A Stable Visual World in Primate Primary Visual Cortex

Highlights

- Neurons in primary visual cortex (V1) are sensitive to the direction of gaze

- This "eye-tracker" signal is encoded similarly during fast and slow eye movements

- This signal is embedded within the classical map of retinal visual space

- V1 encodes the true locations of objects, not only their positions on the retina

Authors

Adam P. Morris, Bart Krekelberg

Correspondence

adam.morris@monash.edu

In Brief

Visual input arrives as a series of snapshots, each taken from a different line of sight, due to eye movements from one part of a scene to another. How do we nevertheless see a stable visual world? Morris and Krekelberg show that in primary visual cortex, the neural representation of each snapshot includes "metadata" that tracks gaze direction.

CellPress

Article

CellPress

A Stable Visual World
in Primate Primary Visual Cortex

Adam P. Morris[1,3,*] and Bart Krekelberg[2]
[1]Neuroscience Program, Biomedicine Discovery Institute, Department of Physiology, Monash University, 26 Innovation Walk, Clayton, Victoria 3800, Australia
[2]Center for Molecular and Behavioral Neuroscience, Rutgers University, 197 University Ave., Newark, New Jersey 07102, USA
[3]Lead Contact
*Correspondence: adam.morris@monash.edu
https://doi.org/10.1016/j.cub.2019.03.069

SUMMARY

Humans and other primates rely on eye movements to explore visual scenes and to track moving objects. As a result, the image that is projected onto the retina—and propagated throughout the visual cortical hierarchy—is almost constantly changing and makes little sense without taking into account the momentary direction of gaze. How is this achieved in the visual system? Here, we show that in primary visual cortex (V1), the earliest stage of cortical vision, neural representations carry an embedded "eye tracker" that signals the direction of gaze associated with each image. Using chronically implanted multi-electrode arrays, we recorded the activity of neurons in area V1 of macaque monkeys during tasks requiring fast (exploratory) and slow (pursuit) eye movements. Neurons were stimulated with flickering, full-field luminance noise at all times. As in previous studies, we observed neurons that were sensitive to gaze direction during fixation, despite comparable stimulation of their receptive fields. We trained a decoder to translate neural activity into metric estimates of gaze direction. This decoded signal tracked the eye accurately not only during fixation but also during fast and slow eye movements. After a fast eye movement, the eye-position signal arrived in V1 at approximately the same time at which the new visual information arrived from the retina. Using simulations, we show that this V1 eye-position signal could be used to take into account the sensory consequences of eye movements and map the fleeting positions of objects on the retina onto their stable position in the world.

INTRODUCTION

During everyday vision, fast eye movements known as "saccades" shift the line of sight to place the high-resolution fovea onto objects of interest. These movements are complemented by slower, smooth eye rotations used to track moving objects

(e.g., a bird in flight) or to lock gaze on an object during locomotion [1]. Though essential, eye movements have stark consequences for the retina's view of the world: objects that are stationary appear to jump or move, and objects that are in motion travel along grossly distorted trajectories (Figure 1). These phenomena reflect the fact that visual input arrives to the nervous system in an *eye-centered* (or retinal) coordinate system, in which objects evoke activity within a local patch of photoreceptors and ganglion cells (i.e., as a place code). Our experience of the world, however, is entirely different: objects do not appear to change positions every time we move our eyes, and we have no trouble planning actions toward them. This suggests that the visual system uses internal knowledge of eye position (and head position in the case of combined eye-head gaze shifts) to compute the *true* locations of objects from their fleeting retinal projections (i.e., their positions in head- and/or body-centered coordinates). Where, and how, this transformation take place in the brain remains unclear.

At first glance, it would seem that a stable representation of visual space is not achieved until late in the cortical hierarchy. For example, in the first stage of cortical vision—the primary visual cortex (V1)—neurons have receptive fields (RFs) that are fixed on the retina [2], and so they move across a visual scene with the eye. Patterns of activity within the V1 retinotopic map thus reflect what is happening on the retina, not in the world itself. Perhaps surprisingly, this remains true throughout the cortex beyond V1, where most visually sensitive neurons have RFs that are anchored to the retina [3–6]. Similarly, neurons that are sensitive to image motion tend to prefer the same direction on the retina regardless of whether it was caused by a moving object or a moving eye [7, 8].

In parietal and temporal areas, however, some of these representations retain information across eye movements by updating patterns of activity, thereby maintaining spatial alignment with the outside world—so-called "remapping" [9, 10]. In addition, the spatial selectivity of some neurons is anchored in head or world coordinates [8, 11–14]. That is, an object at a fixed location in space evokes activity in the same set of neurons irrespective of eye position. Both phenomena—remapping and head-fixed RFs—are rare or absent in earlier visual areas [2, 15], consistent with the idea that stability emerges late in the hierarchy for vision, and even then, only within the dorsal "where" pathway that supports the planning of actions (see [16] for a review). In this view, the ventral "what" pathway that supports object recognition must rely on its dorsal stream counterpart to provide the spatial

Object trajectory in the world

$$Y_w$$

$$o_w(t)$$

$$X_w$$

$$o_w(t) = o_r(t) + e_w(t)$$

Eye movement

Pursuit

$$e_w(t)$$

Saccade

$$X_w$$

$$Y_w$$

Object trajectory on the retina

$$o_r(t)$$

$$X_r$$
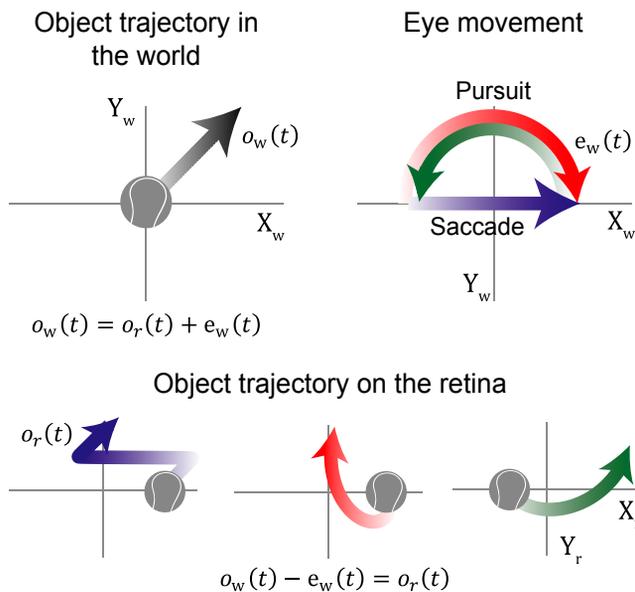
$$Y_r$$

$$o_w(t) - e_w(t) = o_r(t)$$

Figure 1. Our Sense of a Stable Visual World Arises from Retinal Images That Change with Every Eye Movement

In the cartoon, a fixed state of the world (i.e., a ball in motion at 5°/s; black arrow) generates three patterns of retinal stimulation (lower panels) under different eye-movement scenarios: a fixate-saccade-fixate sequence (purple arrow) and smooth pursuit at 10°/s in opposite directions around a half-circle (green and red arrows). To catch the ball, however, the brain would need to know its true location and program the same set of motor commands, regardless of its appearance on the retina. Note that different coordinate systems are used across panels: world coordinates (X_w, Y_w) for the object and eye and retinal coordinates (X_r, Y_r) for the visual images (hereafter, these r and w subscripts are used to indicate the coordinate system of a spatial variable). The object's trajectory on the retina, $o_r(t)$, is roughly equal to the object's trajectory in the world, $o_w(t)$, minus the path traveled by the eyes, $e_w(t)$. The brain, however, must make the reverse inference to recover the object's true trajectory, i.e., $o_w(t) = o_r(t) + e_w(t)$. In this report, we show that V1 neurons represent in parallel both terms on the right side of this equation and hence implicitly also represent $o_w(t)$.

information necessary to piece together snapshots taken from different lines of sight.

We tested the alternative hypothesis that visual stability begins in V1—despite superficial appearances to the contrary—through a distributed population code [17–19]. Such a solution would have the advantage that all downstream computations—in both the dorsal and ventral streams—have direct access to visual information that is robust to eye movements. Our hypothesis builds on previous reports of neurons throughout visual cortex, including in V1 [20–23], that alter their rates of action potentials (spikes) systematically with changes in eye position, even when the image within their RFs is held constant—a property known as a "gain field" [24, 25]. This means that spikes in V1 carry information about not only the retinal image but also the current eye position. This is critical, because to a first approximation, the location of an object in a scene is the sum of its position on the retina and the current eye position (we ignore head movements and other postural changes; Figure 1). Thus, a population of gain-field neurons might allow an eye-centered area like V1 to nevertheless carry a stable representation of visual space. This prediction is supported by influential computational studies

[17, 18, 26] that incorporated artificial, idealized gain-field neurons, but it has to our knowledge never been tested empirically.

For this hypothesis to be true, real gain fields in V1 must support an accurate representation of the position of the eye during natural vision. We and others have shown that neurons in *higher* areas of the dorsal pathway reliably code the direction of gaze during fixation [27–29] and predict the landing point of an impending saccade [28, 30, 31]. Here, we show that not only is a reliable eye-position signal present much earlier—in V1—but also that it can be incorporated into downstream computations in the same way irrespective of whether the eyes are exploring a scene (i.e., saccades) or tracking an object (i.e., pursuit). Thus, this crucial ingredient for visual stability and acting on the world is present in the first cortical area and available to both the dorsal and ventral processing streams.

RESULTS

We recorded extracellular action potentials from 397 units (232 single units and 165 multi-units) using multi-electrode arrays permanently implanted in the parafoveal region of V1 in two adult macaques. Unless otherwise specified, single and multi-units were analyzed together (there were no qualitative differences between their responses; see STAR Methods). On "saccade" trials, the animal fixated a target at one of eight possible starting positions arranged in a circle (15° diameter; only one target was displayed at a time; Figure 2A). After a delay of 1,100 ms, the target stepped to the opposite side of the circle, requiring an immediate saccade and fixation at the new position for a further 1,100 ms. "Pursuit" trials (interleaved with saccade trials in 45/68 sessions) were identical except that after 500 ms of initial fixation, the target moved clockwise (CW) or counter-clockwise (CCW) around the imaginary circle to the opposite side (speed = 10°/s), requiring the animal to track its position with pursuit eye movements (Figure 2A). A final trial type (all sessions) required fixation on a central target for the whole trial.

In most trials, we presented flickering noise on the monitor by setting every pixel to black or white randomly every 13.3 ms. The high spatiotemporal frequency of the noise ensured comparable stimulation of RFs regardless of eye position or velocity (except during the saccade itself, where high velocity retinal slip could, at least partially, gray out the stimulus). Further, it allowed the use of spike-triggered reverse correlation to estimate the RFs of the recorded units (see STAR Methods). Using this approach, we confirmed that the observed RFs were located within the lower parafoveal region in both animals (1.5°–3.0° from the fovea; Figures 2B and 2C). In a subset of experiments (N = 57 units), we also recorded neural activity during trials in which the flickering stimulus was replaced by a uniform gray screen (interleaved with flicker trials).

