Inferring visual space from ultra-fine extra-retinal knowledge of gaze position (preprint)

Zhetuo Zhao[1,2], Ehud Ahissar[3], Jonathan D. Victor[4], Michele Rucci[1,2⋆]

[1]Department of Brain and Cognitive Sciences and [2]Center for Visual Science, University of Rochester, NY, USA. [3]Department of Brain Sciences, Weizmann Institute of Science, Rehovot, Israel. [4]Feil Family Brain and Mind Research Institute, Weill Cornell Medical College, New York, NY, USA. ⋆ Corresponding author (mrucci@ur.rochester.edu).

Abstract. It has long been debated how humans resolve fine details and perceive a stable visual world despite the incessant fixational motion of their eyes. Current theories assume these processes to rely solely on the visual input to the retina, without contributions from motor and/or proprioceptive sources. Here we show that contrary to this widespread assumption, the visual system has access to high-resolution extra-retinal knowledge of fixational eye motion and uses it to deduce spatial relations. Building on recent advances in gaze-contingent display control, we created a spatial discrimination task in which the stimulus configuration was entirely determined by oculomotor activity. Our results show that humans correctly infer geometrical relations in the absence of spatial information on the retina and accurately combine high-resolution extraretinal monitoring of gaze displacement with retinal signals. These findings reveal a sensory-motor strategy for encoding space, in which fine oculomotor knowledge is used to interpret the fixational input to the retina.

Introduction

Our eyes are never at rest. Since fine visual resolution is restricted to a tiny portion of the retina, the fovea, humans use eye movements to inspect objects of interest. Remarkably, the eyes remain in motion even in the intervals between voluntary gaze shifts, the so-called "fixation" periods in which visual information is acquired and processed. In these periods, a persistent eye jitter, known as ocular drift, continually perturbs the direction of gaze, moving the projection of the stimulus on the retina across dozens of receptors.

Given the extent of ocular drift and the temporal responses of retinal neurons, it has long been questioned how the visual system manages to avoid perceptual blurring during fixation and establish stable high-acuity representations[1–3]. Multiple theories have been proposed. Some regard the fixational motion of the eye as a challenge to be overcome through specific decoding strategies[4]. Others argue that eye movements are beneficial for processing spatial information, either by transforming spatial patterns into temporal modulations[5–7] or by following spatial registration strategies similar to those used in computer vision to enhance image resolution[8]. Although the proposed theories differ widely in their specific mechanisms, they all share the common assumption that spatial representations at fixation are established solely based on the visual input signals impinging onto the retina, without making use of information from other sources.

This standard assumption, however, contrasts with the multimodal and sensorimotor integration that is known to occur in the presence of larger eye movements, such as the rapid gaze shifts (saccades) and tracking movements (smooth pursuits) that bring and maintain objects onto the fovea. With these movements, interpretation of retinal activity critically depends on motor and proprioceptive knowledge about how the eyes move. Extraretinal signals are known to modulate visual responses by both enhancing and attenuating sensitivity, often in a dynamic manner at specific times during the movements[9–17]. Extraretinal modulations are deemed to be essential for extracting information from the retinal flow[18,19], establishing spatial representations[20–24], and discarding the motion of the retinal image caused by the eye movements themselves[25].

Various factors have contributed to the current tenet that a similar visuomotor integration does not take place during fixation. From a historical perspective, vision science has traditionally approximated the fixational input to the retina as an image, neglecting the incessant motion of the eye and/or assuming this motion to be too small to yield reliable motor or proprioceptive signal. The eyes appear to wander erratically during fixation, leading many researchers to conclude that ocular drift stems from limits in oculomotor control[26,27] and is, thus, unlikely to be monitored. Reinforcing this idea, previous attempts to identify extraretinal signals associated with fixational drifts reported negative results[28,29], and several studies have argued that retinal signals are solely responsible for establishing stable visual representations during fixation (*e.g.*, [30]).

However, contrary to the mainstream assumption, it has long been proposed that ocular drift may actually represent a form of slow control aimed at delivering a desired amount of retinal image motion[31,32]. This proposal has received renewed support from recent findings, including the observation that drift partly counteracts the physiological instability of the head[33], as well as task- and stimulus-dependent changes in drift characteristics[34–36]. Furthermore, previous studies that searched for motor knowledge of fixational drifts either did not control for the spatial information delivered to the retina[28] or focused on relatively long temporal windows, intervals over which memory decays could have played a role (*e.g.*, [29]). These considerations raise the need for more specific investigations on the mechanisms by which stable high-resolution spatial representations are established during the incessant fixational motion of the eye.

Here we built upon recent advances on high-resolution eye-tracking and gaze-contingent display control, the capability to modify the stimulus in real-time according to the observer's eye movements, to precisely control retinal stimulation. We developed a spatial discrimination task that cannot be accomplished solely based on the visual input signals to the retina, but rather, depended critically on knowledge of eye position. We show that despite the lack of spatial information in the retinal input, the visual system is capable of reconstructing the configuration of the stimulus, and therefore estimating the fixational motion of the eye, with exquisite sensitivity. These results show that humans possess fine motor knowledge of the way the eye drifts during fixation and integrate this information into high-resolution spatial representations.

Results

We developed a task that requires motor knowledge of the direction in which the eye moves to be successfully executed. In a forced-choice task, subjects discriminated the spatial configuration of a stimulus that entirely depended on their performed eye movements. They reported whether the bottom bar of a Vernier appeared to be to the right or left of the top bar (Fig. 1A), but, unlike a conventional spatial judgment, the two bars of the Vernier were never visible simultaneously, and no information about their spatial offset was ever delivered to the retina. This was achieved via a gaze-contingent procedure that rendered the stimulus on the display as seen through a retinally-stabilized aperture, a thin slit that moved under real-time computer control together with the eye to restrict stimulation to a narrow vertical strip on the retina centered on the fovea (Fig. 1B). In this way, as the normal fixational motion of the eye swept the aperture across the stimulus, the two bars appeared sequentially at vertically aligned positions on the retina (Fig. 1C), yielding input signals that—under ideal conditions—are not informative for the task (Fig. 1D).

In practice, unbeknownst to the observer, the two bars were displayed one below and one above the position of the center of gaze at two separate times (T_1 and T_2 in Fig. 1E), the first at a random time from the beginning of the trial, and the second after a fixed delay from the disappearance of the first bar (the inter-stimulus interval, ISI). Subjects moved their eyes normally under these conditions while attempting to maintain fixation at the remembered location of a marker (a $5'$ dot) briefly presented at the beginning of each trial. They alternated occasional small saccades with periods of ocular drift, which moved the eye in its stereotypical, seemingly erratic fashion with characteristics similar to those measured from the same observers when maintaining fixation on a visible marker (Fig. 1F). In this study we specifically focused on fixational drifts and discarded all trials in which subjects performed saccades or microsaccades. With an ISI of 100 ms, ocular drift resulted in displacements of the line of sight distributed around $\pm 4'$ (Fig. 1G).

Remarkably, subjects were highly proficient in reporting the stimulus configuration, even though its spatial layout was never made explicit on the retina (Fig. 1H,I). Their qualitative experience consisted of two successive flashes with a clear spatial offset. Performance was significantly above chance already at the smallest Vernier offset that could be presented, a gap of only $1.4'$ corresponding to the spacing of just one single pixel on the display. Performance further increased with larger gaps, with a two-fold increment in d' as the Vernier offset increased to $2.8'$. These results were highly consistent across individuals: all subjects were able to successfully accomplish the task. In each individual observer, performance was significantly above chance at all Vernier gaps (p <0.021, one-tailed bootstrap test), with the exception of one subject at the smallest offset ($1.4'$) for which the d' was close to significance (p =0.065).

These results were not caused by possible biases—and thereof knowledge—in the individual direction of eye movements, *i.e.,* the realization that perhaps drift was more pronounced in one direction. No obvious directional biases were observed in the recorded data, and horizontal displacements in the two directions were approximately symmetrically distributed (Fig. S1A). Furthermore, performance was high in both the trials in which the eye drifted to the left and to the right (Fig. S1B,C), indicating knowledge of the specific direction of ocular drift in each individual trial. Thus, these data suggest that the visual system has access to high-resolution extraretinal information of how the eye moves during fixation.

[Figure 1 about here]

Given these unexpected results, we wondered whether our methods of visual stimulation inadvertently introduced spurious spatial cues. Meticulous care had been taken to eliminate all obvious retinal cues that could inform about the stimulus configuration. This included conducting the experiments in complete darkness—while preventing dark adaptation with brief light exposure between block of trials—to avoid visual references; using a fast-phosphor high-speed CRT display to minimize persistence; and lowering the monitor intensity to minimum settings to ensure that the edges of the monitor were not visible. We questioned, however, whether more subtle cues, such as the baseline luminance of the CRT display or possible residual phosphor persistence, played a role by providing unwanted visual references. We also wondered whether the aperture had provided some type of motion signal that could inform about the drift direction. For all these reasons, we repeated the experiment using a custom-built display, an array of 110×8 LEDs specifically selected to provide no persistence and no baseline luminance (Fig. 2A). We also made sure to rule out any possible motion signal by exposing each Vernier bar only for a brief interval (5 ms), the shortest detectable exposure allowed by our display.

Comparison between the data in Fig. 2B, C and Fig. 1H, I show that results were little affected by these changes in visual stimulation. The drift behavior changed little from the previous experiment and remained practically identical to that observed during fixation on a visible marker (Fig. S2). Critically, subjects continued to correctly report the stimulus even under these more stringent conditions: performance was already above chance at the smallest possible Vernier offset (in this case $1.9'$, the width of one LED) and further improved as the distance between the two bars increased. These effects were clearly visible in the data from each observer, all of whom individually exhibited above chance performance at all Vernier offsets presented (p <0.011, bootstrap test).

These findings were very robust. As in the experiment of Fig. 1, performance was similar for leftward and rightward ocular drift (Fig. S1D-F), showing access to the specific drift trajectory performed in each trial, rather than knowledge of possible directional biases. Results were also not caused by possible inaccuracies in measuring eye movements. In this regard, it is important to notice that the experiments relied on the relative alignment of the two bars on the retina, not their absolute positions. That is, conclusions do not depend on the accuracy of gaze localization—a notoriously difficult operation—but on the capability to measure changes in gaze position, something that a properly tuned and calibrated DPI eye-tracker accomplishes with sub-arcminute resolution[37]. Monte Carlo simulations show that eye movements would need to be over-estimated by an unrealistic amount, over 100%, to account for our findings (Supplementary Fig. S3). This degree of imprecision is not plausible with our recording apparatus.

