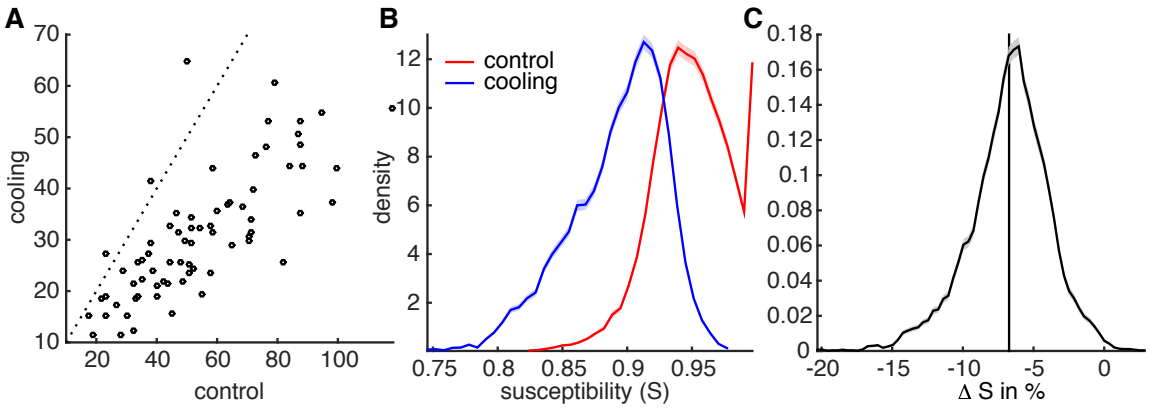

Figure S4: Effect of decrease in firing rate on response correlations in MT (related to Figure 6)

The degree to which input correlations are translated into output (response) correlations depends on the firing rate of a neuron (de la Rocha et al. 2007). This effect of firing rate is well captured by a scaling factor ('susceptibility', S) that depends on the firing rate of the neuron. We used the relationship presented in Figure 3d of de la Rocha et al. 2007 to estimate the size of the effect in our data. To do so, we first computed the distribution of geometric firing rate means separately in the control and the cooling conditions, and used these respective distributions to compute distributions over S. From that we compared the distribution over the change in S (and hence the relative change in MT response correlation) attributable purely to the mean change in firing rate with cooling (Figure 5).

Generally, the magnitude of CPs and DPs is proportional to the square root of the magnitude of the response correlations in the sensory population. If one increases the correlations (and hence the off-diagonal elements of the covariance matrix) by some scaling factor S, DP-0.5 will increase by a factor of $\sqrt{S}$ for large populations in which the diagonal elements of C can be ignored (also see Haefner et al 2013). This proportionality is most explicitly provided in Haefner et al. 2013 in equations (4), (S1.14) and the caption to Fig S1c.

(A) Change in MT firing rate due to cooling. Each circle represents one neuron.
(B) Implied distributions of susceptibility, S, in the control (red) and cool (blue) conditions.
(C) Distribution of relative changes in S in % with the mean of -7% indicated by the black vertical line, which translates into a reduction in DPs of roughly 4%.