Decoding Eye Position during Fixation

We extracted spike counts in 100-ms bins during the initial fixation epoch of the saccade trials (−700 ms to −300 ms relative to saccade onset) and for a matched epoch on the central fixation trials. Consistent with previous reports [20–22], roughly half of the units (219/397) showed significant differences in mean firing rate across the nine fixation positions, even though the RF was stimulated by statistically identical visual noise—that is, they
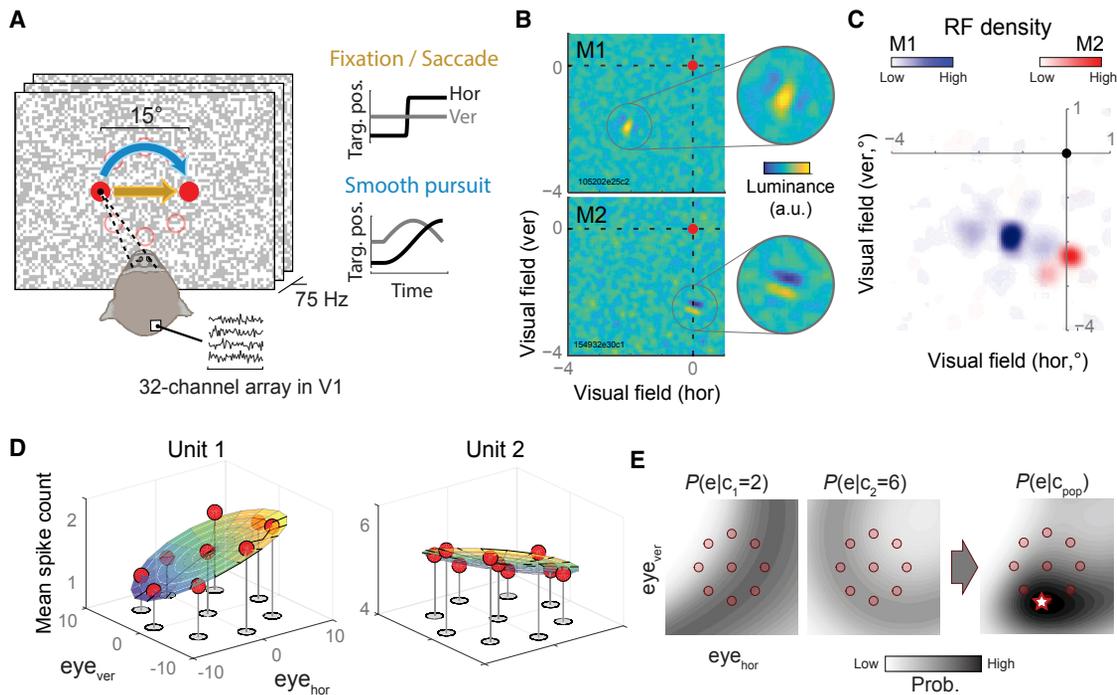
Figure 2. Experimental Design

(A) Animals performed a saccade (yellow arrow) or a pursuit task (blue arrow) over a background of flickering noise pixels. The two tasks were interleaved randomly from trial to trial. Filled red circles show the initial and final positions of the target for the 9-o'clock starting position, and the plots show its 2D position over time in the two tasks. The starting position varied from trial to trial (open circles), with all trials ending at the diametrically opposite position on a virtual circle. Noise pixels are enlarged for illustration.

(B) The noise patterns were used to estimate the receptive fields of neurons (see STAR Methods). An example RF is shown for one unit from each animal (M1 and M2), plotted relative to the fovea (red dot). Each image shows the luminance pattern that, on average, caused the unit to generate a spike, as well as an enlarged view of the RF. The units preferred oriented edges (transitions from bright to dark) presented below the fovea (M2) or below and to the left of the fovea (M1).

(C) The distribution of receptive fields across all recordings for the two animals. Because the recording array spanned a patch of cortex, units recorded from different electrodes had RFs in slightly different parts of the visual field. Color saturation shows the proportion of recorded units that had RFs that included each location in the display. White areas were outside of the RFs of all recorded units.

(D) Example gain fields for two units, plotted as mean spike counts for 100 ms samples during the fixation epoch (gray shaded region in A) at each of nine eye positions (the standard errors of the means are omitted for clarity but were smaller than the symbols in all cases). Under the assumption of Poisson variability, the fitted regression surfaces provide a generative model relating eye position to spike-count probabilities—a "probabilistic gain field" (pGF).

(E) Population decoding using pGFs. The two leftmost images show the probability of all possible eye positions given observed spike counts of 2 and 6 for units 1 and 2 in (D), respectively. The rightmost image shows the combined probability map for this minimalistic population of two units. The position of the peak (star) represents the maximum likelihood estimate of eye position for this example, $[\hat{e}_{wx}, \hat{e}_{wy}]$. In practice, the decoded population included all units with gain fields. See also Figure S1.

showed gain fields (Figure 2D). The mean strength of this modulation was 43% (SD = 42%; median = 28%), defined as the difference between the maximum and minimum mean rate divided by the global mean. In control experiments, we confirmed that these modulations were not driven by differences elsewhere in the retinal image that co-varied with eye-position, such as the edges of the display (which were at least 6° outside of the RFs at all times; Figure S1).

The presence of gain fields hints that as a population, these units might carry enough information to encode 2D eye position (i.e., horizontal and vertical orientation of the eye in the orbit, or equivalently, fixation position on the screen). Building on our previous work [27, 30, 31], we used these neurons to construct a decoder based on maximum-likelihood estimation. The decoder translated a vector of spike counts (one for each unit in our sample) into an estimate of eye position within a finely-spaced 2D grid of possible positions. To do so, it used the gain fields as a

kind of probabilistic "look-up table" (Figures 2D and 2E; STAR Methods).

Figure 3 shows the decoder performance for population responses measured during the fixation epoch (using cross-validation on a dataset not used to setup the decoder). V1 represented the eye accurately and precisely across all of the tested positions, albeit with slight spatial compression (mean *constant error* across conditions was 1.29° [standard error of the mean (STE) across the nine eye positions = 0.16]; *variable error*, defined as the median Euclidean distance of the trial-by-trial estimates from the distribution median, and averaged over conditions, was 3.35° [STE = 0.32°]). Comparable performance was observed when decoding the initial fixation position on pursuit trials (Figure 3), albeit with greater error (mean constant error = 1.97° [STE = 0.32°]), presumably due to a smaller population size (pursuit trials were included during recordings for only 179/219 units). As a further test of robustness, we decoded the
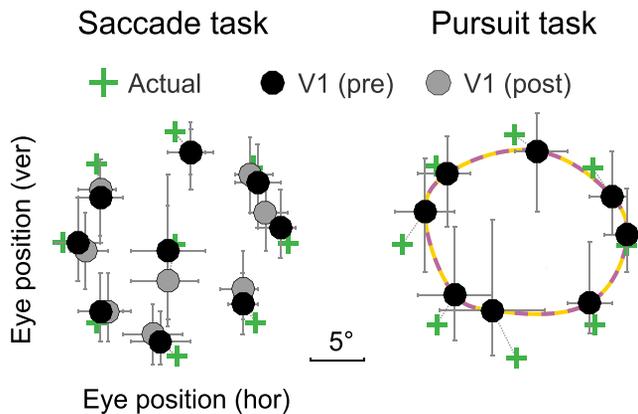
Saccade task   Pursuit task

Figure 3. The Representation of Eye Position in V1 Is Accurate during Fixation

Eye position was decoded from 100 ms samples of neural activity. Left panel: median decoded eye positions during fixation before (pre) and after (post) the saccade, compared with the actual fixation positions. Error bars represent the variability (middle 50% of the distribution) across samples (trials and time). Right panel: decoded positions during the initial fixation phase on pursuit trials. The yellow-purple dashed line is a spline interpolation through the points; it is used in Figure 5 to assess the quality of the decoder during the pursuit.

fixation epoch *after* the saccade (+200 to +500 ms) and again found good accuracy (mean constant error = 1.82° [STE = 0.23°]), even though the decoder had never been exposed to data from these time points.

Decoding Eye Position during Fast Movements

Next, we asked whether this decoder, which was based only on the gain fields measured during fixation, could track the eyes across saccadic and pursuit eye movements without changing any of its parameters. For this to be the case, firing rates must continue to be influenced by the eyes during these other oculomotor behaviors and remain statistically consistent with those observed during fixation. For example, if a neuron showed higher mean firing rates during fixation on the right side compared with the left, it must similarly do so in the immediate wake of a saccade to this position and when the eyes pass smoothly through it during object tracking.

To reveal the representation of eye position in V1 during saccades, we applied the fixation decoder sequentially and independently to each point in time throughout the fixation-saccade-fixation sequence (Figures 4A and 4B). The transition from initial to final eye position—the neural "saccade"—occurred shortly after the true change in eye position and with approximately sigmoid-shaped dynamics, much like the actual eye. This was true for all eight saccade directions and starting positions (Figure 4A).

We summarized the decoder's performance for saccades by rotating the data in Figure 4A so that the horizontal coordinate corresponded to the saccade direction and then averaging across conditions. In this format, the time courses are represented as if all saccades were to the right. This averaged eye-position signal closely matched the true eye position (Figure 4B). To quantify the latency of the V1 signal, we fitted the decoded eye position and the true eye position data in Figure 4B with sigmoids (see STAR Methods). Using parameters from the fits, we
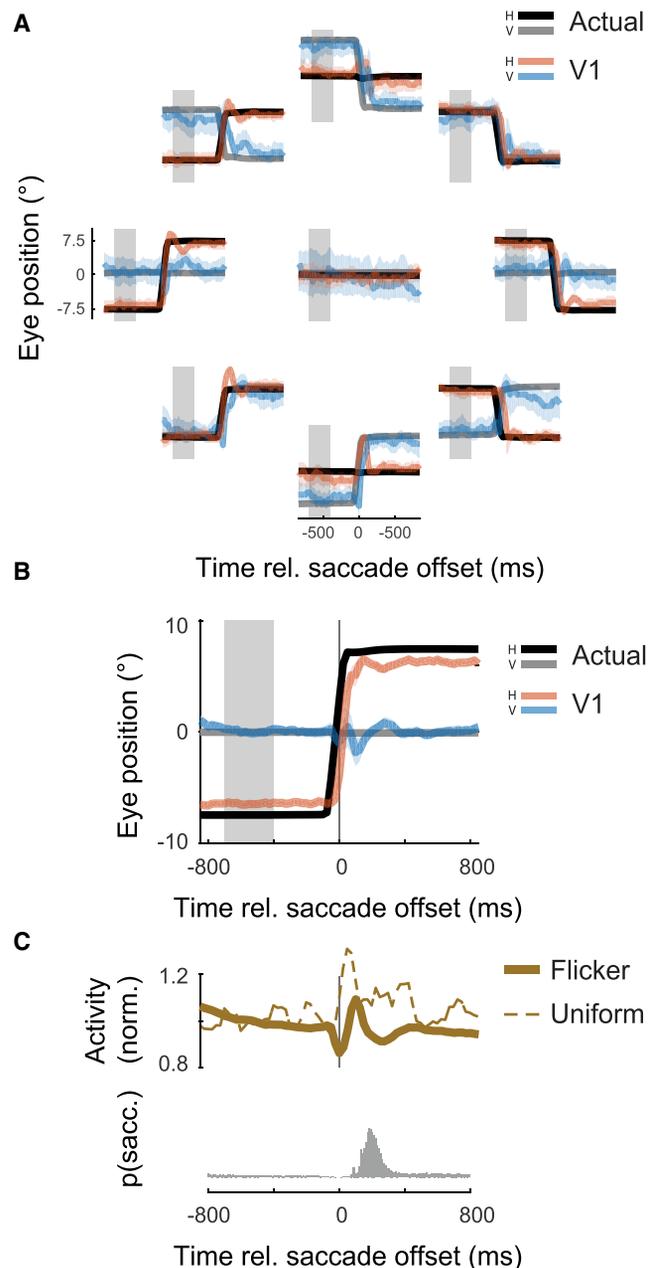