Furthermore, analysis of residual errors in the alignment of retinal stimuli revealed that these cues cannot account for the experimental data. To be perfectly aligned on the retina, each bar needs to be rendered exactly at the current location of gaze. In practice, however, the precision of this operation is limited by the resolution of the display, as the stimulus can only be shown at the closest pixel/LED location, resulting in a small offset (X_R in Fig. 2A). This misalignment did exert a perceptual influence. For each Vernier offset X on the display (diagonal lines in Fig. 2D), perceptual reports exhibited a subtle but systematic influence from X_R: the probability of reporting the bottom bar to the right was slightly larger when the misalignment was consistent with this interpretation ($X_R > 0$) than when it was in the opposite direction ($X_R < 0$; diagonal arrow in Fig. 2D). However, this cue could not possibly account for the general pattern of results obtained as X varied. Its influence was small relative to that exerted by the gaze displacement X_E (horizontal arrow in Fig. 2D), and overall, perceptual reports were driven by X_E irrespective of X_R (Fig. 2E). In fact, X_R was overall poorly correlated with subject responses (average correlation coefficient across observers: $\rho = -0.016 \pm 0.093$), and subjects were able to successfully accomplish the task even in the trials in which the misalignment indicated the wrong response, the trials in which the X_R was in the opposite direction of the Vernier offset on the display (Fig. 2F). All these analyses further support the conclusion that humans incorporate fine oculomotor knowledge in the establishment of spatial representations.

[Figure 2 about here]

The small stimulus offsets caused by the display resolution provide an opportunity to examine how the visual system integrates retinal and extraretinal signals at fixation. To gain insight into this process, we compared the perceptual reports recorded in the experiments to the responses of an ideal observer that inferred the most likely configuration of the stimulus from sensory measurements of both the eye displacement and the retinal misalignment. The ideal observer assumes uncertainty in sensory signals (modeled as additive Gaussian noise) and possesses only general knowledge about eye movements. Specifically, it assumes that ocular drift evolves as Brownian motion and, therefore, the variance of the probability of gaze displacement increases proportionally to time[38,39]. For each individual observer, the diffusion constant of this motion was directly estimated from their eye movements. In each trial, the model weighted the measured probability of eye displacement by its prior and estimated the most likely configuration of the stimulus ("Left" or "Right") by comparing the overall probability (the integral of the 2D posterior probability distribution) on the two sides of the zero displacement line (diagonal cyan line in Fig. 3A).

The ideal observer closely replicated the way subject's responses varied as a function of X_E and X_R (*cf.* Fig. 3B and Fig. 2D). As in the empirical data, the overall pattern of response was primarily driven by the eye displacement, but a dependence on retinal misalignment was also visible for each gap of the Vernier on the display. Across all data points, the model accounted for 93% of the variance in subject's responses (green dots in Fig. 3E) and accurately predicted the d′ observed in the experiments (green line Fig. 3F; individual data in Fig. S4A). Critically, both motor information about drift displacement and retinal information about bars alignment were necessary to replicate experimental data. Discarding the retinal signal led to a reduction in performance, but the model was still able to account for about 56% of the variance in perceptual reports. In contrast, performance dropped to chance level and the model could only account for 12% of the variance following elimination of extraretinal information (Fig. 3E,F; see also log-likelihood data in Fig. S4A). Thus, subjects performed very similarly to the predictions of a Bayesian combination of retinal and extraretinal sensory signals, with a predominant influence exerted by motor knowledge of eye movements.

[Figure 3 about here]

The previous results indicate that motor knowledge of eye drifts during fixation is incorporated into spatial judgements. To gain insight into the mechanisms responsible for monitoring gaze position at this level of resolution, we examined the temporal course of this process. Specifically, as illustrated in Fig. 4*A*, we searched for the interval W over which the gaze displacement ΔX best correlated with the subject's responses. To this end, we systematically varied both the duration of the window of observation (T_W) and its timing (Δt_W) measured as the lag between the onset of the first bar and the window center.

Fig. 4*B* shows the average correlation between gaze shifts and perceptual reports as a function of both position (horizontal axis) and duration (vertical axis) of the window of observation. As shown by these data, the correlation peaked for a short window of approximately 100 ms that slightly preceded the bar presentations. This finding was highly consistent across individual observers, all of whom exhibited a similar timing, resulting in a statistically significant anticipation of the onset of the window of observation relative to the onset of the first Vernier bar (-13 ms on average; Fig. 4*C*). Thus, during fixation, retinal signals appear to be combined with motor estimation of gaze position that slightly precedes retinal exposures, suggesting a predictive use of extraretinal signals.

Given that the duration of the optimal window in Fig. 4 was similar to the interval between bar exposures (100 ms), we wondered whether this window indicates continuous oculomotor monitoring throughout the ISI or represents a fixed internal temporal scale over which drift is estimated. Both possibilities can be mediated by a mechanism of integration of noisy velocity signals, a process similar to the one believed to occur for smooth pursuit[40–42]. However, these two hypotheses lead to distinct predictions as the ISI is further increased. If drift displacement is integrated across the entire interval between bar exposures, we would expect the uncertainty in the extraretinal measurement of displacement (*i.e.*, its standard deviation σ_E) to increase no faster than $\sqrt{t}$, as accumulation of temporally uncorrelated noise progressively disrupts the position estimate. In contrast, if gaze displacement is estimated from the movement measured over a shorter interval of the ISI, we would expect σ_E to increase proportionally to t, as the consequence of a temporal extrapolation process.

To address this question, we repeated the experiment while increasing the ISI between the bar exposures by a factor of five, to 500 ms. Except for this longer interval, the paradigm was otherwise identical to that of Fig. 2, with 5 ms exposures delivered by our custom LED display. Increasing the ISI profoundly affected performance. Proportion of correct responses at the smallest Vernier offsets dropped drastically and were at now chance level with a $1.9'$ gap. Furthermore, even at the much larger Vernier offsets resulting from the longer ISI, performance remained considerably lower than the levels measured in the 100 ms ISI condition (Fig. 4E). Results were highly consistent across subjects, all of whom exhibited substantial and significant reductions in performance in the 500 ms condition (Fig. 4F).

These data closely matched the predictions of our ideal observer model under the "extrapolation" hypothesis, *i.e.*, when the uncertainty in the extraretinal measurement for the 500-ms ISI was assumed to be five times larger than for the 100-ms ISI ($\sigma_E(500) = 5\ \sigma_E(100)$; red curve in Fig. 4E). In contrast, model predictions fell far from the data under the "integration" hypothesis, *i.e.*, when extraretinal uncertainty was increased proportionally to the squared root of time ($\sigma_E(500) = \sqrt{5}\ \sigma_E(100)$; yellow curve in Fig. 4E). In this case, the model significantly overestimated performance in all observers (Fig. S5). In keeping with these data, the duration of the temporal window over which gaze displacement best correlated with perceptual reports did not increase, but remained similar to that observed for the shorter ISI (Fig. 4D).

These findings are not compatible with continuous monitoring of eye position throughout the ISI. They suggest that extraretinal estimation of ocular drift is conducted over a short window of approximately 100 ms duration. When asked to estimate ocular displacement over a longer interval, subjects recur to a process of extrapolation, presumably because this is the only possible strategy under our experimental conditions.

[Figure 4 about here]

Discussion

The eyes drift incessantly in the intervals between saccades, even when attending to a single point, raising fundamental questions on how the visual system avoids perceptual blurring, resolves fine detail, and establishes stable high-acuity spatial representations. Existing theories assume these processes to rely exclusively on the output signals from the retina[4,8,43]. Contrary to this idea, our results show that the human visual system has access to high-resolution motor knowledge about eye movements and integrate this information with signals from the retina to estimate fine spatial relations. These findings challenge the standard view of passive processing of a retinal image during fixation and indicate that the computations responsible for representing visual space are intrinsically sensorimotor.

To unveil extraretinal contributions, our study relied on methods for gaze-contingent display control, the updating of the stimulus according to the observer's eye movements. This approach enables both precise control of visual input signals and manipulation of visuomotor contingencies. Specifically, in our experiments, we tailored the visual input to create a stimulus configuration on the display that conveyed no spatial information on the retina. Our data show that even under these stringent conditions, humans retain sufficient knowledge of their oculomotor activity to reconstruct the direction in which gaze drifts over a short interval. This knowledge is specific, enabling detection of gaze displacements with arcminute resolution. Furthermore, this oculomotor signal is evaluated in the light of general knowledge of eye drift statistics, so that spatial judgements closely follow the predictions of an ideal observer that assigns uncertainty to the estimated spatial representation (the addition of retinal and extraretinal information) based on the reliability of the extraretinal signal.

Our results also indicate that the extraretinal signal is continually estimated over a short temporal interval of approximately 100 ms. This interval systematically precedes visual stimulation by more than 10 ms, likely yielding an even larger lag once the response delays of visual neurons are taken into account. The duration of the window of integration appears inflexible, forcing subjects in our experiment with a long ISI to infer gaze displacement via extrapolation (Fig. 4E). This strategy likely represents an adaptation to the unnatural conditions of our experiments, where visual information is not continuously present. It is remarkable that the duration of this window approximately matches the interval of integration of neurons in the early stages of the visual system[44], a matching that may have important computational consequences. Note, however, that a persistent trace of visual stimulation may not be necessary for the extraretinal signal to exert its effect, as suggested by the above chance performance measured with the 500 ms ISI (Fig. 4E).

It is important to emphasize that our findings cannot plausibly be explained by inaccurate positioning of stimuli on the retina nor inaccuracies in eye-tracking. Extensive care was taken to eliminate all informative spatial cues in the retinal input and ensure that results were not contaminated by corollary discharges associated to saccades, microsaccades, or other types of smooth eye movements. Our analyses confirm that our apparatus is highly precise, as we can reliably measure the perceptual consequence of the stimulus misalignment caused from display resolution, the tiny mismatch between the measured gaze position and the actual position of the stimulus on the display. While this retinal cue exerts a clear influence at every presented Vernier offset, it cannot account for perceptual reports, as performance in the task varied primarily with the measured gaze displacement. Modeling of eye-tracking errors showed that gaze displacement would need to be over-estimated by an unrealistic amount to yield retinal cues that could account for our results.

Furthermore, the stimulus duration was too short—more than one order of magnitude[45,46]—to provide useful motion signals. Our custom display was designed to switch on/off within tens of microseconds (Fig. 2A), and the median displacement of the stimulus on the retina during the resulting brief exposures was only $\sim$14 arcseconds, well below the thresholds reported in the literature for similar tasks[47]. Results did not change when selecting only the trials with minimal displacement during exposure, and the instantaneous velocities measured around the times of bar exposures were only weakly correlated with perceptual reports (Fig. 4B). All these observations indicate that retinal image motion played no role in our experiments.