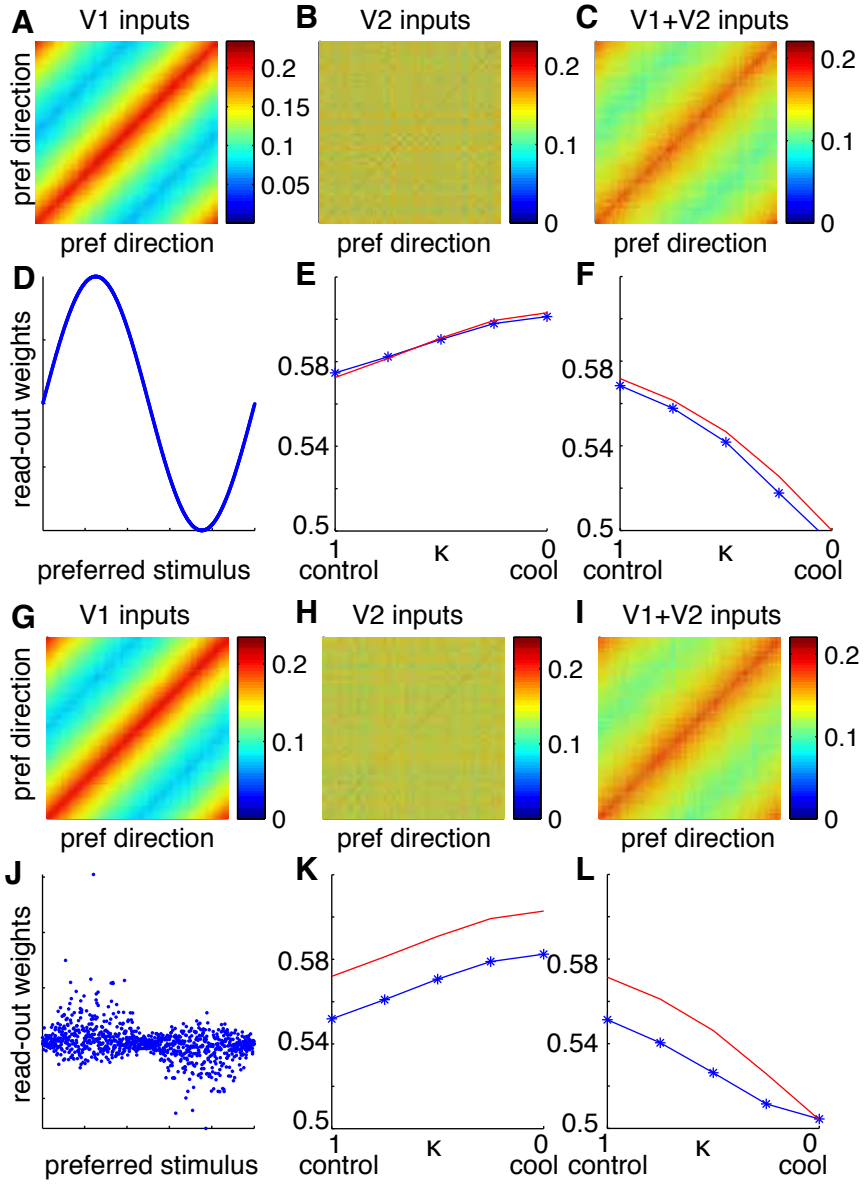

Figure S5: Simulations of heterogeneous populations with various read-out schemes (related to Figure 7)

Same conventions as in Figure 7 in main text.
(A-F) Simulation under the assumption of a heterogeneous instead of a homogenous population (as in Figure 7), for which response variances were drawn from the χ^2-distribution around a mean of 10 spikes^2/sec^2
(G-L) Simulation with the heterogeneous population as in A-F but weights are assumed to be linear optimal. This leads to an extremely large variability in the weights such that even some neurons closely tuned to the target stimulus have negative weights. The general underestimation of the actual DPs by the analytical formula for optimal weights has previously been discussed in Haefner et al. 2013 and only affects the overall scale, not the change with cooling relevant here.

	Monkey S	Monkey Q
Mean Behavioral Threshold (%)		
Scotoma		
Depth pre	36	50
Depth cool	49	88
Motion pre	35	45
Motion cool	41	66
Ipsilateral Control		
Depth pre	37	48
Depth cool	36	38
Motion pre	37	42
Motion cool	35	43
Change in Behavioral threshold: median EI		
Scotoma		
Depth	0.47	0.78
Motion	0.18	0.66
Depth – Motion (paired)	0.29; sign test p = 0.0005	-0.14*; sign test p = 0.38
Ipsilateral control		
Depth	0.00; Wilcoxon rank sum test† p = 1 (n=7)	-0.01; Wilcoxon rank sum test p = 1 (n=16)
Motion	-0.05; Wilcoxon rank sum test p = 1 (n=7)	-0.01; Wilcoxon rank sum test p = 1 (n=7)
Breaks in fixation: median difference (cool – pre-cool)		
Depth	-4%; sign test p = 6.6×10^{-6}	-0.1%; sign test p = 0.76
Motion	1%; sign test, p = 0.0002	0%; sign test p = 1
False alarms: median difference (cool – pre-cool)		
Depth	-4%; sign test p = 0.009	0%; sign test p = 1
Motion	-5%; sign test p = 0.003	-2%; sign test p = 0.27