Figure 4. The Representation of Eye Position Updates Rapidly with Each Saccade

(A) The decoded eye position over time for each saccade direction (red and blue curves), plotted against the average eye trace (black and gray). Plots are arranged spatially according to the starting position (no saccade was required for the central fixation position). Curves represent the median decoded position and shading represents the variability (middle 50% of distribution) across trials. The gray shaded region indicates the fixation epoch used to build the decoder (i.e., to estimate a pGF for each neuron).

(B) The data from (A), averaged over all eight starting positions (after rotating to align their starting positions). Error shading represents ±1 STE.

(C) The upper panel shows the mean normalized firing rate (see STAR Methods) across the population in response to the flickering background and for a uniform gray background. The lower panel shows the distribution of saccade onset times (relative to the offset of the primary saccades) across all sessions, excluding the primary saccade of the task. Most occur in the wake of the primary saccade. See also Figure S2.
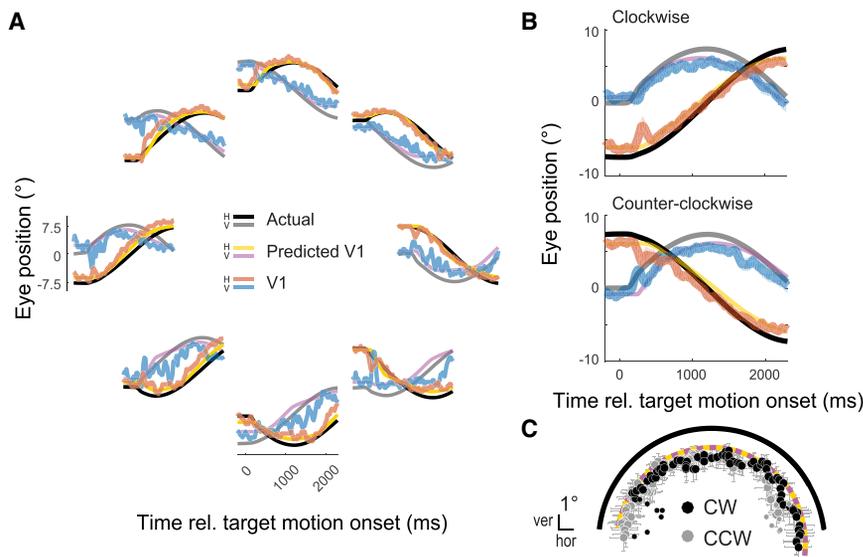
A

B

Clockwise

Counter-clockwise

Time rel. target motion onset (ms)

C

Eye position (°)

Actual
Predicted V1
V1

Time rel. target motion onset (ms)

1°
ver
hor

● CW
● CCW

Figure 5. V1 Neurons Track the Eye during Pursuit Using the Same Neural Code as during Fixation

(A) Decoded eye position over time for clockwise pursuit, plotted as in Figure 4A, but with the addition of curves (yellow and purple) showing what the decoder performance would look like if the spatial errors in decoder performance seen during fixation (Figure 3) were recapitulated during pursuit. This predicted V1 signal corresponds to the yellow-purple spline shown in Figure 3 but plotted over time (see STAR Methods).

(B) As in (A) but averaged over the eight conditions for each pursuit direction. Error shading represents ±1 STE. The curves for the predicted signal are mostly obscured by those for the decoded eye, reflecting the close match.

(C) The data from (B) plotted in space (black and gray filled circles). Time points before the onset of steady pursuit are plotted as smaller circles. (CW, clockwise; CCW, counter-clockwise.) The black curve indicates the true eye position, averaged over clockwise and counter-clockwise directions. The yellow-purple curve indicates the performance predicted on the basis of fixation decoding. Error bars represent ±1 STE.

found that the V1 saccade had a similar duration (13 ms longer than the real eye) and lagged the actual eye by just 45 ms. This short lag is well within the duration of a typical fixation [32], suggesting that the V1 representation of the eye is able to keep up with the eye during exploratory vision. At this lag, the decoded eye-position signal accounted for 97% of the variance in the true eye position over the course of the trial. Interestingly, this lag closely matched the visual latency in these neurons (53.11 ms; SD = 9.93 ms; see STAR Methods). This implies that, immediately after a saccade, the new retinal information and the new eye-position information arrive in V1 at around the same time.

There were, however, brief errors in the representation of the saccade step. In Figure 4A, these errors are most visible for saccades along the cardinal axes, but they were present in all conditions and both parallel and orthogonal to the direction of eye movement (Figure S2). This not surprising, because neural responses underwent considerable changes around the time of the saccade, including a transient reduction in rate, followed by an enhancement (Figure 4C). Further, these effects on firing rate echoed (weakly) after the primary saccade due to the presence of corrective saccades on some trials (Figure 4C). Such shifts in spiking regime are challenging for fixed-parameter decoders such as ours and can lead to volatility and bias. There was no pattern to the observed errors across conditions, and they largely cancelled out in the averaged time course. This suggests that the eye would likely be represented accurately for all saccade directions with a larger sample of neurons.

In addition to the transient effects around the time of the saccade, there was a gradual reduction in firing rate during the fixation intervals, totaling 6.0% over the course of the trial (SD = 6.7%). These changes were tolerated well: the representation of eye position remained steady and accurate, reflecting a degree of robustness to global changes in the firing rate in the decoder.

Decoding Eye Position during Pursuit Movements

The previous section showed that V1 faithfully encodes eye position across saccadic eye movements (albeit with a short lag). We tested whether the same neural code also represents eye position during pursuit—an oculomotor behavior that differs greatly from saccades both in terms of kinematics and neural control mechanisms [33]. In the pursuit task, the eye tracked the target along a circular path at a constant speed, passing through the positions measured during the fixation task on other trials. The horizontal (e_{wx}) and vertical (e_{wy}) coordinates of the eye therefore traced out a (half-period) sinusoid over time with an amplitude of approximately 7.5°.

Figure 5 plots the decoded eye position during pursuit for each of the eight starting positions (CW only; Figure 5A) and averaged over conditions (separately for CW and CCW; Figure 5B). The decoder tracked the position of the eye tightly but with a centripetal bias toward the central position. We quantified this performance with simultaneous sine and cosine fits to the horizontal and vertical components of the decoded eye position (not shown in the figure; Equations 3 and 4). The amplitude of the fit showed that the decoded signal had a pursuit radius of 5.62 (i.e., 75% of the true value).

Some of this bias, however, was expected on the basis of the decoder performance for the fixation epochs of pursuit trials (Figure 3, right panel). We thus generated a *predicted* pursuit curve—under the hypothesis that the neural code for eye position is identical for stationary and moving eyes—by interpolating the decoded fixation data points with a 2D cubic spline. This provided a good match to the decoded data (R^2 = 0.91 and 0.97 for CW and CCW, respectively; Figure 5), showing that pursuit introduces few, if any, additional biases into the decoder.

The sine and cosine fits suggested that the V1 representation *led* the real eye by 35 ms and 61 ms for CW and CCW pursuit, respectively. The decoded trajectories, however, were not
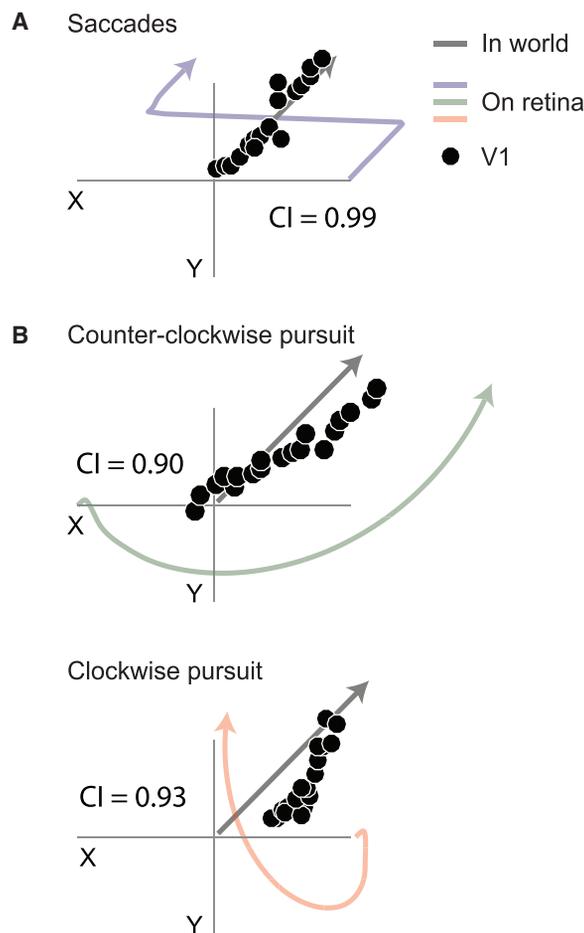
CellPress

Figure 6. Population Codes for Eye Position in V1 Allow the Area as a Whole to Represent the True Locations of Objects in the World

We simulated an observer that localizes objects by combining their instantaneous positions in the V1 retinotopic map with our experimentally measured eye-position signal. The simulated scenarios were identical to our real experiment—including the saccades and pursuit eye movements—but with the addition of an object in linear motion (black arrow) as in Figure 1. The observer's "perception" (black circles), which indexes the representation of visual space in V1, was accurate even though its projection on the retina differed greatly under different eye movements (colored arrows). CI shows the compensation index (see STAR Methods) for each of the scenarios.

(A) A saccade, identical to that in our real experiment, was performed mid-way through the object's trajectory.