At first sight, the finding that eye drift is monitored appears to contradict widespread assumptions in the field. An obvious conflict is with the notion that drift is not controlled—the popular idea that this motion results from noise at the neural and/or muscular level[26,48]. Although less known, however, it has long been proposed that the smooth fixational motion of the eye actually represents a form of slow control, a sort of pursuit of a stationary target aimed at maintaining ideal visual conditions[31,32,49] and eliciting neural responses[10,50]. This view has received strong support in the recent literature. It is now known that during natural fixation, when the head is free to move normally, ocular drift partially compensates for the physiological instability of the head, severely constraining retinal image motion[33,51]. Furthermore, changes in the characteristics of ocular drift have been observed in high-acuity tasks, as when looking at a 20/20 line of an eye-chart or when judging the expression of a distant face[34,52]. These changes appear to be functional, as they increase the power of the luminance modulations impinging onto retinal receptors, an effect consistent with theories arguing for temporal representations of fine spatial details[6,7]. The present study goes beyond this previous body of work by showing that the signals involved in exerting control at this scale also contribute extraretinal information that is integrated in spatial representations.

Our conclusions also appear to contrast with those reached by previous studies with similar paradigms. Classical experiments with asynchronously displayed Verniers concluded that drift is not monitored because performance declines with increasing delays between exposures[28,53]. Fig. 4 replicates this effect, but our data show that other factors (*e.g.*, memory decays and/or the window over which drift is monitored) must be responsible for the measured decrement in performance. More recently, support to the notion that fixational drift is not monitored has come from systematic localization errors observed with stimuli briefly displayed in complete darkness. When reporting the position of a previously displayed reference by selecting between two probes, one at the same reference's location on the display (spatiotopic probe) and one at its same position on the retina (retinotopic probe), subjects systematically select the retinotopic one[29]. These errors are, in fact, predicted by our ideal observer model, but attributed to the specific perceptual choice presented to the observer rather than lack of extraretinal knowledge of eye drift (see Fig. S7). Thus, the present study suggests alternative explanations for the previous reports in the literature.

Our findings lead to a critical question: why are eye movements monitored at such high level of resolution? There are several complementary ways in which an extraretinal drift signal could contribute to visual processing. A possibility is by facilitating visual stability during fixation, *i.e.*, by helping disentangling the visual motion signals resulting from external objects from those generated by eye movements. Studies on how the visual system discards motion signals resulting from egomotion have primarily focused on larger eye movements, saccades and smooth pursuits[23–25,54,55], often in the context of the establishment of spatiotopic representations[56–58]. These studies have emphasized the interaction between retinal and extraretinal signals, both efference copies of motor commands[59] and proprioceptive information from extraocular muscles[60]. The eye drift that occurs during fixation is commonly assumed to be too small for extraretinal compensation, and early suggestions that the receptive fields of neurons in the primary visual cortex counteract this motion[61] were not supported by later experimental measurements[62]. Thus, the resulting visual motion signals are believed to be perceptually canceled solely on the basis of the retinal input[29,30].

This idea, however, is at odds with the motion perceived during exposure to retinally-stabilized objects, stimuli that move with the eye to remain immobile on the retina[29,63].

Furthermore, it has been observed that motion perception is biased to the direction of eye movements, so that stimuli that move opposite to ocular drift on the retina tend to appear stable even if their motion is amplified[63,64]. Such bias requires knowledge of drift direction, information that could be provided by the extraretinal signal uncovered in our experiments. There are several ways in which extraretinal knowledge of ocular drift can help improving perceptual stability. One possibility is that the visual system estimates drift motion as the lowest instantaneous velocity on the retina that is also congruent with drift direction, a view that would explain not only the perceived motion with stabilized images and the directional anisotropy in motion perception, but also the perceived jittery motion of a stationary stimulus following adaptation to dynamic noise patterns[30]. This approach is similar to the one proposed to explain jitter after-effects, but differs from a purely retinal cancellation mechanism for also requiring directional consistency with extraretinal measurements.

Our findings also suggest another way in which extraretinal drift information could contribute to visual perception, which is by directly participating in the establishment of high-acuity spatial representations. In our experiments, observers were able to infer geometrical arrangements purely based on extraretinal information. Until now, spatial information during fixation has been assumed to be extracted solely from the responses of retinal neurons[4,8,43,65]. While several methods have been proposed for registering afferent visual information into spatial maps as the eye drifts, all these methods exclusively rely on the retinal input. However, this process presumably depends on the richness of visual stimulation and requires temporal accumulation of evidence, difficulties that an extraretinal drift signal could alleviate. Thus, motor knowledge of ocular drift may be particularly valuable when visual stimulation is sparse and following saccades, when new visual content is introduced on the retina. Interestingly, an extraretinal contribution makes this process similar to the coordinate transformation underlying the establishment of head-centered spatial representations during larger eye movements[56,66–69], emphasizing a general computational strategy and supporting a similarity between fixational drift and pursuit movements[32]. Further work is needed to assess the origins of the extraretinal signal unveiled by our experiments and its specific role in representing space.

Methods

Subjects

A total of 13 subjects (5 males and 8 females; age range: 20-35), all naïve about the purpose of the study, participated in the experiments. All subjects were emmetropic, with at least 20/20 visual acuity in the right eye as measured by a Snellen eye chart, and were compensated for their participation. Informed consent was obtained from all participants following the procedures approved by Institutional Review Boards at Boston University and the University of Rochester.

Stimuli

Stimuli consisted of standard Verniers, with two vertical bars separated by a horizontal gap (the Vernier offset; Fig. 1A). The two bars were never simultaneously visible: they were exposed at different times at the current location of the line of sight on the display, so that the offset was determined by the gaze displacement that occurred in between the two exposures. In this way, the bars always appeared vertically aligned on the retina, whereas the gap on the display varied across trials based on the eye movements performed by the observer.

In the experiment of Fig. 1 (Experiment 1), each bar was $28'$ long and $1.4'$ wide and exposed at luminance of 14.2 cd/m^2. Bars were $19' \times 1.9'$ and possessed luminance of 49.6 cd/m^2 in the experiments of Figs. 2 and 4 (Experiments 2 and 3). These dimensions were the outcome of adjusting the distance of the display so that each bar could be as thin as possible (one pixel wide in Experiment 1 and one LED wide in Experiments 2-3), while at the same time retaining clear visibility when briefly exposed at maximum intensity. Stimuli were examined in total darkness, carefully removing all light sources that could serve as potential spatial references and all visual cues that could provide information about the Vernier configuration.

Apparatus

Stimuli were rendered by means of EyeRIS, a hardware/software system for real-time gaze-contingent display that enables precise synchronization between eye movement data and the refresh of the display[70]. They were viewed monocularly with the right eye, while the left eye was patched. A dental imprint bite-bar and a headrest minimized head movements and maintained the observer at a fixed distance from the monitor.

Different displays were used in Experiment 1 and in Experiments 2-3. In Experiment 1, stimuli were rendered on a fast-phosphor CRT monitor (Iiyama HM204DT) at a resolution of 800 × 600 pixels and 200 Hz refresh rate. This monitor has fast phosphors with decay time shorter than 2 ms. A completely dark background and tuning of the monitor at minimum settings ensured that the edges of the display were never visible.

To further control for possible influences from phosphor persistence, residual background luminance, and retinal image motion, in Experiments 2-3 stimuli were displayed on a custom LED display specifically developed for this study (Fig. 2A). LEDs are not affected by lingering activity like phosphors and have zero baseline illumination when not active. The custom display consisted of 880 LEDs, 800 rectangular elements arranged into two rows of 4 LED, and a 3 × 3 array of circular LED used for eye-tracking calibration. Each Vernier bar was given by the simultaneous activation of a column of 4 LED in either the top or the bottom row. This display also offered lower latency relative to a CRT (3 ms vs. 7.5 ms, on average) and more precise timing, since each LED could be controlled independently without having to wait for the rasterization of a frame to be completed, as in a CRT. LED activation triggered a digital signal that was sampled synchronously with oculomotor data, so that the timing of stimulus presentation could be reconstructed offline with high precision.

To measure eye movements with the precision necessary to align stimuli on the retina, we used a dual Purkinje Image (DPI) eye-tracker (Fourward Technology), an analog system with high spatiotemporal resolution and minimal delay. This specific eye-tracker has been customized over the course of two decades to refine its dynamics and minimize sources of noise. It resolves movements smaller than $1'$ as tested with an artificial eye controlled by a galvanometer. Analog eye movements data were first low-pass filtered at 500 Hz, then sampled at 1 kHz, and recorded for off-line analysis. Note that since the DPI directly estimates gaze position, measurement errors do not accumulate over time. That is, measurements of similar displacements estimated over different ISIs, like in Fig. 4E, are expected to possess similar accuracy.

Experimental procedures

Data were collected in multiple experimental sessions, each lasting approximately 1 hour. Each session consisted of several blocks of trials, with each block containing approximately 50 trials. Every block started with preparatory procedures to ensure optimal eye-tracking.

These steps included positioning the subject in the apparatus, tuning the eye-tracker, and performing calibration procedures to accurately localize gaze. Frequent breaks between blocks allowed the subject to rest. Lights were turned on during these breaks to prevent dark adaptation and minimize visibility of the edges of the CRT as well as the influence of any possible residual light.

Subjects were told that the two bars of a Vernier would be presented sequentially in random order and were asked to report whether the bottom bar was to the left/right of to the top bar by pressing a corresponding button on a joypad. Each trial started with the subject fixating on a $5'$ red dot at the center of the display for 1 s. The fixation marker was then turned off, and after a uniformly-distributed random delay of 1-2 s, the first Vernier bar was exposed either above or below the current gaze position (equal probability across trials). The second bar then followed with a fixed delay (the inter-stimulus interval, ISI) at the current gaze location. In this way, the two Vernier bars were aligned on the retina and separated on the display by the gaze shift that occurred during the ISI (both horizontal and vertical displacements in Experiment 1; only horizontal displacement in Experiments 2 and 3). The ISI was 100 ms in Experiments 1 and 2 and 500 ms in Experiment 3.

Slightly different procedures were adopted in Experiments 1 and 2-3. In Experiment 1, the image was continually updated on the CRT display to replicate the visual consequences of viewing stimuli through a thin slit aperture that moved with gaze (*i.e.*, a retinally-stabilized aperture; Fig. 1C,D). This implied that the stimulus exposure varied across trials, as each bar remained visible as long as it was aligned with the aperture. One bar was displayed in the top half of the aperture, and one in the bottom half. In Experiments 2 and 3, to eliminate possible motion signals, each Vernier bar was only displayed for 5 ms, the shortest exposure at the maximum intensity afforded by our LED display. In every trial, two columns of LED were activated, one in the top and one in the bottom row of the display. Columns were selected as the ones closest to the horizontal gaze position measured at the time of exposure. Except for these points, the paradigm was otherwise identical in the two experiments.