Table S1: Behavioral Performance (related to Figure 2)

Raw values and statistics for the mean behavioral thresholds, changes in behavioral thresholds, breaks in fixation, and false alarms for each task and animal.

* Note this value is negative, indicating that motion was impaired more than depth in the paired comparison, even though the median effect size was bigger for depth than motion when they were considered separately (the two rows above)

† Here we included data from some sessions in which behavioral performance could be assessed in only one of the two tasks. We therefore use an unpaired statistical test, unlike the paired tests for the effects in the scotoma.

	Combined	Monkey S	Monkey Q
Neurometric performance: median effect index (EI)			
Depth	-0.31	-0.32	-0.30
Motion	-0.18	-0.16	-0.20
$EI_{depth} - EI_{motion}$	-0.13; Wilcoxon rank sum test p = 0.03	-0.16; Wilcoxon rank sum test p = 0.03	-0.1; Wilcoxon rank sum test p = 0.24
Tuning: median difference in discrimination index (DI) (cool – pre-cool)			
Binocular disparity (BD)	-0.14	-0.07	-0.19
Direction	-0.03	-0.01	-0.06
BD – direction	-0.08; sign test p = 3.7×10^{-8}	-0.07; sign test p = 0.005	-0.15; sign test p = 4.1×10^{-5}

Table S2: Neuronal effects of inactivation (related to Figure 3)

Raw values and statistics for the effect index (EI) for neurometric performance, tuning, and changes in the EI with cooling for each animal and task.

	DP		
	Combined	**Monkey S**	**Monkey Q**
Depth pre	0.60	0.61	0.58
Depth cool	0.56	0.58	0.55
Depth pre – cool p-value	0.001	0.02	0.02
Motion pre	0.58	0.58	0.58
Motion cool	0.59	0.59	0.59
Motion pre – cool p-value	0.80	0.73	0.81
P-value for the difference between depth and motion being different from zero	0.003	0.04	0.01

Table S3: Choice-related effects of inactivation (related to Figure 4)

Raw DP values and statistics for each task and animal.

Supplemental Experimental Procedures

Before training, a scleral search coil in each eye and a custom titanium head-post were implanted under general anesthesia. Following training on a motion and then a depth signal detection task a Cilux recording cylinder (Crist Instrument Co.) was implanted in the right hemisphere to access MT via an anterior approach. This was followed by recovery and then by implantation of cryoloops, the devices used for cooling (Ponce et al., 2008). Cooling sessions began after complete recovery, within about one month after implantation.

Visual Stimuli

Visual stimuli were presented through a Wheatstone stereoscope (Wheatstone, 1838) on two Viewsonic 21" G220F CRT monitors such that the virtual image of the fixation point was 57 cm in front of the monkey's eyes. The display subtended 39 x 29 degrees of visual angle at a resolution of 1600 x 1200 and was updated at 75 Hz. All visual stimuli were drawn using the Cogent Graphics Matlab toolbox, developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience (http://www.vislab.ucl.ac.uk/cogent.php).

Stimuli were random dot kinetograms drawn with the red phosphor on both monitors with a luminance of 6.97 cd/m^2 as seen in each of the mirrors. Binocular disparity stimuli were generated by drawing the image of a single dot with a horizontal offset between the two monitors. The magnitude and sign of this offset specifies the depth of the stimulus. By convention, negative disparities are nearer to the observer than the fixation plane and positive disparities are farther.