(B) Object motion was viewed during clockwise or counter-clockwise pursuit.

perfectly sinusoidal (Figure 5A), complicating a simple interpretation of lag parameters. Calculating lag with respect to the predicted pursuit trajectory (see STAR Methods) resulted in different values: 58 ms for CW conditions (consistent with the estimate from the saccade task) and −163 ms for CCW (i.e., anticipatory). The lag during pursuit thus remains unclear.

How Well Does V1 Code the True Positions of Visual Objects?

What does the reliable representation of eye position mean for the coding of visual space in V1? Geometrically, the true location of an object (o_w) is the sum of its position on the retina (o_r) and

eye position (e_w; we ignore head movements; Figure 1). This means that, by definition, a region of cortex that simultaneously encodes o_r and e_w implicitly also represents o_w (through the availability of $o_r + e_w$, as long as downstream neurons can extract this information). The presence of a robust place code for o_r in V1 is well established: neurons have RFs that remain at a fixed position on the retina irrespective of eye position [2]. Hence, a point stimulus evokes a narrow hill of activity in the V1 retinotopic map that moves across the cortical sheet according to its own movements and those of the eyes. Our results show that neurons within a local region of the map also encode e_w. This implies that V1 as a whole could hold a stable representation of visual space even when the eyes move. For this to be true, the visual and eye-position signals would also need to be updated at the same time.

To quantify how well V1 would represent the *true* location of a visual object (o_w), we simulated an observer-viewing scenario like those in Figure 1. The eye movements in the simulations were identical to those in our real experiment. The simulated observer located the moving visual object by combining its momentary position on the retina, $o_r(t)$, with the *empirical* eye position we extracted from neural activity in V1 (i.e., Figures 4B and 5B), $\hat{e}_w(t)$, both represented in our analysis as 2D vectors. To take into account the visual latency of our V1 neurons, we lagged $o_r(t)$ by 53 ms (see STAR Methods). $o_r(t)$ was thus used as a proxy for the (delayed) hill of activity this hypothetical object would evoke in the V1 map rather than its position on the retina, per se.

This hypothetical observer perceived the movements of objects to be close to their real-world movements, despite grossly distorted trajectories across the retina and V1 map (Figure 6). To benchmark this performance, we compared these reconstructions with a second hypothetical observer who used only the retinal information to locate the object (i.e., by equating retinal coordinates with world coordinates). We defined a compensation index (CI; see STAR Methods), in which a value of 1 corresponds to a perfect reconstruction of the object's trajectory and 0 corresponds to the error expected using only its retinal trajectory. Under the traditional view of V1—in which neurons support only a retinotopic position code—the CI values should all be close to zero. Our results, however, showed a value of 0.99 for saccades, and 0.90 and 0.93 for clockwise and counter-clockwise pursuit, respectively, showing that this hypothetical observer compensated almost completely for the effects of eye movements.

DISCUSSION

Our results show that neurons in primary visual cortex carry a distributed signal that serves as an intrinsic eye tracker, representing the moment-to-moment direction of gaze. This population code was nimble enough to keep up with the eye during exploratory, saccadic eye movements, and it also tracked the eye during smooth pursuit movements. Together with V1's well-established representation of retinal position (i.e., the retinotopic map), this signal provides a plausible neural substrate for the stability of visual perception across eye movements, and it suggests that the sensorimotor transformations needed for action begin at the earliest stage of cortical vision.

Where in the Brain Is the Stable World Represented?

Figure 6 showed that V1 contains the information necessary to represent visual space in a way that is much more robust to the effects of eye movements than is suggested by its retinotopic organization and eye-centered receptive fields [2]. That is, V1 neurons represent not only the locations of features on the retina (which change every time the eyes move) but also their true locations in the world (technically, their positions relative to the head). By "represent," we mean that an algorithm can extract the information from activity in V1 alone, without accessing any other area. We showed this by construction; the algorithm used in Figure 6 added the 2D vectors of eye position decoded from our data with a (hypothetical, but well-established) 2D vector of a feature's retinal position.

In the brain, such a computation can likely be performed through appropriate pooling of activity in a single synaptic step, as shown by extensive modeling work [17, 18, 26, 34]. Further, our results suggest this could occur with fixed pooling weights irrespective of the oculomotor state of the animal. Pooling V1 activity would allow, for example, the construction of an RF that is anchored in head coordinates rather than retinal coordinates, like those seen in higher visual areas [8, 11–14]. More likely, however, downstream areas continue to represent stable visual information in distributed, retina-centered representations like that shown here. The finding that some visual illusions depend on eye position is consistent with this view and the results reported here [18, 35, 36].

Note that this interpretation does not preclude other uses of the eye position signals in V1, including prioritization of visual signals arising from directly in front of an observer's body [22, 37], stabilization of image representations during fixational eye movements [38, 40, 41] (A.G. Anderson et al., 2016, 50th Asilomar Conference on Signals, Systems, and Computers, conference), or calibration of oculomotor signals [42].

Robustness in Rich Visual Scenes

In our simulations (Figure 6), and in the classical modeling work [17, 18, 26, 34, 43], visual scenes consisted of a single, isolated object. In contrast, real scenes are cluttered and have high dimensionality (e.g., color, orientation, luminance, etc.). This has consequences for the representation of an object's retinal position, as there is not a single hill of activity in the cortical map but rather a mountain range reflecting the structure of the scene. Similarly, scene structure could affect the representation of eye position, because neural activity will be affected by the image features that happen to be in the RFs. This raises the question of whether the V1 eye-position signal is robust in the face of rich visual scenes.

Two aspects of our results suggest that this should be the case. First, our experiments used white-noise patterns that spanned a large set of black-and-white scenes through their random pixel structure. Despite these changes, the decoder estimated eye position with good accuracy and precision. Second, the decoder was relatively insensitive to global changes in the population activity while a given eye position was maintained (Figures 4B and 4C). This reflects a key computational advantage of distributed population codes [17, 18]: information is mostly encoded in the *patterns* of activity across neurons and hence is robust to factors that affect all neurons similarly.

Finally, we note that the issue of robustness (or invariance) to scene structure is not unique to eye-position coding [44, 45]. The classical tuning properties of V1 neurons include many dimensions of the retinal image, including position, orientation, motion, spatial frequency, color, disparity, and so on. The spike rate of a single neuron—as a one-dimensional quantity—clearly cannot code all these visual dimensions simultaneously. Consider a classical V1 neuron that responds strongly to a grating when it is oriented horizontally but weakly when it is oriented vertically. What can we conclude about orientation if the neuron is responding weakly? The grating *could be* vertical, but it could also be that the grating has been moved to the margin of the receptive field or switched to one of a non-preferred spatial frequency. Hence, even classical properties of V1 are encoded ambiguously at the level of single neurons and must be extracted through selective pooling of population activity. We suggest that eye-position information in V1 should be viewed similarly, as one of many features encoded in V1 that can be incorporated into computations downstream.

The Origin of Eye-Position Signals

There are many known sources of eye-position information in the brain that could underlie the V1 signal we have observed. Proprioceptive signals from the eye muscles target visually sensitive neurons in sub-cortical and cortical structures [46], including the main feedforward driver of V1 (i.e., lateral geniculate nucleus [19, 47]). In somatosensory cortex (area 3a), proprioceptive signals lag the eye by an amount similar to that observed here for V1 [48, 49] (compare our Figure 4B with Figure 3F in [48]). Such a source could also explain why a single decoder generalized across different oculomotor behaviors (fixation, saccades, and pursuit). A further hint that proprioceptive signals play a role in V1 is that blocking these afferent signals impairs the development of orientation selectivity and stereo vision [50]. Ultimately, experiments in which the proprioceptive afferents are selectively abolished are needed to fully quantify their contribution to the V1 eye-position signal [48, 49].

Alternatively, or in addition, V1 could be informed of an impending change in eye position by efference copy signals originating in movement-planning areas, such as the superior colliculus or the frontal eye fields [51–53]. These are the only sources that could explain predictive eye-position signals, such as those found in parietal cortex [30, 31, 54]. Although there is evidence of presaccadic changes in V1 activity [55], there was little sign of such effects in our data. Hence, it is unclear whether efference copy signals contribute to the eye-position signals reported here.

Interestingly, eye position can also be estimated from changes in visual responses alone, provided that the scene structure remains constant. V1 might use such an algorithm to stabilize a representation of the retinal image during fixation, where tiny eye movements would otherwise reduce acuity [38, 41, 52]. In our experiments, however, the image was changed randomly every 13 ms, and eye movements were large, suggesting that this type of algorithm is unlikely to have contributed to our results. Conversely, however, the V1 signal we have observed could be used to improve the performance of such stabilization mechanisms. The precision of the V1 signal (decoded from just 219 units) in our experiments was too low to track such tiny

eye movements (Figure 3), but this might be feasible for V1 as a whole.

Beyond V1

Visual stability during eye movements is a multi-faceted problem. Our data suggest that V1 solves one aspect: it establishes a correspondence between the current retinal input and its location in the world and thereby generates a representation of visual information in a reference frame suitable for actions in the world. This representation was nimble and accurate in the immediate wake of saccades, thus addressing one of the main criticisms levelled against gain-field models of visual stability [42]. Indeed, it was updated considerably faster than the ∼150 ms delay seen in many neurons in the lateral intraparietal area (LIP) of parietal cortex [42] (though LIP as a whole also supports faster, even predictive, eye-position signals [30]). Of course, our study does not show that the representation is perfect under all conditions, and it would be interesting to find its limitations. For instance, what happens in low light or even complete darkness? Are there errors that can explain behavioral phenomena that suggest the use of an incorrect reference frame [56–58]?

Other aspects of visual stability may not be resolved at the level of V1. For instance, saccades briefly interrupt the visual input, and these interruptions could be perceptually jarring. Some neurons mitigate this problem by remapping activity at different times around the saccade, improving the temporal continuity of visual signals [9, 59]. Remapping is exceedingly rare in V1, but it is relatively common in other areas [15]. Hence, the temporal continuity and reference frame problems are not only conceptually distinct (solving one does not necessarily solve the other) but also seemingly solved by different neural substrates.

Our finding that V1 is home to a stable, distributed representation of visual space does not preclude the existence of similar representations in other visual areas. In fact, we speculate that the signals we observed in V1 propagate through the cortical hierarchy along with the feedforward visual information. This would maintain the spatiotemporal alignment needed for stability within each area (and hence at every spatial scale and at every level of motion or form analysis). The presence of eye-position signals across the visual hierarchy supports this view [12, 21, 25, 27, 29–31, 60, 61].

Conclusion

V1 is usually thought to encode only what is on the retina, with no information about where objects are in the world. Instead, the true locations of objects are represented in V1 implicitly through a reliable population code for eye position. This stable spatial information is available to both the dorsal (where) pathway and the ventral (what) pathway, ideally placed to support action planning, object recognition, and decision making in the face of eye movements.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Subjects
  - Animal preparation
- METHOD DETAILS
  - Stimuli and apparatus
  - Behavioral tasks
  - Electrophysiology and data processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Estimation of receptive fields
  - Normalized population activity
  - Eye position gain-fields
  - Population decoding of eye position
  - Comparison of single unit versus multi-unit activity
  - Control experiments
  - Assessing visual stability
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.03.069.