Data analysis

Oculomotor data. Periods of blinks and poor tracking were automatically detected by the DPI eye-tracker. Only trials with optimal, uninterrupted eye-tracking and no blinks were selected for data analysis. Recorded oculomotor traces were first automatically segmented into separate periods of drift and saccades based on speed threshold of $3°/s$ and validated by human experts. Segmentation based on eye speed is very accurate with the high-quality data provided by the DPI during head immobilization. In this study, we specifically focus on ocular drift. All trials that contained other types of eye movements besides ocular drift, like saccades and microsaccades, were excluded from data analysis.

Evaluation of performance. At every Vernier offset, performance was quantified by means of both proportion correct and d′. For each individual observer, we used bootstrap to evaluate statistical significance across conditions and differences from chance levels (Figs. S4 and S5). The data reported in Figs. 1-4 are averages across observers and corresponding statistics.

In Fig. 2, performance is also examined as a function of both the horizontal eye displacement (X_E) and the estimated misalignment of the two bars on the retina ($X_R < 2'$). Ideally the two Vernier bars need to be perfectly aligned. However, each bar could only be displayed at the pixel/LED closest to the estimated gaze position, so that the Vernier offset X on the display was not X_E, but equal to $X_E + X_R$. We assessed the joint influence of X_E and X_R by binning trials according to their values to uniformly sample the space and examined how perceptual reports varied across bins. In the space (X_E, X_R), a Vernier offset X on the display corresponds to a -45° tilted line, as the same X could be reached with various cue combinations ($X_E + X_R = X$). The 5 lines in Fig. 2D corresponds to the 5 Vernier offsets reached in the experiment ($0'$, $\pm 1.9'$, $\pm 3.8'$). The data in Fig. 2D represent averages obtained by pooling data across subjects, so that each bin contained on average 60 trials.

In Fig. 4B, D, the correlation between gaze displacement and perceptual reports was examined as a function of both lag Δt_W and duration T_W of the temporal window of observation. To this end, we first converted subject's responses into a binary format (-1 and 1) and then computed the Pearson correlation coefficient with the horizontal displacement in the interval $[\Delta t_W - \frac{T_W}{2}, \Delta t_W + \frac{T_W}{2}]$. Highly similar results were obtained over sets of trials collected early or late in the experiments, suggesting little influence from training (Fig. S6).

Ideal observer model. To gain insight into the mechanisms by which extraretinal estimation of ocular drift contributes to representing space, we compared the perceptual reports measured in the experiments to those of an ideal observer that adds noisy sensory measurements of spatial cues on the retina and eye movements (X_R and X_E) to establish head-centered representations. The ideal observer assumes ocular drift to resemble Brownian motion with a specific diffusion rate. This assumption is incorporated in the joint prior distribution $p(X_R, X_E)$, which is uniform along x_R and follows a Gaussian distribution with zero mean and standard deviation $\sqrt{2DT}$ along x_E, where D is the diffusion coefficient of the individual's drift process and T the ISI. In a Brownian process the variance evolves proportionally to time. For each subject, we estimated D from the recorded eye traces via linear regression of the variance of the gaze displacement over the considered ISI. Sensory measurements of X_E and X_R were assumed to be corrupted by independent additive white noise processes with Gaussian distributions: $p(x_R|X_R) = N(X_R, \sigma_R)$ and $p(x_E|X_E) = N(X_E, \sigma_E)$.

In every trial, the ideal observer estimates the joint posterior probability of the retinal and extraretinal displacement:

$$p\left(\widehat{X}_E, \widehat{X}_R\right) = p(x_E, x_R|X_E, X_R)\ \ p(X_E, X_R) = p(x_E|X_E)\ \ p(x_R|X_R)\ \ p(X_E, X_R). \tag{1}$$

Thus, $p\left(\widehat{X}_E, \widehat{X}_R\right)$ is a two-dimensional Gaussian with mean and covariance given by:

$$\boldsymbol{\mu} = \begin{bmatrix} \frac{2DT}{\sigma_E^2+2DT} X_E \\ X_R \end{bmatrix},\ \Sigma = \begin{bmatrix} \frac{2DT\sigma_E^2}{\sigma_E^2+2DT} & 0 \\ 0 & \sigma_R^2 \end{bmatrix}, \tag{2}$$

The probability of any given Vernier offset X, $p\left(\widehat{X}\right)$, can then be estimated by integrating the joint posterior probability $p\left(\hat{X}_E, \hat{X}_R\right)$ along the line $X_E + X_R = X$ (Fig. 3A):

$$p\left(\hat{X}\right) \sim N\left(X_R + \frac{2DT}{\sigma_E^2 + 2DT} X_E, \sqrt{\sigma_R^2 + \frac{2DT\sigma_E^2}{\sigma_E^2 + 2DT}}\right) \tag{3}$$

The probabilities of reporting the bottom bar of the Vernier to the left or to the right of the top bar are then given by $\mathrm{P}\left(\widehat{X} < 0\right)$ and $\mathrm{P}\left(\widehat{X} > 0\right)$, respectively.

The two free parameters of the model, σ_E and σ_R, determine the uncertainty of sensory measurements. The larger is σ_E, the weaker is its perceptual influence, with no trial-specific extraretinal knowledge of ocular drift in the limit case of $\sigma_E = \infty$. These parameters were estimated individually for each subject to maximize the log likelihood ($L = \sum_i \log P_i$) of the model replicating the subject's perceptual reports across all trials:

$$(\sigma_E, \sigma_R) = \arg\max_{\sigma_E, \sigma_R} L \tag{4}$$

where P_i represents the probability that the model responds in the same way as the observer in trial i: $P_i = \mathrm{P}\left(\widehat{X} < 0\right)$ if the subject responded ”Left” and $P_i = \mathrm{P}\left(\widehat{X} > 0\right)$ if he/she responded ”Right”.

Evaluation of model performance. We evaluated the model in several ways. The data in Fig. 3F compare the overall performance measured in the experiments to that predicted by the model. Predictions were first obtained for each individual observer (Fig. S4*A*) and then averaged across subjects in Fig. 3. The log-likelihood L by which the model accounts for subject's perceptual responses is reported in Fig. S4*A*. We also examined the model's capability to reproduce the pattern of perceptual responses as a function of the measured retinal and extraretinal cues. Fig. 3*B* compares the output of the model to perceptual reports for each of the groups of trials of Fig. 2*D*. The overall accuracy of the model is summarized by the coefficient of determination R^2 in Fig. 3*E*.

Furthermore, in Fig. 3 we compared both performance and perceptual reports to an ideal observer that operates on just one of the two cues, either X_E or X_R. In this case, parameters were optimized with the model reduced to estimating the Vernier offset from the marginal posterior probability along the considered axis:

$$p\left(\hat{X}\right) = p\left(\hat{X}_E\right) \sim N\left(\frac{2DT}{\sigma_E^2 + 2DT} X_E, \sqrt{\frac{2DT\sigma_E^2}{\sigma_E^2 + 2DT}}\right) \tag{5}$$

or

$$p\left(\hat{X}\right) = p\left(\hat{X}_R\right) \sim N(X_R, \sigma_R) \tag{6}$$

where parameters were obtained via the same maximum likelihood procedure used for the full model.

Dynamics of drift estimation. Distinct predictions emerge if gaze displacement is estimated over the entire interval between bar exposures or by extrapolating measurements obtained over a shorter interval. In the former case, the error in estimating gaze displacement will progressively accumulate because of the noise in the measurement. Specifically, the standard deviation of the estimate will grow as $\sigma_E \propto \sqrt{t}$ under the assumption of temporally uncorrelated sensory noise. In contrast, if drift is estimated over an interval shorter than the ISI, we would expect the displacement error to grow proportionally to time as a consequence of extrapolation: $\sigma_E \propto t$.

In Fig. 4*C*, we tested which of these two alternative hypotheses best fit the data when the ISI, T, was increased from 100 ms to 500 ms. In the 500 ms condition, the standard deviation of the prior was correspondingly increased by a factor of $\sqrt{5}$ to reflect the five-fold increment in the interval between bar exposures, as dictated by the assumption that ocular drift resembles Brownian motion. The uncertainty in the retinal signal (σ_R) remained the same as in the 100 ms condition. The uncertainty in the extraretinal cue (σ_E) was either enlarged by a factor a $\sqrt{5}$ or 5 as suggested by the two hypotheses. Individual subjects data and model predictions are reported in Fig. S5.

Data availability. Data are available from the Harvard Dataverse at https://doi.org/10.7910/DVN/FYKP2L. Source data are provided with this paper.

Code availability. The Matlab code for analyzing the data and generating the figures is available at:

References

[1] F. Ratliff and L. A. Riggs. Involuntary motions of the eye during monocular fixation. *J. Exp. Psychol.*, 40(6):687–701, 1950.

[2] R. W. Ditchburn. Eye movements in relation to retinal action. *Opt. Acta*, 1:171–176, 1955.

[3] R. M. Steinman, J. Z. Levinson, H. Collewijn, and J. Van der Steen. Vision in the presence of known natural retinal image motion. *J. Opt. Soc. Am. A*, 2(2):226–233, 1985.

[4] Y. Burak, U. Rokni, M. Meister, and H. Sompolinsky. Bayesian model of dynamic image stabilization in the visual system. *Proc. Natl. Acad. Sci. USA*, 107(45):19525–19530, 2010.

[5] W. H. Marshall and S. A. Talbot. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. In H. Kluver, editor, *Biological Symposia—Visual Mechanisms*, volume 7, pages 117–164, Lancaster, PA, 1942. Cattel.

[6] E. Ahissar and A. Arieli. Figuring space by time. *Neuron*, 32(2):185–201, 2001.

[7] M. Rucci, E. Ahissar, and D. Burr. Temporal coding of visual space. *Trends Cogn. Sci.*, 22(10):883–895, 2018.

[8] A. G. Anderson, K. Ratnam, A. Roorda, and B. A. Olshausen. High-acuity vision from retinal image motion. *J. Vis.*, 20(7):1–19, 2020.

[9] J. B. Reppas, W. M. Usrey, and R. C. Reid. Saccadic eye movements modulate visual responses in the lateral geniculate nucleus. *Neuron*, 35(5):961–974, 2002.

[10] I. Kagan, M. Gur, and D. M. Snodderly. Saccades and drifts differentially modulate neuronal activity in V1: Effects of retinal image motion, position, and extraretinal influences. *J. Vis.*, 8(14):1–25, 2008.