The stimuli in both detection tasks comprised 0.1° dots drawn at a density of 1-2 dots/deg^2 with a 3-frame (40 ms) dot lifetime. The motion task stimulus has been described (Cook and Maunsell, 2002a). Briefly, before signal onset all dots were "noise" dots that moved in random directions at a fixed speed. After signal onset, a proportion of the dots became signal dots, which moved in one direction at a fixed speed; the rest remained noise dots. The proportion of signal dots dictated the motion signal strength. Dots maintained their direction for their entire 3-frame lifetime such that no dot ever changed direction. The change from noise to signal happened for each dot only after it "died" and was re-plotted in a new location. Because of the limited lifetime, on every frame a third of the dots was re-plotted in a new location; therefore the true motion

signal between any two frames was at most 66%. We only report the true motion signal in the results presented here. Throughout each motion trial all dots were presented at a fixed binocular disparity (*i.e.* 100% coherent in depth).

The depth task stimulus was designed analogously to the motion stimulus. Throughout the trial all dots moved coherently in one direction. Signal dots were presented at a particular binocular disparity and noise dots were scattered in depth in the range -1.5 to 1.5°. The proportion of signal dots dictated the signal strength. This stimulus was also drawn with a 3-frame *motion* lifetime but this did not limit the maximum binocular disparity signal strength. During signal onset, dots changed binocular disparity only upon being plotted in a new location. Stimuli for both tasks were confined to fixed apertures of equal size in the two eyes to eliminate monocular cues to the depth change.

Recording Methods

Tungsten microelectrodes (impedance, 0.5-3 MΩ at 1 kHz) were advanced through a trans-dural guide tube; the signals were amplified and passed through a band pass filter and window discriminator (BAK electronics) for on line spike detection. The analog voltage signals from the extracellular recordings were digitized at 25 kHz by a Cambridge Electronic Design 1401 data acquisition system. Spike2 software was used for offline spike sorting.

The influence of internal threshold on DP

Changes in DP cannot be explained by changes in the animal's internal threshold during inactivation. In previous models of perceptual decision-making (reviewed in Nienborg et al. 2012) and change detection in particular (Smith et al. 2011), a linear combination of the sensory responses is compared to an internal threshold to elicit a decision. It can be shown that the DP measured in such a framework depends on the internal threshold only through the fraction of trials in which the subject detected the stimulus, p^{detect} (Haefner 2014, Methods). We always adjusted the task difficulty in both the cooling and the control condition to approximate the balanced 50/50 case, in which DPs are most comparable to the choice probabilities in discrimination tasks. All our data was derived from signal strengths where $0.3 \leq p^{detect} \leq 0.7$,

which affected our observed DPs by less than ±5% (measured as DP deviation from 0.5). Therefore, possible changes in internal threshold between cooling and control condition cannot explain the observed decrease in DP with cooling in the depth task.

Controls for involuntary eye movements

Although fixation was enforced throughout the entire trial period, differences in the number of small involuntary eye movements (microsaccades) or in the vergence angle between correct and missed trials can affect estimates of DP (Herrington et al., 2009; Nienborg and Cumming, 2006). Although there was a significant difference in the proportion of trials containing microsaccades between missed trials and correct trials on most days (difference of 10-30% depending on the monkey and task), there was no significant relationship between DP in either condition and the number of microsaccades during correct trials, missed trials, or the ratio between the numbers of microsaccades between correct and missed trials ($|\rho| \leq 0.26$, $p > 0.19$ in all cases). In only 5 sessions we found that there was a significant difference in the mean vergence angle between correct and missed trials in one of the tasks or conditions. We computed DP with the exclusion of all neurons for which the animal exhibited this difference and found that this had no effect on the results: DP was reduced significantly more during the depth task than the motion task (depth DP: pre-cool = 0.59, cool = 0.55; motion DP: pre-cool = 0.57, cool = 0.58; $p = 0.02$ resampling procedure). Together, these results demonstrate that microsaccades and vergence state did not affect our estimates of DP. Similar analyses revealed that changes in eye movements were not related to the animals' behavioral performance.