AUTHOR CONTRIBUTIONS

A.P.M. and B.K. performed all aspects of this work, including conceptual development and experimental design, surgical procedures, animal training, programming for experimental control and analysis, and preparation of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Kowler, E. (2011). Eye movements: the past 25 years. Vision Res. 51, 1457–1483.

2. Gur, M., and Snodderly, D.M. (1997). Visual receptive fields of neurons in primary visual cortex (V1) move in space with the eye movements of fixation. Vision Res. 37, 257–265.

3. Andersen, R.A., and Buneo, C.A. (2002). Intentional maps in posterior parietal cortex. Annu. Rev. Neurosci. 25, 189–220.

4. Colby, C.L., and Goldberg, M.E. (1999). Space and attention in parietal cortex. Annu. Rev. Neurosci. 22, 319–349.

5. Hartmann, T.S., Bremmer, F., Albright, T.D., and Krekelberg, B. (2011). Receptive field positions in area MT during slow eye movements. J. Neurosci. 31, 10437–10444.

6. Morris, A.P., Liu, C.C., Cropper, S.J., Forte, J.D., Krekelberg, B., and Mattingley, J.B. (2010). Summation of visual motion across eye movements reflects a nonspatial decision mechanism. J. Neurosci. *30*, 9821–9830.

7. Chukoskie, L., and Movshon, J.A. (2009). Modulation of visual signals in macaque MT and MST neurons during pursuit eye movement. J. Neurophysiol. *102*, 3225–3233.

8. Inaba, N., Miura, K., and Kawano, K. (2011). Direction and speed tuning to visual motion in cortical areas MT and MSTd during smooth pursuit eye movements. J. Neurophysiol. *105*, 1531–1545.

9. Duhamel, J.R., Colby, C.L., and Goldberg, M.E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. Science *255*, 90–92.

10. Morris, A.P., Chambers, C.D., and Mattingley, J.B. (2007). Parietal stimulation destabilizes spatial updating across saccadic eye movements. Proc. Natl. Acad. Sci. USA *104*, 9069–9074.

11. Duhamel, J.R., Bremmer, F., Ben Hamed, S., and Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. Nature *389*, 845–848.

12. Galletti, C., Battaglini, P.P., and Fattori, P. (1995). Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. Eur. J. Neurosci. *7*, 2486–2501.

13. Inaba, N., Shinomoto, S., Yamane, S., Takemura, A., and Kawano, K. (2007). MST neurons code for visual motion in space independent of pursuit eye movements. J. Neurophysiol. *97*, 3473–3483.

14. Mullette-Gillman, O.A., Cohen, Y.E., and Groh, J.M. (2005). Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. J. Neurophysiol. *94*, 2331–2352.

15. Nakamura, K., and Colby, C.L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. Proc. Natl. Acad. Sci. USA *99*, 4026–4031.

16. Hall, N.J., and Colby, C.L. (2011). Remapping for visual stability. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 528–539.

17. Pouget, A., and Sejnowski, T.J. (1997). Spatial transformations in the parietal cortex using basis functions. J. Cogn. Neurosci. *9*, 222–237.

18. Pouget, A., Fisher, S.A., and Sejnowski, T.J. (1993). Egocentric spaw representation in early vision. J. Cogn. Neurosci. *5*, 150–161.

19. Lal, R., and Friedlander, M.J. (1989). Gating of retinal transmission by afferent eye position and movement signals. Science *243*, 93–96.

20. Trotter, Y., and Celebrini, S. (1999). Gaze direction controls response gain in primary visual-cortex neurons. Nature *398*, 239–242.

21. Rosenbluth, D., and Allman, J.M. (2002). The effect of gaze angle and fixation distance on the responses of neurons in V1, V2, and V4. Neuron *33*, 143–149.

22. Durand, J.B., Trotter, Y., and Celebrini, S. (2010). Privileged processing of the straight-ahead direction in primate area V1. Neuron *66*, 126–137.

23. Weyand, T.G., and Malpeli, J.G. (1993). Responses of neurons in primary visual cortex are modulated by eye position. J. Neurophysiol. *69*, 2258–2260.

24. Andersen, R.A., Essick, G.K., and Siegel, R.M. (1985). Encoding of spatial location by posterior parietal neurons. Science *230*, 456–458.

25. Andersen, R.A., and Mountcastle, V.B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. J. Neurosci. *3*, 532–548.

26. Deneve, S., Latham, P.E., and Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. Nat. Neurosci. *4*, 826–831.

27. Morris, A.P., Bremmer, F., and Krekelberg, B. (2013). Eye-position signals in the dorsal visual system are accurate and precise on short timescales. J. Neurosci. *33*, 12395–12406.

28. Graf, A.B., and Andersen, R.A. (2014). Inferring eye position from populations of lateral intraparietal neurons. eLife *3*, e02813.

29. Boussaoud, D., and Bremmer, F. (1999). Gaze effects in the cerebral cortex: reference frames for space coding and action. Exp. Brain Res. *128*, 170–180.

30. Morris, A.P., Bremmer, F., and Krekelberg, B. (2016). The dorsal visual system predicts future and remembers past eye position. Front. Syst. Neurosci. *10*, 9.

31. Morris, A.P., Kubischik, M., Hoffmann, K.P., Krekelberg, B., and Bremmer, F. (2012). Dynamics of eye-position signals in the dorsal visual system. Curr. Biol. *22*, 173–179.

32. Becker, W. (1991). Saccades. In Eye Movements, R.H.S. Carpenter, ed. (MacMillan), pp. 95–137.

33. Lisberger, S.G., Morris, E.J., and Tychsen, L. (1987). Visual motion processing and sensory-motor integration for smooth pursuit eye movements. Annu. Rev. Neurosci. *10*, 97–129.

34. Zipser, D., and Andersen, R.A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature *331*, 679–684.

35. Nishida, S., Motoyoshi, I., Andersen, R.A., and Shimojo, S. (2003). Gaze modulation of visual aftereffects. Vision Res. *43*, 639–649.

36. Mayhew, J.E.W. (1973). After-effects of movement contingent on direction of gaze. Vision Res. *13*, 877–880.

37. Durand, J.B., Camors, D., Trotter, Y., and Celebrini, S. (2012). Privileged visual processing of the straight-ahead direction in humans. J. Vis. *12*, 34.

38. McFarland, J.M., Bondy, A.G., Cumming, B.G., and Butts, D.A. (2014). High-resolution eye tracking using V1 neuron activity. Nat. Commun. *5*, 4605.

40. Burak, Y., Rokni, U., Meister, M., and Sompolinsky, H. (2010). Bayesian model of dynamic image stabilization in the visual system. Proc. Natl. Acad. Sci. USA *107*, 19525–19530.

41. Pitkow, X., Sompolinsky, H., and Meister, M. (2007). A neural computation for visual acuity in the presence of eye movements. PLoS Biol. *5*, e331.

42. Xu, B.Y., Karachi, C., and Goldberg, M.E. (2012). The postsaccadic unreliability of gain fields renders it unlikely that the motor system can use them to calculate target position in space. Neuron *76*, 1201–1209.

43. Pouget, A., Deneve, S., and Duhamel, J.R. (2002). A computational perspective on the neural basis of multisensory spatial representations. Nat. Rev. Neurosci. *3*, 741–747.

44. Gerstner, W., Kreiter, A.K., Markram, H., and Herz, A.V. (1997). Neural codes: firing rates and beyond. Proc. Natl. Acad. Sci. USA *94*, 12740–12741.

45. Roth, N., and Rust, N.C. (2019). Rethinking assumptions about how trial and nuisance variability impact neural task performance in a fast-processing regime. J. Neurophysiol. *121*, 115–130.

46. Donaldson, I.M. (2000). The functions of the proprioceptors of the eye muscles. Philos. Trans. R. Soc. Lond. B Biol. Sci. *355*, 1685–1754.

47. Donaldson, I.M., and Dixon, R.A. (1980). Excitation of units in the lateral geniculate and contiguous nuclei of the cat by stretch of extrinsic ocular muscles. Exp. Brain Res. *38*, 245–255.

48. Wang, X., Zhang, M., Cohen, I.S., and Goldberg, M.E. (2007). The proprioceptive representation of eye position in monkey primary somatosensory cortex. Nat. Neurosci. *10*, 640–646.

49. Xu, Y., Wang, X., Peck, C., and Goldberg, M.E. (2011). The time course of the tonic oculomotor proprioceptive signal in area 3a of somatosensory cortex. J. Neurophysiol. *106*, 71–77.

50. Buisseret, P. (1995). Influence of extraocular muscle proprioception on vision. Physiol. Rev. *75*, 323–338.

51. Chen, C.Y., Ignashchenkova, A., Thier, P., and Hafed, Z.M. (2015). Neuronal response gain enhancement prior to microsaccades. Curr. Biol. *25*, 2065–2074.

52. Morris, A.P. (2015). Neuroscience: tiny eye movements link vision and attention. Curr. Biol. *25*, R769–R771.

53. Sommer, M.A., and Wurtz, R.H. (2002). A pathway in primate brain for internal monitoring of movements. Science *296*, 1480–1482.

54. Dowiasch, S., Blohm, G., and Bremmer, F. (2016). Neural correlate of spatial (mis-)localization during smooth eye movements. Eur. J. Neurosci. *44*, 1846–1855.

55. Ibbotson, M., and Krekelberg, B. (2011). Visual perception and saccadic eye movements. Curr. Opin. Neurobiol. *21*, 553–558.

56. Klingenhoefer, S., and Krekelberg, B. (2017). Perisaccadic visual perception. J. Vis. *17*, 16.

57. Bridgeman, B., Dassonville, P., and Lester, B.D. (2018). The Roelofs and induced Roelofs effects. Conscious. Cogn. *64*, 6–12.

58. Ross, J., Morrone, M.C., Goldberg, M.E., and Burr, D.C. (2001). Changes in visual perception at the time of saccades. Trends Neurosci. *24*, 113–121.

59. Mirpour, K., and Bisley, J.W. (2016). Remapping, spatial stability, and temporal continuity: from the pre-saccadic to postsaccadic representation of visual space in LIP. Cereb. Cortex *26*, 3183–3195.

60. Bremmer, F. (2000). Eye position effects in macaque area V4. Neuroreport *11*, 1277–1283.

61. Nowicka, A., and Ringo, J.L. (2000). Eye position-sensitive units in hippocampal formation and in inferotemporal cortex of the macaque monkey. Eur. J. Neurosci. *12*, 751–759.

62. Duijnhouwer, J., and Krekelberg, B. (2016). Evidence and counterevidence in motion perception. Cereb. Cortex *26*, 4602–4612.

63. Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput. *16*, 1661–1687.