[11] M. A. Sommer and R. H. Wurtz. Brain circuits for the internal monitoring of movements. *Annu. Rev. Neurosci.*, 31:317–338, 2008.

[12] M. Ibbotson and B. Krekelberg. Visual perception and saccadic eye movements. *Curr. Opin. Neurobiol.*, 21(4):553–558, 2011.

[13] J. M. McFarland, A. G. Bondy, R. C. Saunders, B. G. Cumming, and D. A. Butts. Saccadic modulation of stimulus processing in primary visual cortex. *Nat. Commun.*, 6(1):1–14, 2015.

[14] H. H. Li, A. Barbot, and M. Carrasco. Saccade preparation reshapes sensory tuning. *Curr. Biol.*, 26(12):1564–1570, 2016.

[15] A. Benedetto and M. C. Morrone. Saccadic suppression is embedded within extended oscillatory modulation of sensitivity. *J. Neurosci.*, 37(13):3661–3670, 2017.

[16] J. Intoy, N. Mostofi, and M. Rucci. Fast and nonuniform dynamics of perisaccadic vision in the central fovea. *Proc. Natl. Acad. Sci. USA*, 118(37):1–9, 2021.

[17] L. M. Kroell and M. Rolfs. The peripheral sensitivity profile at the saccade target reshapes during saccade preparation. *Cortex*, 139:12–26, 2021.

[18] M. Nawrot. Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. *Vision Res.*, 43(14):1553–1562, 2003.

[19] J. W. Nadler, M. Nawrot, D. E. Angelaki, and G. C. DeAngelis. MT neurons combine visual motion with a smooth eye movement signal to code depth-sign from motion parallax. *Neuron*, 63(4):523–532, 2009.

[20] M. Lappe, F. Bremmer, and A. V. Van den Berg. Perception of self-motion from visual flow. *Trends Cogn. Sci.*, 3(9):329–336, 1999.

[21] M. Rolfs, D. Jonikaitis, H. Deubel, and P. Cavanagh. Predictive remapping of attention across eye movements. *Nat. Neurosci.*, 14(2):252–256, 2011.

[22] M. Poletti, D. C. Burr, and M. Rucci. Optimal multimodal integration in spatial localization. *J. Neurosci.*, 33(35):14259–14268, 2013.

[23] L. D. Sun and M. E. Goldberg. Corollary discharge and oculomotor proprioception: Cortical mechanisms for spatially accurate vision. *Annu. Rev. Vis. Sci.*, 2(1):61–84, 2016.

[24] R. H. Wurtz. Corollary discharge contributions to perceptual continuity across saccades. *Annu. Rev. Vis. Sci.*, 4:215–237, 2018.

[25] P. Binda and M. C. Morrone. Vision during saccadic eye movements. *Annu. Rev. Vis. Sci.*, 4:193–213, 2018.

[26] T. N. Cornsweet. Determination of the stimuli for involuntary drifts and saccadic eye movements. *J. Opt. Soc. Am.*, 46(11):987–993, 1956.

[27] A. Fiorentini and A. M. Ercoles. Involuntary eye movements during attempted monocular fixation. *Atti Fondazione Giorgio Ronchi*, 21:199–217, 1966.

[28] J. M. Findlay. Direction perception and human fixation eye movements. *Vision Res.*, 14(8):703–711, 1974.

[29] M. Poletti, C. Listorti, and M. Rucci. Stability of the visual world during eye drift. *J. Neurosci.*, 30(33):11143–11150, 2010.

[30] I. Murakami and P. Cavanagh. A jitter after-effect reveals motion-based stabilization of vision. *Nature*, 395(6704):798–801, 1998.

[31] J. Nachmias. Determiners of the drift of the eye during monocular fixation. *J. Opt. Soc. Am.*, 51(7):761–766, 1961.

[32] R. M. Steinman, G. M. Haddad, A. A. Skavenski, and D. Wyman. Miniature eye movement. *Science*, 181(102):810–819, 1973.

[33] M. Poletti, M. Aytekin, and M. Rucci. Head-eye coordination at the microscopic scale. *Curr. Biol.*, 25(24):3253–3259, 2015.

[34] J. Intoy and M. Rucci. Finely tuned eye movements enhance visual acuity. *Nat. Commun.*, 11(795):1–11, 2020.

[35] L. Z. Gruber and E. Ahissar. Closed loop motor-sensory dynamics in human vision. *PLoS ONE*, 15(10):1–18, 2020.

[36] Y. C. Lin, J. Intoy, A. Clark, M. Rucci, and J. D. Victor. Cognitive influences on fixational eye movements during visual discrimination. *J. Vis.*, 21(9):1894, 2021.

[37] H. K. Ko, D. M. Snodderly, and M. Poletti. Eye movements between saccades: Measuring ocular drift and tremor. *Vision Res.*, 122:93–104, 2016.

[38] R. Engbert, K. Mergenthaler, P. Sinn, and A. Pikovsky. An integrated model of fixational eye movements and microsaccades. *Proc. Natl. Acad. Sci. USA*, 108(39):765–770, 2011.

[39] X. Kuang, M. Poletti, J. D. Victor, and M. Rucci. Temporal encoding of spatial information during active visual fixation. *Curr. Biol.*, 22(6):510–514, 2012.

[40] R. J. Krauzlis and S. G. Lisberger. A model of visually-guided smooth pursuit eye movements based on behavioral observations. *Front. Comput. Neurosci.*, 1(4):265–283, 1994.

[41] M. Spering and A. Montagnini. Do we track what we see? Common versus independent processing for motion perception and smooth pursuit eye movements: A review. *Vision Res.*, 51(8):836–852, 2011.

[42] J. J. O. de Xivry, S. Coppe, G. Blohm, and P. Lefevre. Kalman filtering naturally accounts for visually guided and predictive smooth pursuit dynamics. *J. Neurosci.*, 33(44):17301–17313, 2013.

[43] M. Rucci and J. D. Victor. The unsteady eye: an information-processing stage, not a bug. *Trends. Neurosci.*, 38(4):195–206, 2015.

[44] E.A. Benardete and E. Kaplan. The receptive field of the primate P retinal ganglion cell, I: Linear dynamics. *Visual Neurosci.*, 14(1):169–185, 1997.

[45] T. Tayama. The minimum temporal thresholds for motion detection of grating patterns. *Perception*, 29(7):761–769, 2000.

[46] B. G. Borghuis, D. Tadin, M. J. Lankheet, J. S. Lappin, and W. A. van de Grind. Temporal limits of visual motion processing: psychophysics and neurophysiology. *Vision*, 3(1):1–17, 2019.

[47] C. A. Johnson and R. P. Scobey. Foveal and peripheral displacement thresholds as a function of stimulus luminance, line length and duration of movements. *Vision Res.*, 20(8):709–771, 1980.

[48] R. W. Ditchburn and B. L. Ginsborg. Vision with a stabilized retinal image. *Nature*, 170(4314):36–37, 1952.

[49] J. Epelboim and E. Kowler. Slow control with eccentric targets: Evidence against a position-corrective model. *Vision Res.*, 33(3):361–380, 1993.

[50] E. Riva Sanseverino, C. Galletti, M.G. Maioli, and S. Squatrito. Single unit responses to visual stimuli in cat cortical areas 17 and 18: III. responses to moving stimuli of variable velocity. *Arch. Ital. Biol.*, 117:248–267, 1979.

[51] M. Aytekin, J. D. Victor, and M. Rucci. The visual input to the retina during natural head-free fixation. *J. Neurosci.*, 34(38):12701–12715, 2014.

[52] N. Shelchkova, C. Tang, and M. Poletti. Task-driven visual exploration at the foveal scale. *Proc. Natl. Acad. Sci. USA*, 116(12):5811–5818, 2019.

[53] L. Matin, J. Pola, E. Matin, and E. Picoult. Vernier discrimination with sequentially-flashed lines: Roles of eye movements, retinal offsets and short-term memory. *Vision Res.*, 21(5):647–656, 1981.

[54] E. Kowler, J. F. Rubinstein, E. M. Santos, and J. Wang. Predictive smooth pursuit eye movements. *Annu. Rev. Vis. Sci.*, 5:223–246, 2019.

[55] J. Fooken, P. Kreyenmeier, and M. Spering. The role of eye movements in manual interception: A mini-review. *Vision Res.*, 183:81–90, 2021.

[56] C. Galletti, P. P. Battaglini, and P. Fattori. Parietal neurons encoding spatial locations in craniotopic coordinates. *Exp. Brain Res.*, 96(2):221–229, 1993.

[57] D. Melcher and M.C.. Morrone. Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nat. Neurosci.*, 6(8):877–881, 2003.

[58] G. d'Avossa, M. Tosetti, S. Crespi, L. Biagi, D.C. Burr, and M.C. Morrone. Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nat. Neurosci.*, 10(2):249–255, 2007.

[59] R. W. Sperry. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.*, 43(6):482–489, 1950.

[60] I. Donaldson. The functions of the proprioceptors of the eye muscles. *Phil. Trans. R. Soc. Lond. B*, 335(1404):1685–1754, 2000.

[61] B.C. Motter and G.F. Poggio. Dynamic stabilization of receptive fields of cortical neurons (VI) during fixation of gaze in the macaques. *Exp. Brain Res.*, 83(1):37–43, 1990.

[62] M. Gur and D.M. Snodderly. Visual receptive fields of neurons in primary visual cortex (V1) move in space with the eye movements of fixation. *Vision Res.*, 37(3):257–265, 1997.

[63] D. W. Arathorn, S. B. Stevenson, Q. Yang, P. Tiruveedhula, and A. Roorda. How the unstable eye sees a stable and moving world. *J. Vis.*, 13(10):1–19, 2013.

[64] L. A. Riggs, F. Ratliff, J. C. Cornsweet, and T. N. Cornsweet. The disappearance of steadily fixated visual test objects. *J. Opt. Soc. Am.*, 43(6):495–501, 1953.

[65] Alexander Rivkind, Or Ram, Eldad Assa, Michael Kreiserman, and Ehud Ahissar. Visual hyperacuity with moving sensor and recurrent neural computations. In *International Conference on Learning Representations*, 2021.

[66] U. J. Ilg, S. Schumann, and P. Thier. Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron*, 43(1):145–151, 2004.

[67] M. Spering and K. R Gegenfurtner. Contrast and assimilation in motion perception and smooth pursuit eye movements. *Brain Res.*, 98(3):1355–1363, 2007.

[68] T. C. Freeman, R. A. Champion, and P. A. Warren. A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Curr. Biol.*, 20(8):757–762, 2010.

[69] A. R. Bogadhi, A. Montagnini, and G. S. Masson. Dynamic interaction between retinal and extraretinal signals in motion integration for smooth pursuit. *J. Vis.*, 13(13):1–26, 2013.

[70] F. Santini, G. Redner, R. Iovin, and M. Rucci. EyeRIS: A general-purpose system for eye movement contingent display control. *Behav. Res. Methods*, 39(3):350–364, 2007.

Acknowledgments: This work was supported by the National Institutes of Health grants EY18363 (MR) and EY07977 (JV). We thank Claudia Cherici and David Richters for their help in preliminary experiments and Janis Intoy and Martina Poletti for helpful comments and discussions.

Author Contributions: ZZ implemented the experiments and the model, collected and analyzed experimental data, and ran simulations. EA and MR conceived the original idea. JV helped interpret data, formalize the ideas, and develop the model. MR supervised the project. All authors contributed to the writing of the article.

Competing Interests: The authors declare no competing financial interest.

Figure Captions

Figure 1. Estimating spatial relations via eye movements.

Figure 2. Controlling for visual cues

Figure 3. Integration of visual and motor cues.

Figure 4. Characteristics of the extraretinal signal.

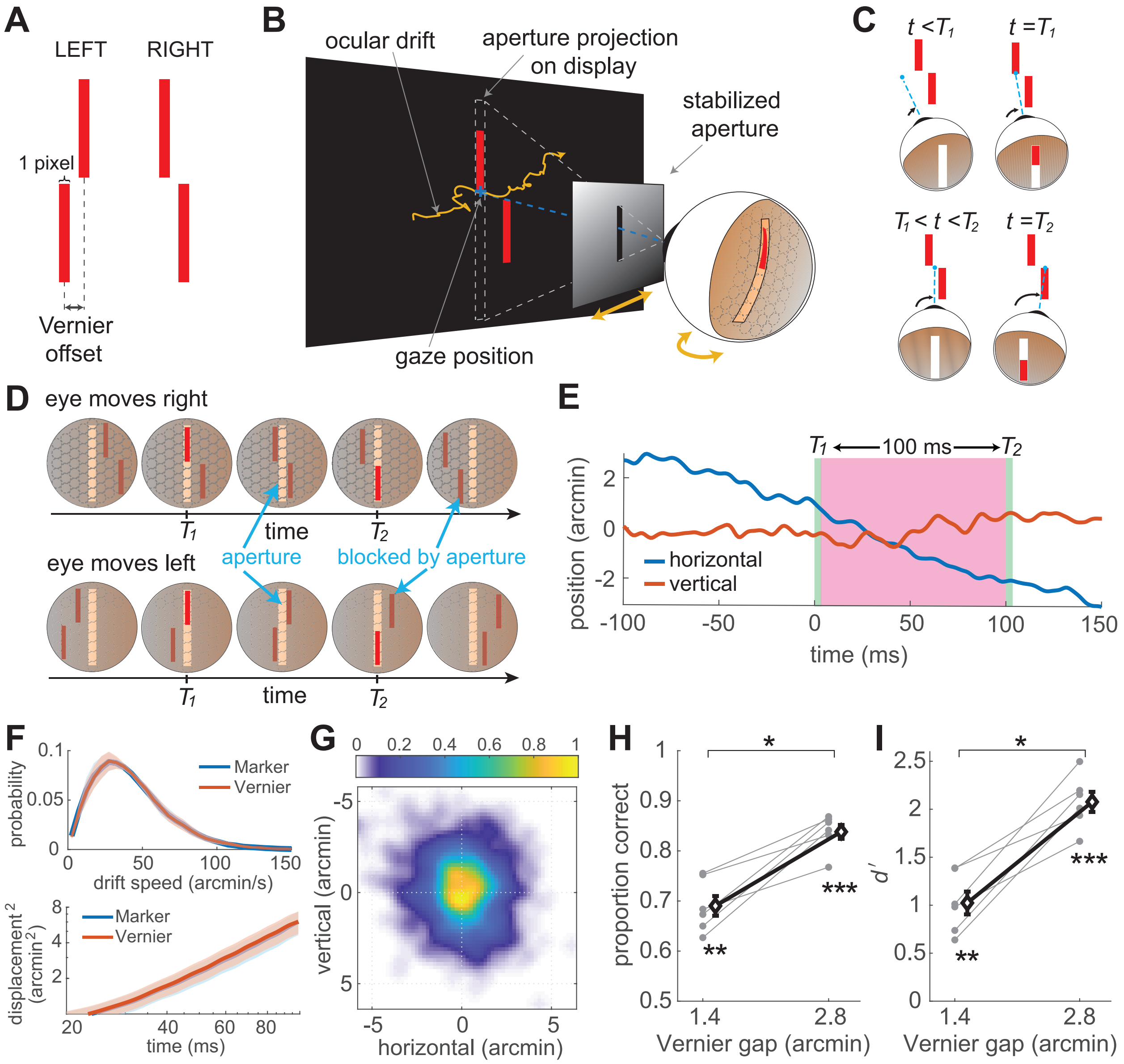

Fig.1: **Estimating spatial relations via eye movements.** (**A-C**) Experimental design. (*A*) Subjects reported the spatial configuration of a Vernier (left or right) viewed through a retinally-stabilized aperture. (*B*) The aperture moved together with the eye, to allow stimulation of only a thin vertical strip on the retina. The width of the aperture was equal to that of each bar in the Vernier (28′ long; 1.4′, the angle covered by one pixel on the CRT). (*C*) In this way, each Vernier bar was visible only when it directly overlapped with the aperture, resulting in vertically-aligned bar exposures on the retina. (**D**) Motor knowledge of eye movements is required to accomplish this task. The same visual input signals can be obtained with different configurations of the stimulus, when the eye drifts in opposite directions. (**E**) Example trace of eye movements in a trial. The shaded green regions mark the periods of exposure of each Vernier bars. The pink region indicates the inter-stimulus interval (ISI), here 100 ms. (**F-I**) Ocular drift characteristics and performance in the task. Data from $N = 6$ human observers. (**F**) Mean eye speed and displacement are virtually identical to those

measured in the same subjects while fixating on a marker. Shaded regions represent ± one SEM across subjects. (**G**) Average probability distribution of gaze displacement in between bar exposures. (**H-I**) Subjects correctly reported the configuration of the stimulus. Both proportion of correct responses and discriminability index were significantly above chance (H: $^{\star\star}p = 3.16 \times 10^{-4}$; $^{\star\star\star}p = 3.44 \times 10^{-6}$; I: $^{\star\star}p = 5.02 \times 10^{-4}$, $^{\star\star\star}p = 9.16 \times 10^{-6}$, two-tailed t-test) and improved as the Vernier gap increased (H: $^{\star}p = 0.0024$; I: $^{\star}p = 0.0016$, paired two-tailed t-test). Gray circles are the individual subjects data. Diamonds and associated error bars represent averages ± one SEM across subjects. Source data are provided as a Source Data file.

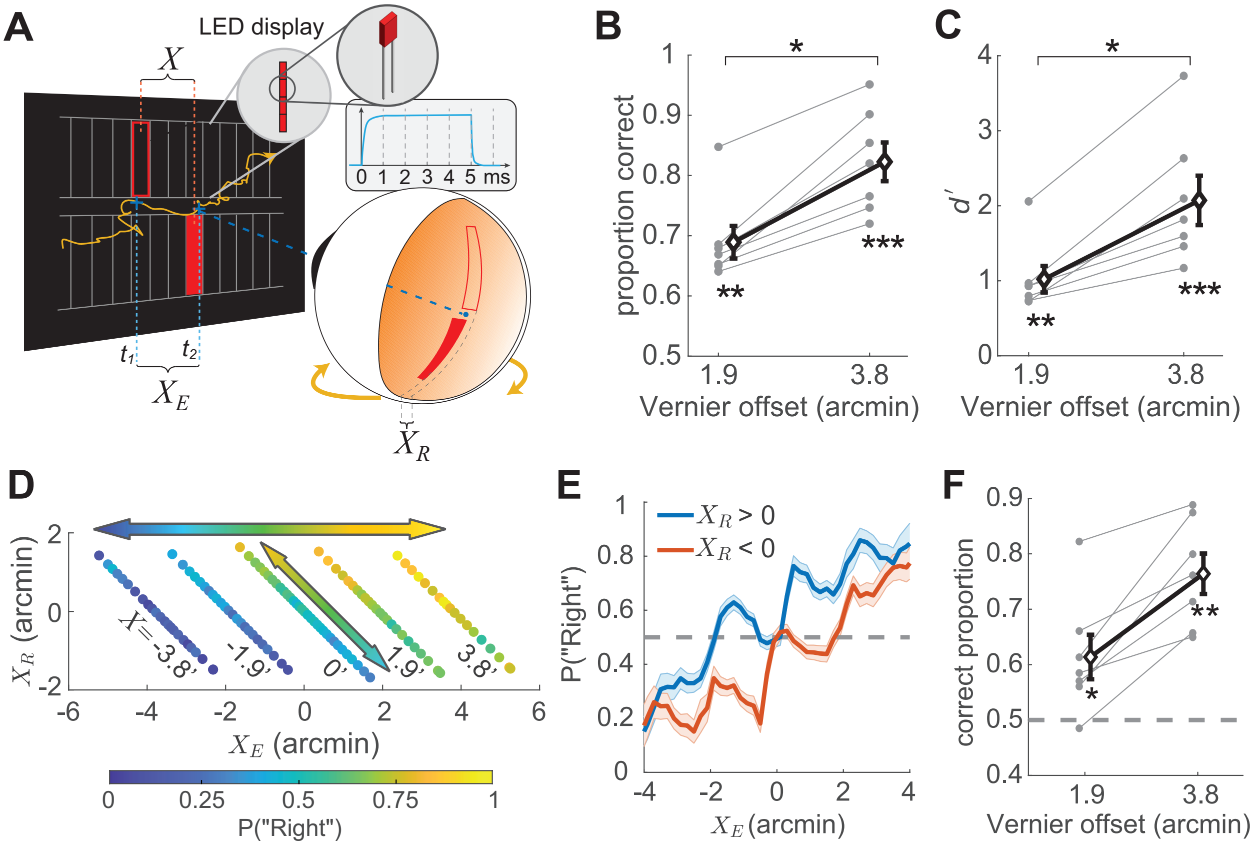

Fig. 2: **Controlling for visual cues.** (**A**) A custom LED display developed specifically for this study. This display, an array of 110× 8 LEDs, each covering 1.9′ on the horizontal meridian, was designed to provide no persistence and no background luminance. The insert shows the time-course of activity of one of the LEDs with the brief exposures used in our experiments, measured with a high-speed photocell. (**B-F**) Performance measured with 5 ms exposures (N=7 subjects). Both (**B**) proportions of correct responses and (**C**) the discriminability index improved as the Vernier gap increased ($^{\star}p = 7.21 \times 10^{-4}$ in B and 1.65×10^{-3} in C; paired two-tailed t-tests) and were significantly above chance (B: $^{\star\star}p = 4.5 \times 10^{-4}$, $^{\star\star\star}p = 5 \times 10^{-5}$; C: $^{\star\star}p = 1.17 \times 10^{-3}$, $^{\star\star\star}p = 7.49 \times 10^{-4}$, two-tailed t-tests). Graphic conventions are as in Fig. 1H-I, with diamonds representing mean values ± SEM across subjects. **(D)** Probability of ”Right” responses as a function of both the eye displacement in a trial (X_E) and the small misalignment on the retina caused by the display resolution ($X_R < 1.9'$; one LED, see panel A). Negative and positive X_R indicate that, on the retina, the bottom bar was shifted to the left or right, respectively. Each diagonal line represents a Vernier offset X on the display. **(E)** Marginal probability of ”Right” responses as a function of the eye displacement in a trial for both $X_R < 0$ and $X_R > 0$. The shaded regions represent one SEM. Perceptual reports are influenced by X_R (the oscillations in both curves) but primarily driven by X_E (the overall trend). **(F)** Mean performance ± SEM in the trials in which X_E and X_R possessed opposite signs. Subjects successfully completed the task even when X_R predicted the wrong response ($^{\star}p = 0.0297$ and $^{\star\star}p = 3.54 \times 10^{-4}$ above chance; two-tailed t-test). Source data are provided as a Source Data file.