64. Friedman, H.S., and Priebe, C.E. (1998). Estimating stimulus response latency. J. Neurosci. Methods *83*, 185–194.

65. Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nat. Rev. Neurosci. *7*, 358–366.

66. Graf, A.B., and Andersen, R.A. (2015). Predicting oculomotor behaviour from correlated populations of posterior parietal neurons. Nat. Commun. *6*, 6024.

67. Dai, J., and Wang, Y. (2012). Representation of surface luminance and contrast in primary visual cortex. Cereb. Cortex *22*, 776–787.

68. Kinoshita, M., and Komatsu, H. (2001). Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. J. Neurophysiol. *86*, 2559–2570.

69. McFarland, J.M., Bondy, A.G., Saunders, R.C., Cumming, B.G., and Butts, D.A. (2015). Saccadic modulation of stimulus processing in primary visual cortex. Nat. Commun. *6*, 8110.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Rhesus (Macaca Mulatta)	Covance, Inc., Texas	
Software and Algorithms		
MATLAB, 2014-2017a	Mathworks Inc.	https://www.mathworks.com
Neurostim Experimental Control software	SourceForge	https://sourceforge.net/projects/neurostim/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Adam Morris (adam.morris@monash.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects
Behavioral and electrophysiological measurements were performed in two adult, male macaques (*Macaca Mulatta*; M1: weight = 14 Kg; M2: weight = 9.9 Kg). All procedures conformed to the National Institute of Health guidelines for the Humane Care and Use of Laboratory animals and were approved by the Rutgers University Animal Care and Use Committee.

Animal preparation
Surgery was performed under aseptic conditions and general anesthesia (isoflurane), and analgesics (ibuprofen and buprenorphine) and antibiotics were delivered during recovery. Animals were first fitted with a titanium head-post (Gray Matter Research) to stabilize the head during recording sessions using standard surgical procedures. After recovery, and completion of simple fixation training, animals were prepared for chronic, subdural implantation of one (M2; left-hemisphere) or two (M1; right-hemisphere) 32-channel, planar microelectrode arrays (Floating Microelectrode Arrays; MicroProbes for Life Science) into the parafoveal region of V1. Each array consisted of a 4 × 8 arrangement of platinum/iridium electrodes (0.6–1.5 mm in length [tiered]), spanning 1.2 × 3.4 mm and separated by a distance of 0.44 mm (0.8–1 MΩ impedance at 1 kHz). A craniotomy was performed immediately anterior to the occipital ridge and lateral from the midline, and a small patch of dura was removed to provide access to the parafoveal region of V1. Each array was implanted normal to and flush with the cortical surface using manual descent and depth control via a stereotax-mounted micromanipulator [62]. Neural signals were communicated via a fine bundle of wires that was routed out and along the margin of the craniotomy to Omnetics connector(s) mounted in a stainless steel chassis (M2) or dental acrylic (M1). The two arrays in M1 were implanted directly side-by-side in the same hemisphere. Full closure of the craniotomy was achieved using a layer of DuraGen (Integra) and dental acrylic to reduce risk of infection (no subsequent signs of infection were observed). The animals were allowed to fully recover before the commencement of electrophysiological recordings.

METHOD DETAILS

Stimuli and apparatus
Visual stimuli were presented on a CRT display (Sony GDM-520, NVidia 8800 GT graphics card) at a vertical refresh rate of 150 Hz and resolution of 1024 × 768. The visible display area subtended 43.7° × 33.6° at a center-aligned viewing distance of 50 cm and was linearized for luminance at [CIEx, CIEy] = [0.33, 0.33]. Stimulus presentation and experimental flow was controlled by custom software written in C++ (Neurostim, available at https://sourceforge.net/projects/neurostim/; last accessed March, 2019). The animal was seated in a primate chair with its head stabilized using a head-post. During most experimental trials, neurons were stimulated visually by dynamic binary noise, in which every pixel in the display was switched to black (0.6 cd/m^2) or white (50 cd/m^2) randomly and independently on every second retrace of the display (and maintained its value for the intervening frame). This resulted in an effective framerate of 75 Hz and a mean luminance of 25.3 cd/m^2. In a subset of trials, the screen was a uniform gray matching this mean luminance rather than filled with noise.

Fixation and pursuit targets were red spots (diameter = 0.25°) placed atop a larger black spot (diameter = 0.63°) that masked the surrounding noise pixels. Left and/or right eye position was recorded using an infrared video-based eye tracker (Eyelink II, SR Research) at a sample rate of 250 or 500 Hz, interpolated offline to 1 kHz, and temporally aligned to the neural data.

Behavioral tasks

Each trial began with a fixation target presented at one of nine positions (eight equally spaced around a virtual circle [radius = 7.5°] and one in the center of the display) against a gray background (25.3 cd/m^2; Figure 2A). When ready, the animal fixated the target to initiate the trial proper. The onset of fixation triggered the presentation of binary noise across every pixel in the display, which continued until the end of the trial. On saccade trials, the animal maintained fixation for 1100 ms before the target stepped to the diametrically opposite position on the circle. The animal was required to saccade to the target within 450 ms and hold fixation for a further 1100 ms. On pursuit trials, the target started to move after animal had maintained initial fixation for 500 ms, and then moved at constant speed (10°/s [i.e., degrees of visual angle]) around the circle (clockwise or counter-clockwise) to the diametrically opposite position (at which time the trial ended). In the case of the central fixation target, the animal was required to maintain fixation for the duration of the trial. Trials in which the animal failed to maintain gaze within 1.5° of the target were terminated immediately without reward, while successfully completed trials were rewarded with diluted fruit juice. The screen returned to gray during the 1200 ms inter-trial interval and the animal was free to look in any direction. Animals M1 and M2 completed an average of 71.26 (SD = 16.88) and 50.90 (SD = 23.69) saccade trials correctly per starting position, per session. They also completed an average of 36.51 (SD = 7.95) and 21.46 (SD = 4.87) pursuit trials correctly per starting position, per pursuit direction (clockwise, counter-clockwise), per session.

Electrophysiology and data processing

Signals from the V1 microelectrode array were digitized (14 bit at 25 kHz) and band-passed (Butterworth, 120 Hz at 12 dB/octave to 6 kHz at 24 dB/octave) using AlphaLab recording hardware (Alpha Omega Co.). Candidate action potentials were detected using a threshold of 3.5 standard deviations separately for each electrode and sorted into clusters using an automated algorithm (superparamagnetic clustering [63]). Each cluster was categorized by hand as a single-unit, multi-unit, or junk on the basis of its (1) mean cluster waveform; (2) inter-spike interval histogram, and (3) visual responsivity (assessed by peristimulus time histograms of spike counts aligned to the onset of the binary noise). All analyses were performed on single- and multi-units without distinction unless otherwise specified.

QUANTIFICATION AND STATISTICAL ANALYSIS

Estimation of receptive fields

The linear receptive-field (RF) of each unit was estimated from the fixation/saccade trials using a standard spike-triggered average (STA). The approximate locations of the RFs were known from preliminary analyses and previous recordings. Accordingly, the STA was calculated only for noise pixels within a 10.7° × 10.7° (251 × 251 pixels) square region of interest (ROI) centered on the position [−4°, −1°] relative to the fixation point (i.e., from different screen locations depending on the fixation location). Each pixel subtended 2' 33'' (0.04°) of visual angle.

The luminance values (coded as 0 [black] and 1 [white]) of every pixel were collated into an $m \times n$ matrix, L, where m is the number of pixels in the ROI and n is the total number of frames (pooled across time and all trials). The spike times for each unit were aligned to the first frame of the stimulus and binned by counting the number of spikes that occurred within each stimulus frame (i.e., 13.33 ms bins). To take into account visual latency, the resulting n-length column vector, C, was replicated 11 times, each with an additional frame of time lag (i.e., shifted from 13 to 147 ms), and compiled into an $m \times 11$ matrix. The space-time STA ($m \times 11$) was then calculated through matrix multiplication ($L*C$). The strongest STAs were found in the three lags from 54 to 80 ms prior to the time of the spike, hence we calculated the spatial receptive field maps in Figure 2 by averaging these three time lags and smoothing with a 2D Gaussian filter (5 × 5 pixels, SD = 2). These visual latencies estimated using the STA approach were consistent with those estimated from onset responses to the noise patterns (53.11ms) using the method of Friedman and Priebe (1998); that is, using only data from 150 ms before to 200 ms after stimulus onset [64].

To identify pixels in the STA that were statistically significant, we created an empirical null distribution of STA magnitudes (including the effects of smoothing and other averaging) by repeating our analysis across the entire 1024 × 768 stimulus. These 786,432 pixels could be used as a measure of sampling variability because we knew from previous recordings that almost of them were well outside the tiny V1 RFs. Pixels in the ROI STA were deemed significant if their absolute value exceeded 3.5 times the standard deviation of this null distribution (i.e., p < 0.0002). The fact that a small subset of pixels was inside the RF made this a conservative estimate.

To summarize the RFs across all units in an animal (Figure 2C), we calculated population RF densities by convolving each unit's significance map (where significant and non-significant pixels had values of one and zero, respectively) with a 2D Gaussian filter (10 × 10 pixels, SD = 4) and then summing across all units. Pixels in individual STAs that were false positives would occur at random positions and thus would tend not to accumulate across units and be largely absent in the population density image.

Normalized population activity

To compute the normalized activity in Figure 4C, we calculated the firing rate in sliding windows (100 ms, stepped by 25 ms), divided this measure by the mean response in the fixation window (−700 ms to −300 ms relative to saccade onset), and then averaged across all neurons. This measure was calculated separately for trials in which the flickering background stimulus was shown and for trials in which the background was uniform gray.

Eye position gain-fields

Gain-fields were estimated from the fixation/saccade trials. Spike times were aligned to the offset of the primary saccade from the first fixation position to the opposite side of the circle. This eye movement was found offline and defined as the first saccade (detected by the Eyelink system) after the change in target position that had an amplitude of at least 75% of the required distance (15°). Spike times on central fixation trials, which did not require a saccade, were aligned to the mean saccadic offset time from the saccade trials (i.e., the time of the target step plus mean saccade latency and duration) to maintain consistent alignment. Spike times were then binned into 100 ms windows stepped in 25 ms increments to provide a spike-count time course for each trial.