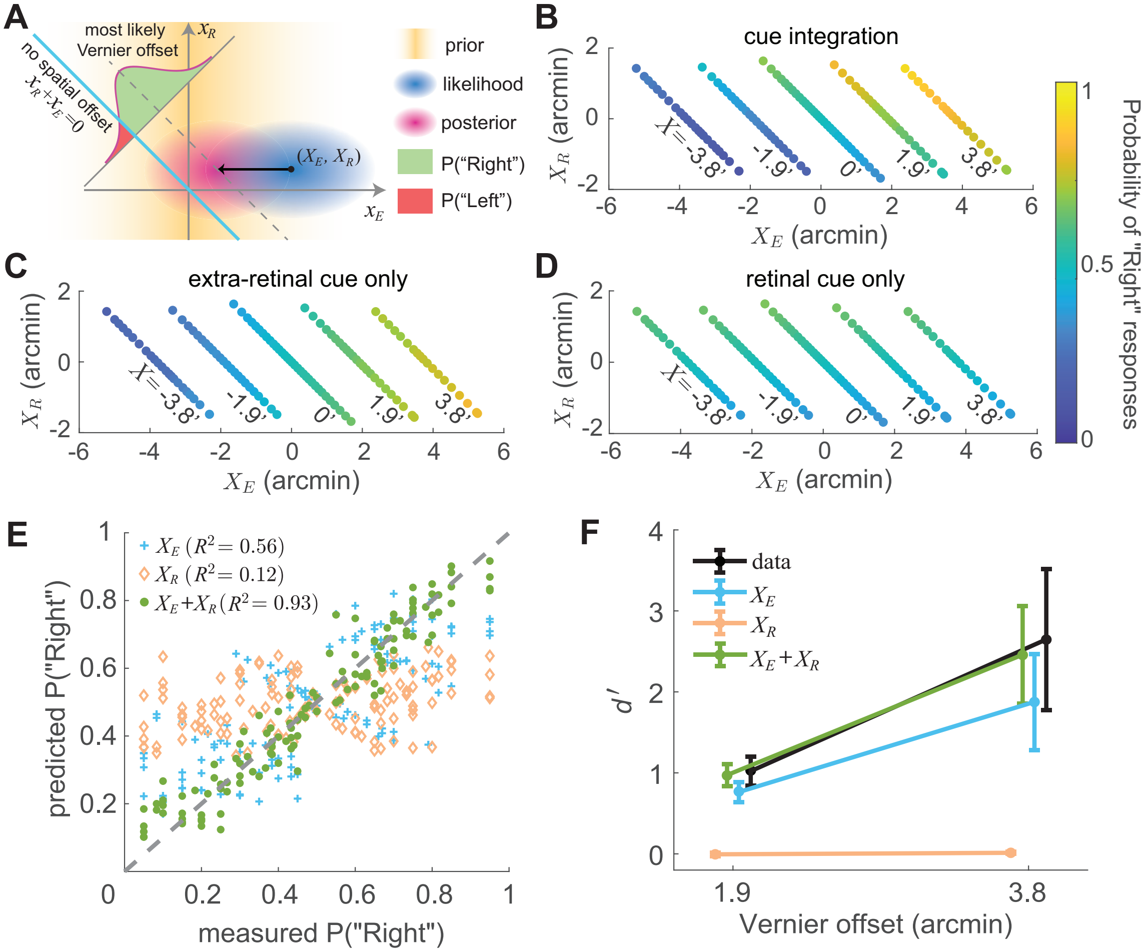

Fig. 3: **Integration of visual and motor cues.** **(A)** An ideal observer model that combines retinal (x_R) and extraretinal (x_E) signals. The model assumes sensory measurements to be corrupted by unbiased additive Gaussian noise (standard deviations σ_R and σ_E) and applies a uniform prior to x_R and a zero-mean Gaussian prior to x_E ($\sigma = \sqrt{2DT}$, where T is the ISI), the latter based on the assumption that ocular drift resembles Brownian motion. In each trial, the likelihood of any given combination (X_E, X_R) is first converted into a joint posterior probability distribution and then integrated on the -45° line $x_E + x_R = X$ (dashed line) to evaluate the probability of any given Vernier offset X. The left/right perceptual report in a trial is determined by which side of the zero-offset line (cyan line) gives higher probability. **(B-D)** Response patterns are best predicted by the model when combining both cues (*B*); the model (*C*) that only uses extraretinal information does not account for the dependence on X_R, and the model (*D*) that uses only retinal information fits poorly. Graphic conventions are as in Fig. 2*D*. **(E-F)** Comparison of experimental data and model predictions of responses and d′: (*E*) Probability of responding "Right" for various combinations of X_R and X_E (the same data points as in *B-D*). (*F*) Average performance measured as d′ (N=7 subjects). Error bars represent ± one SEM across subjects. The cue integration model predicts both subject responses and overall performance with significantly greater accuracy than the single-cue models. Source data are provided as a Source Data file.

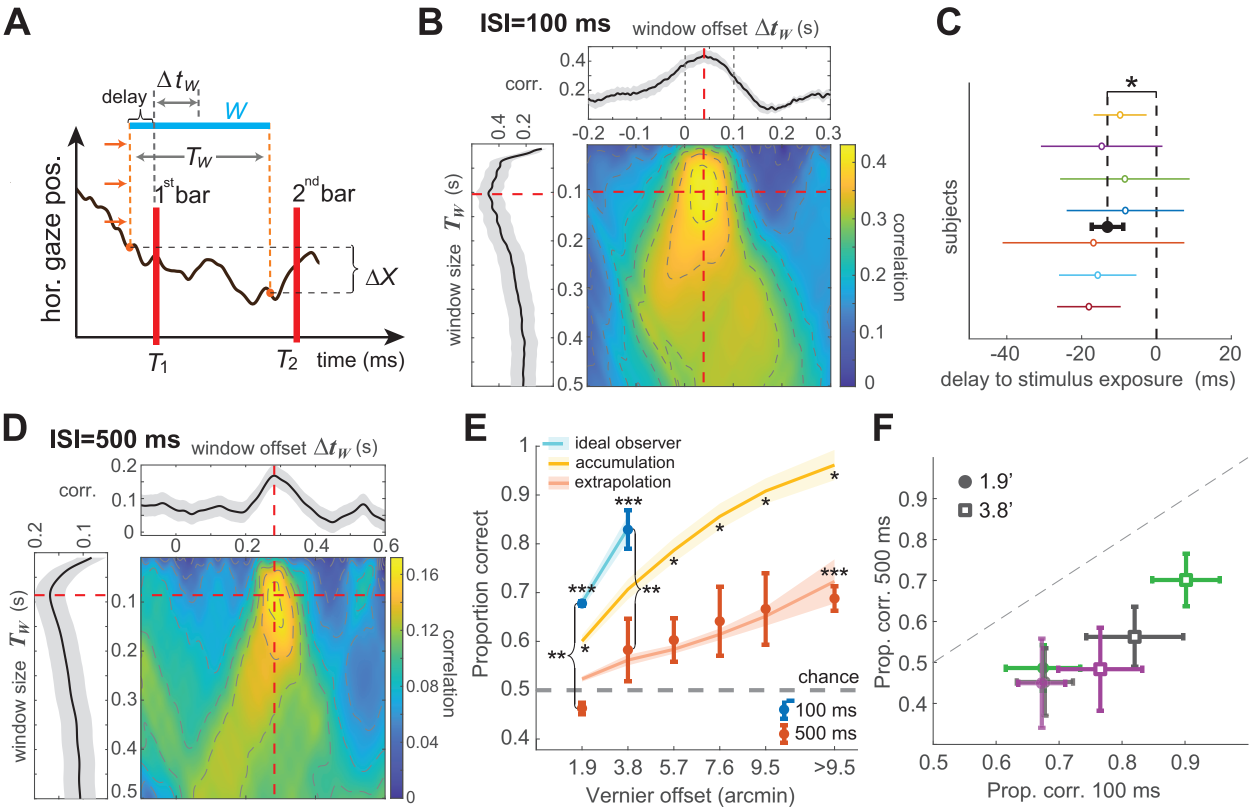

Fig. 4: **Characteristics of the extraretinal signal.** **(A)** Determination of the horizontal gaze displacement $\Delta X = X(\Delta t_W + \frac{T_W}{2}) - X(\Delta t_W - \frac{T_W}{2})$ that best correlates with perceptual reports. T_W represents the duration of the window of observation, Δt_W the temporal lag between the onset of the first bar and the window center. **(B-F)** Results obtained with a 100 ms ISI (panels *B-C*, N=7 subjects) and with a 500 ms ISI (panels *D-F*, N=3 subjects). Data were collected using the custom LED display with 5 ms flashes. **(B, D)** Correlation between ΔX and perceptual reports as a function of both window parameters (T_W; vertical axis) and (Δt_W; horizontal axis). The highest correlation is achieved for a 100 ms window that slightly precedes the first bar. Side plots are sections at the optimal T_W and Δt_W (red dashed lines). Shaded regions represent ± one SEM across subjects. **(C)** The timing of maximum correlation for each individual subject. On average, subject responses are best correlated with a 100-ms window that anticipated the stimulus by 13 ms (filled black circle; $^\star p = 1.52 \times 10^{-4}$, two-tailed t-test). Error bars represent ± one SD. **(E)** Comparison of performance with 100 and 500 ms ISIs. Performance was lower in the 500 ms condition ($^{\star\star}p < 0.005$, paired one-tailed t-test) and improved marginally with increasing Vernier offset (***above chance; $p < 0.009$, one-tailed t-test). Empirical data are consistent with the prediction from the ideal observer model with σ_E adjusted to increase proportionally to time (red curve) and lower than predicted by increasing $\sigma_E \propto \sqrt{t}$ (yellow curve; $^\star p < 0.037$, one-tailed paired t-test). Note that these fits have no free parameters: all parameters were obtained from those estimated over the 100 ms ISI in Figure 3. Error bars and shaded regions represent ± one SEM. **(F)** For each observer (different colors) performance at both 1.9′ and 3.8′ gaps was always lower in the 500 ms condition ($p < 0.027$, one-tailed bootstraps over an average of N=87 trials across subjects and gaps). Error bars represent ± one SEM. Source data are provided as a Source Data file.