Gain-fields are defined as a systematic change in the mean firing rate of a visual neuron as a function of fixation position despite constant stimulation of the receptive field. Accordingly, we calculated the mean spike-count during the initial fixation period (from −700 ms to −400 ms, equivalent to −570 to 870 ms after the onset of the visual noise on average) separately for each of the nine fixation positions. For plotting, we calculated the standard error of the mean (STE) for each of the nine eye positions, as shown in Figure. 2D. A statistically significant gain-field was identified using stepwise regression, in which a second–order polynomial was used to model the effect of eye-position (azimuth [X] and elevation [Y]) on mean spike counts, $\widehat{c}$ (Equation 1). Coefficient estimates and the associated F-tests were performed using the "stepwiseglm" function in MATLAB (Mathworks Inc.). To prevent cardinal biases all terms were retained for decoding (see below), even if the coefficients were individually non-significant [27].

$$\widehat{c}(X, Y) = a_0 + a_1 X + a_2 Y + a_3 X^2 + a_4 Y^2 + a_5 XY \qquad \text{(Equation 1)}$$

Population decoding of eye position

Gain-fields provide an intuitive quantitative link between the activity of a neuron and eye-position. Such a link can form the basis of decoding algorithms that predict or estimate eye position using only neural activity across one or more neurons. We have used this approach successfully to decode eye position from populations of neurons across many regions in the dorsal visual system during stationary gaze [27], and to examine the dynamics of these representations when the eyes move [30, 31] (see also [54]). In one such study [27], we developed a decoding approach that estimates the *most likely* eye position given a set of observed spike counts across a sample of neurons. The method is built on the well-established statistical approach of maximum-likelihood estimation, which takes into account not only how the mean firing rates depends on eye position, but also the associated variability (sometimes referred to as neural "noise"). Here we applied this decoding approach to all V1 units that had significant gain-fields (219/397). The full method has been described in our previous work [27] but is summarized here.

Maximum-likelihood estimation

The traditional definition of a gain-field quantifies only how the mean firing rate of a neuron depends on eye position and says nothing about spike-count variability. We introduced a richer statistical quantification known as a *probabilistic gain-field* (pGF) [27]. Given an eye position, the relative likelihood of observing each possible spike-count (i.e., $c = 0, 1, 2, 3$, etc.) is captured by a probability density function (pdf), $P(C \mid x, y)$. Our approach is to estimate this generative pdf at each of the experimentally measured eye positions and then to interpolate (in 2D) to unmeasured fixation positions. By doing so with a continuous expression, we can express $P(C \mid x, y)$ at arbitrary eye positions within or near the sampled range – this representation is the pGF. For decoding, however, we want to come from the other direction: given an observed spike-count, we want to know the relative probability of all possible eye positions, $P(X, Y \mid c)$. This *a posteriori* density can be computed from the pGF using Bayes Rule (we assumed a uniform prior over all eye positions) and then used to generate a point estimate by choosing the most likely current eye position.

The approach extends naturally to a population of neurons. With a pGF in hand for each of N recorded V1 units, a vector of observed spike-counts across the sample, c_{pop}, can be converted into a population estimate of $P(X, Y \mid c_{pop})$ by combining the information across units, implemented as the sum of the (log) posterior probabilities:

$$\log P(X, Y \mid c_{pop}) = \sum_{i=1}^{N} \log P(X, Y \mid c_i) \qquad \text{(Equation 2)}$$

The final estimate of eye-position is taken as the position associated with the highest likelihood, $[x_{max}, y_{max}]$.

For this study, we decoded across a discrete grid of possible eye positions spanning ± 30° in 0.5° steps both horizontally (azimuth) and vertically (elevation). The spike-count vector, c_{pop}, was compiled by taking a random trial for each unit from a common experimental condition and treating them as if they occurred simultaneously. This pseudo-population response approach was preferred because our dataset included only 7.49 visually-responsive units per recording session on average (SD = 5.35), which was insufficient to decode the eye reliably for all positions. A consequence of this randomization, however, is that it removes correlated variability across the population of neurons, which can influence the information that is available to a decoder (for better or worse, depending on its structure) [65, 66]. Our estimates of decoder precision (i.e., moment-to-moment variability) thus incorporate this potentially incorrect assumption of independence and should be interpreted accordingly.

As described above, we needed an estimate of $P(C \mid x, y)$ at each measured eye position. Ideally, we would use the empirically observed distribution of spike counts for a given unit for this purpose. In studies of this kind, however, it is usually not possible to perform enough measurements to estimate the underlying probability model reliably, particularly for rare spike-count values (which can then cause instability in the decoder). Instead, as in previous work [27, 28] we made the simplifying assumption that the variability can be approximated by Poisson distributions, in which the mean and variance are equal. This has the convenient property that the

traditional gain-field function, which quantifies how mean evoked activity varies with eye position, can be reinterpreted as a continuous expression of the value of the single-parameter (λ) Poisson pdf. Accordingly, Equation 1 was used as a generative model for the pGF of each unit.

Decoding during fixation

The previous section described how we built the decoder from the gain-fields of each recorded unit. To assess how well V1 represented the eye during fixation, the decoder was applied to spike-counts in 100 ms wide windows during the same interval that was used to fit the gain-fields (i.e., −700 ms to −400 ms relative to saccade offset). This was achieved using leave-one out cross-validation, in which all but one trial from each of the nine eye-positions was used to fit the gain-fields (and hence, produce a pGF for each unit). The resultant decoder was then applied to the remaining trial from each condition and the estimated eye position was recorded. This was repeated 1000 times to produce a distribution of estimates for each eye position. The 2D median of each of these distributions (plotted in Figure 3) was used to summarize the accuracy of the V1 representation, and the span of the middle 50% of the data (horizontally and vertically) was used to represent its precision (error bars in Figure 3).

For all subsequent analyses (described below) we used a decoder built from this initial fixation interval but without leaving out any trials. This fixation decoder was then locked down – no further learning could take place – and applied to either different time points to which it was naive (e.g., during the saccade), or to different trial types (e.g., pursuit trials).

Decoding across saccades

One-thousand pseudo-population responses were compiled for each starting position in the saccade task and then decoded using the decoder described above. Each time point, from 900 ms before the saccade to 900 ms after, was decoded sequentially and independently and with no knowledge of the true eye position. Hence, only changes in the activity of V1 neurons over time could lead to changes in the decoded eye positions. The result was a distribution of 1000 decoded time courses for each condition. As for the fixation analysis, we used the median and width of these distributions at each time point to assess the accuracy and precision of the decoder. To summarize overall performance, we combined all conditions by rotating them onto a common starting position (the 9 o'clock position) and then taking the mean (of the medians; Figure 4B).

Of key interest was the timing and speed with which the recorded units transitioned from a representation of the initial fixation to that of the final position at the diametrically opposite position on the circle. Like the real eye, the decoded eye showed a sigmoid-like change over time in the saccading channel, while remaining relatively stationary in the orthogonal channel. To quantify this transition, we fitted a cumulative Gaussian (using MATLAB's "lsqcurvefit" function) to the saccade channel. The mean parameter (μ) of the fit corresponded to the mid-point of the represented saccade and so was used as the time at which updating occurred. The duration was taken as the interval between the 5th and 95th percentile (i.e., saccade duration = 2*1.65*σ). The same method was used to quantify the real eye, which was also binned in 100 ms for consistency with the neural data. This binning dampens the signals, which is why the average saccade duration reported here (114 ms) is longer than its true duration (33.1 ms). This does not affect our conclusions because identical binning was used for the eye and neural data. The difference between the two updating times ($\sigma_{decoded} − \sigma_{real}$) was used as the estimated latency of the eye-position signals represented in V1 (i.e., the delay from a true change in eye position to when that change is represented in cortex).

Decoding during smooth pursuit

The spike times on pursuit trials were aligned to the onset of the target motion and binned in the same way as for saccade trials. Because pursuit trials were not included in all recording sessions, some units had to be removed from the decoder. In all other respects, however, the decoder remained unchanged and hence had never been exposed to a single data point from pursuit trials. We decoded the neural data in the same manner as described above for saccade trials and extracted the same measures of accuracy and precision over time.

Both the real and decoded eye position traced out approximate sine/cosine functions over time for the horizontal and vertical channels. We fitted such functions (Equations 3 and 4) to the averaged decoded time courses (Figure 5B) and to the real eye and used the difference in phase between them to estimate the latency of the V1 eye-position signal. Data from the first 400 ms after the target motion onset (where pursuit was not yet stable) was excluded.

$$\widehat{x}(t) = \alpha\sin(\beta t + \varphi) \tag{Equation 3}$$

$$\widehat{y}(t) = \alpha\cos(\beta t + \varphi) \tag{Equation 4}$$

To ensure that all temporal differences manifest in the phase (φ), the period parameter (β) was constrained to be the same for the real and decoded eye position traces. This was achieved by first fitting Equations 3 and 4 to the two datasets separately, and then refitting them with β fixed to the average of the two values from the first stage.

Decoder performance during the initial fixation period of pursuit trials (i.e., −200 ms to +100 ms relative to target motion onset) was examined for multiple purposes. First, it allowed us to compare the spatial errors with those of the saccade trials for this reduced population size (Figure 3). Second, and more importantly, we used these errors to construct baseline time courses for how the decoder should perform during smooth pursuit *if the neural code for eye position remained identical to that of a stationary eye* – that is, if neural firing rates depended only on the instantaneous eye position (Figure 5).

Each pursuit condition passed through five of the positions measured during fixation (including the start- and end-points) during its half-circle trajectory, which we refer to as "waypoints." To generate expected trajectories for each condition, we first fitted a 2D cubic

spline through the eight decoded positions associated with each waypoint (i.e., the dotted line in Figure 3). Spatially, the predicted pursuit signal should travel around this spline, but how it should do so in time is complex and requires explication.

As a hypothetical example, suppose the decoded eye position for the 9 o'clock position during fixation was 1° above the true position. What would happen if this same error occurred during pursuit at the moment the eye passed *through* the 9 o'clock waypoint? The decoder would appear to *lead* the eye during clockwise pursuit because the decoded position would already be ~1° ahead of the upward moving eye. Conversely, the same error would look like a temporal *lag* when the eyes were moving counter clockwise because the decoded position would remain ~1° behind the downward moving eye. Thus, although the real eye travels from one waypoint to the next with approximately constant speed and distance, the signal that is predicted on the basis of the fixation performance should travel different distances depending on the errors associated with the departure and arrival waypoints.

With this in mind, we computed a predicted pursuit signal by (1) noting the times at which the eye crossed each waypoint, and (2) moving around the cubic spline from the (erroneous) position associated with each waypoint to that of the next within the same time interval as the real eye. These predictions are plotted in Figure 5, and the pursuit decoding performance is expressed as the correlation with these predictions.

Comparison of single unit versus multi-unit activity

Most analyses were done on a dataset in which we combined single and multi-unit activity. This is warranted given that the properties of single units and multi-unit activity were quite similar. For instance, 55% of single units (127/232) and 56% of multi-units (92/165) had significant eye position modulation. We also found that the decoding results were highly similar. For instance, decoding the eye position across fast eye movements using only the single-units (127/219) in our sample led to an almost identical time course compared to the one shown in Figure 4B (the correlation was 0.997; data not shown). This suggests that neighboring neurons share eye-position modulation (otherwise, one would expect the modulation to average out in multi-unit activity). The limited spatial extent of the multielectrode recording arrays (2 × 3 mm), however, prevented us from assessing whether this reflects a local clustering or a global organization across the cortical map such as the bias for straight ahead locations that has been found in the peripheral visual field [37].

Control experiments

The binary noise background was designed to stimulate neurons in the same way irrespective of eye position. The CRT display, however, subtended 43.7° × 33.6° of visual angle and so there was a peripheral visible edge between the luminance of the display and surrounding darkness that changed position on the retina depending on the fixation position. This edge was at least 6° away from the receptive fields at all eye positions, making it unlikely to have much influence on firing rates. Trotter & Celebrini [20], who found gain-fields in V1 in similar proportions to ours even when moving the display with each change in eye position to keep the retinal projection constant. Nevertheless, effects of luminance and contrast in the far periphery have been reported in V1 [67–69]. Control experiments (one animal only; 15 sessions) quantified any such effects, and ultimately, ruled out an influence of the display edge on the findings reported in the main paper.

We interleaved trials of the fixation task from our main experiment (but without requiring a saccade) with simulated eccentric fixation trials. In the simulated trials, the animal fixated centrally but a dark band (0.04 cd/m^2) was presented on the display in a way that mimicked the darkness beyond the screen edge for a particular eye position. For example, for a simulated upper-right fixation, the band was presented along the upper and right margins of the display, at the same position and distance relative to the fovea as when truly fixating eccentrically. If the screen edge was the source of eye position effects in our V1 neurons, these bands in the visual periphery should drive similar modulations in firing rates. That is, mean firing rates for *real* and *simulated* eye positions should correlate and a regression line-of best-fit should have a slope of ~1. This was not the case (Figure S1).

Assessing visual stability

A neural code for vision is stable if it remains spatially accurate even when the eyes move. In the current context, this requires encoding an object's position relative to the head (or equivalently, because the animal was head-fixed, relative to the display), because this position remains the same irrespective of eye position. To a good approximation, the head-centered location of an object is equal to the sum of two quantities: its projected position onto the retina (i.e., its horizontal and vertical distance from the fovea, in degrees of visual angle), and eye position (i.e., the horizontal and vertical angles of rotation in the orbit). Assuming a stationary object, both of these quantities change every time the eyes move but their sum remains constant.

Given that objects move in the real world, however, all terms in this linear relation are functions of time. Moreover, the cortical representation of an object's position on the retina, $\widehat{r}$, presumably encoded via retina-fixed receptive fields [2], does not reflect its current position, but rather where it was some time ago, reflecting the transmission delay (visual latency, V_L) from the retina to the cortex. That is, assuming a perfect cortical code:

$$\widehat{r}(t - V_L) = h(t - V_L) - e(t - V_L) \qquad \text{(Equation 5)}$$

where h is the true location of the object in the world and e is the true eye position.

Similarly, the brain does not have access to eye position directly but must estimate it from an internal signal (e.g., efference copy from the motor system or proprioception originating in the eye musculature). To maintain spatiotemporal alignment with the incoming visual information, this internal eye-position signal, $\widehat{e}$, must but also lag behind the eye by the visual latency (as well as faithfully

representing the dynamics of the eye). In that case, the combination of retinal and eye-position signals gives rise to a stable internal representation of the outside world (at least, as it was V_L milliseconds ago).

To assess visual stability in V1 within this framework, we simulated an object moving across the display during the saccade and pursuit tasks, and used the observed V1 data to reconstruct a neural estimate of its position over time. The simulated object moved at a constant velocity in the fronto-parallel plane, which varied in speed and direction across simulated trials (8 evenly spaced directions between 0 and 2π, 5 speeds [0°/s, 2°/s, 4°/s, 8°/s, 16°/s]). A neural estimate of the object's position was calculated by combining the decoded eye position signal (i.e., the time course shown in Figure 4B or Figure 5B) with an idealized representation of the object's position on the retina. The latter was calculated as the difference between the position of the simulated object and the true eye position (as reported by the eye tracker, also shown in Figure 4B). This representation of the retinal projection was then lagged by the visual latency.

The two quantities – visual and eye-position – were represented as [x,y] vectors and summed to provide an estimate of the object's position at each moment in time, across saccades and during pursuit. That is:

$$\widehat{h}(t) = \widehat{r}(t) + \widehat{e}(t) \qquad \text{(Equation 6)}$$

We compared the estimated and true object positions to quantify visual stability in V1. We first calculated a traditional sum-of-squared errors to quantify the spatial mismatch (i.e., 2D Euclidean distance) between the predicted object trajectory and its true trajectory:

$$d(t) = \sqrt{\left(\widehat{h}_x(t) - h_x(t)\right)^2 + \left(\widehat{h}_y(t) - h_y(t)\right)^2}$$
$$SSE_{\widehat{r}+\widehat{e}} = \sum_{t=1}^{N} d(t)^2 \qquad \text{(Equation 7)}$$

To provide a meaningful reference, we compared this error value with that observed if we used only the retinal signal for localization (i.e., only the object's position within the V1 map) and ignored the eye-position signal. That is:

$$d(t) = \sqrt{\left(r_x(t) - h_x(t)\right)^2 + \left(r_y(t) - h_y(t)\right)^2}$$
$$SSE_{\widehat{r}} = \sum_{t=1}^{N} d(t)^2 \qquad \text{(Equation 8)}$$

We defined a compensation index (CI) to capture how well gain-fields allowed V1 to compensate for the effects of eye movements:

$$CI = 1 - \frac{SSE_{\widehat{r}+\widehat{e}}}{SSE_{\widehat{r}}} \qquad \text{(Equation 9)}$$

The index has a value of zero if the prediction is no better than using V1 receptive fields alone, and 1 in the case of perfect prediction (the upper-bound). The index has no lower-bound (negative values are possible).

The CI index was computed for all of the simulated object trajectories during saccades, and during both pursuit directions (clockwise and counter-clockwise).

DATA AND SOFTWARE AVAILABILITY

Data were analyzed in MATLAB (Mathworks Inc.) using a range of built-in, external, and in-house toolboxes, classes, and scripts, as well as in-house databases. Details and access to in-house resources are available on request. Processed data files, including spike counts and eye tracking data, are available on request.

Supplemental Information

A Stable Visual World

in Primate Primary Visual Cortex

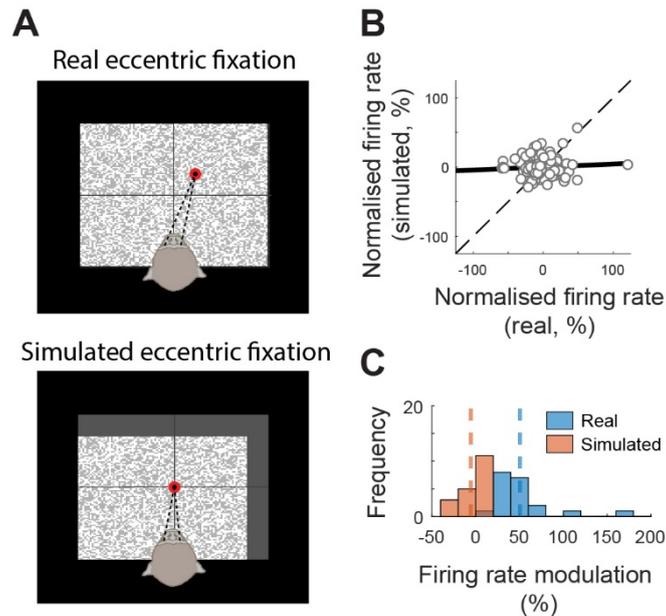Adam P. Morris and Bart Krekelberg

Figure S1. Gain-fields are not an artefact of the visible screen edge, related to Figure 2. (A) Control experiments interleaved trials of real eccentric fixation (at each of the 8 positions on the circle used in the main experiment) and a corresponding set of conditions in which eccentric fixation was simulated by matching the pattern of retinal stimulation. Here, a schematic of the experimental configuration for the upper-right eye position is shown (the display is drawn to scale; quadrant lines are shown for illustration only and were not present on the display). In the simulated trials, a dark band was presented on the display (here shown as grey for clarity) to mimic the darkness beyond the screen edge (black region) and was approximately matched for luminance. (B) Firing rates were calculated in a time window that approximately matched those used to define fixation in the main experiment (i.e. 600-900 ms after the onset of the noise stimulus). Significant gain-fields were observed in 20/56 units for real changes in eye position, but only one of these had an effect in the simulated condition (in that case, the two GFs were anti-correlated). Each data point in the scatterplot represents the mean firing rate for one unit at one real/simulated eye position (expressed as a percentage difference relative to the unit's global mean rate; N = 20 × 8). The correlation between these values was close to zero at the population level (Pearson's r = 0.05, p=0.54) and non-significant for all but one unit when calculated individually (the anti-correlated case noted above). None were significant after correcting for multiple comparisons. Further, the line of best fit (orthogonal regression, black line) to the population data had a slope close to zero (0.05), showing that real changes in eye position accounted for the overwhelming majority of response variance. (C) As a further test, we compared the strength of modulation between the "best" and "worst" eye position for real changes in eye position with that observed in the corresponding simulated conditions. Strong modulation was observed for real changes (mean = 49%) but not simulated changes (mean = -6%%). Finally, seven units that did not have a significant gain-field for real changes in eye position did have a significant effect of simulated eye positions. In those cases, again, there were no significant correlations between the two conditions. In sum, although there were weak visual effects in a small number of neurons, they were statistically dissimilar to

the changes in firing rate that occurred for real changes in eye position and thus cannot account for the decoding results presented in the main paper (which themselves were cross-validated to ensure statistical reliability).
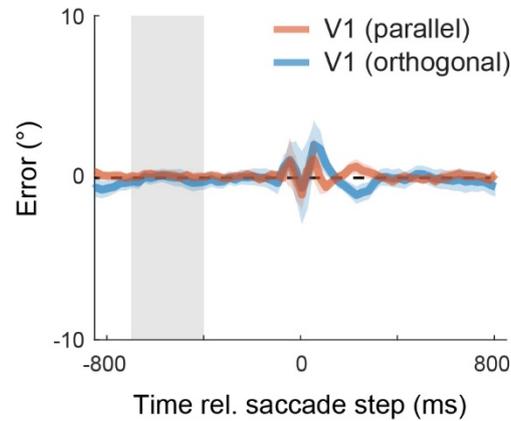

Figure S2. **Transient errors in the representation of the saccade step, related to Figure 4.** Curves represent the error between the true eye position and the decoded eye position, averaged over the eight saccade directions (± 1 STE of the mean across conditions). The data were rotated before averaging (as if all saccade were from left-to-right), as in Figure 4, so that errors parallel and orthogonal to the direction of the saccade are represented by respective components of the 2D vectors. Because the aim here was to assess the spatial aspects of the step, the lag of the decoded signal behind the eye (45 ms) was removed (by shifting the data in time by that amount) before calculating the residual error between the V1 representation and the real eye. Similarly, we removed the slight compression of decoded eye positions seen during fixation, visible in Figure 3, and in Figure 4B as an under-estimation of eccentricity. We did so by scaling the decoded data by the ratio of step amplitudes from the sigmoid fits.