Supplementary Information: Inferring visual space from ultra-fine extra-retinal knowledge of gaze position

Zhetuo Zhao[1,2], Ehud Ahissar[3], Jonathan D. Victor[4], Michele Rucci[1,2]

[1]Department of Brain and Cognitive Sciences and [2]Center for Visual Science, University of Rochester, NY, USA. [3]Department of Brain Sciences, Weizmann Institute of Science, Rehovot, Israel. [4]Feil Family Brain and Mind Research Institute, Weill Cornell Medical College, New York, NY, USA

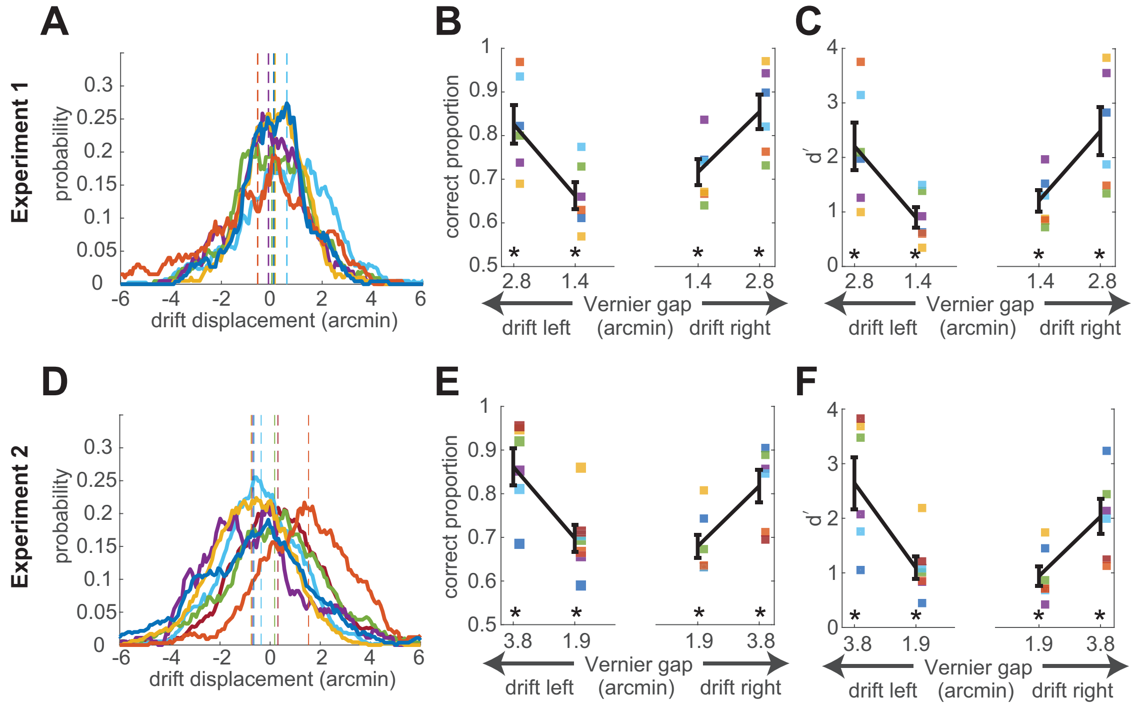

Fig. S1: **Performance as a function of ocular drift direction.** **(A-C)** Results from Experiment 1 (N=6 subjects). **(A)** Distributions of horizontal eye displacement in the 100 ms inter-stimulus interval of Experiment 1. Data from individual subjects are shown in separate curves. The vertical dashed lines mark the means of the distributions. Note that for all observers means are close to zero, *i.e.*, drift displacements were unbiased. **(B-C)** Performance in Experiment 1 measured as both proportion correct (B) and d' (C) for displacements in both directions. Black lines represent averages ± one SEM across subjects. Squares are data from individual subjects ($^\star p < 0.0035$ in B and < 0.0052 in C, two-tailed t-tests). Results with drifts in both directions were similar. **(D-F)** Similar analyses for the data from Experiment 2 (N=7 subjects) Graphic conventions are identical to the panels above, with black lines representing mean values ± SEM across subjects ($^\star p < 7.2 \times 10^{-4}$ in E and < 0.0027 in F, two-tailed t-tests). Source data are provided as a Source Data file.

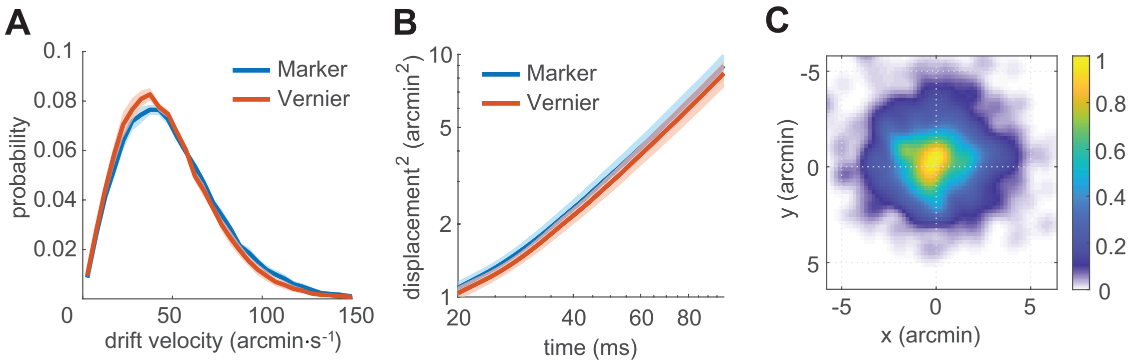

Fig. S2: **Ocular drift characteristics in Experiment 2.** (**A**) Average distribution of eye speed. (**B**) Squared displacement as a function of time. (**C**) 2D probability of overall drift displacement. Data represent averages across individuals and refer to the 100 ms ISI interval. Shaded regions represent ± one SEM. For comparison, the same measurements obtained while maintaining fixation on a 5′ dot (marker) are also shown in *A* and *B*. Source data are provided as a Source Data file.

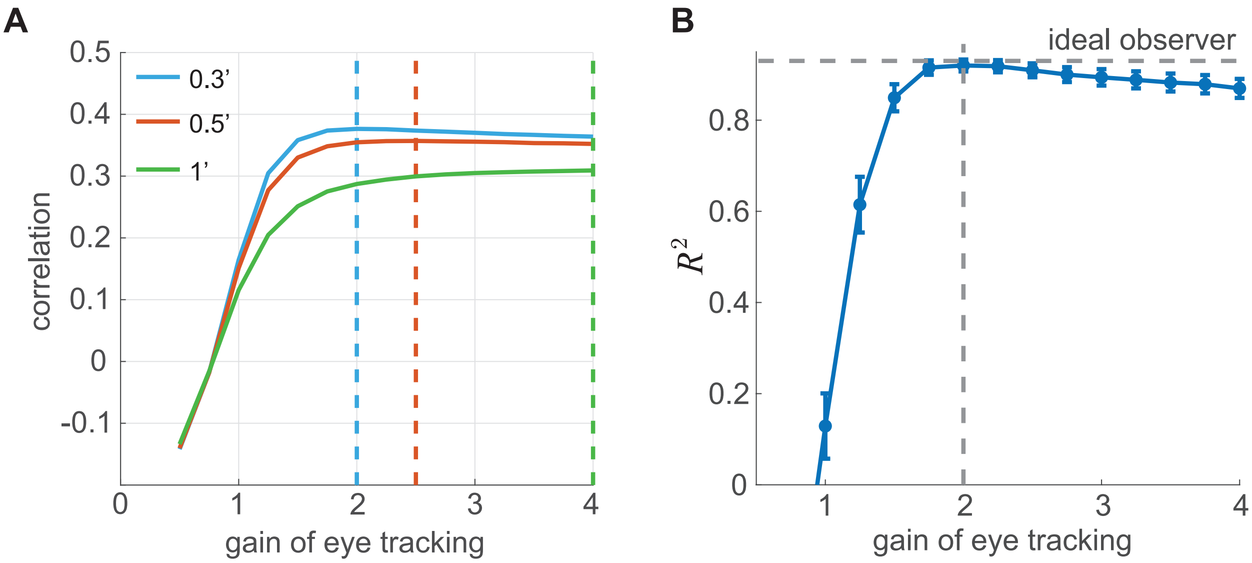

Fig. S3: **Consequences of over-estimating eye movements.** Results of Monte Carlo simulations that modeled the eye-tracker output as $\widehat{x} = \gamma x_e + \eta$, where x_e is the horizontal gaze displacement; γ represents the eye-tracker gain; and $\eta = N(0, \sigma)$ is a Gaussian noise term with zero mean and standard deviation σ_η. (**A**) Correlation between subject's responses and the resulting retinal misalignment $(X - \widehat{x})$ as a function of γ. The three curves represent results with different σ_η. The lower boundary for σ_η, as measured with a stationary artificial eye is 0.3. Note that the correlation never exceeds 0.4. Vertical dashed lines show the gains for which the curves reach their maximum. **(B)** Maximum variance in perceptual reports that could be explained by an ideal observer only using this retinal cue. For each γ, the perceptual uncertainty in the retinal measurement was estimated to maximize the R^2 as in Eq. 4. To account for subject's responses, the eye-tracker would need to overestimate the gaze displacement by approximately a factor of 2 (vertical line), which is unrealistic. The dashed horizontal line marks the variance accounted by the ideal observer in Figure 3, which assumes measurements of eye drifts to be veridical ($\gamma = 1$). For each gain, R^2 was evaluated over the $N = 124$ cue combinations of Fig. 3B. Errorbars represent ± one SEM from bootstrap. Source data are provided as a Source Data file.

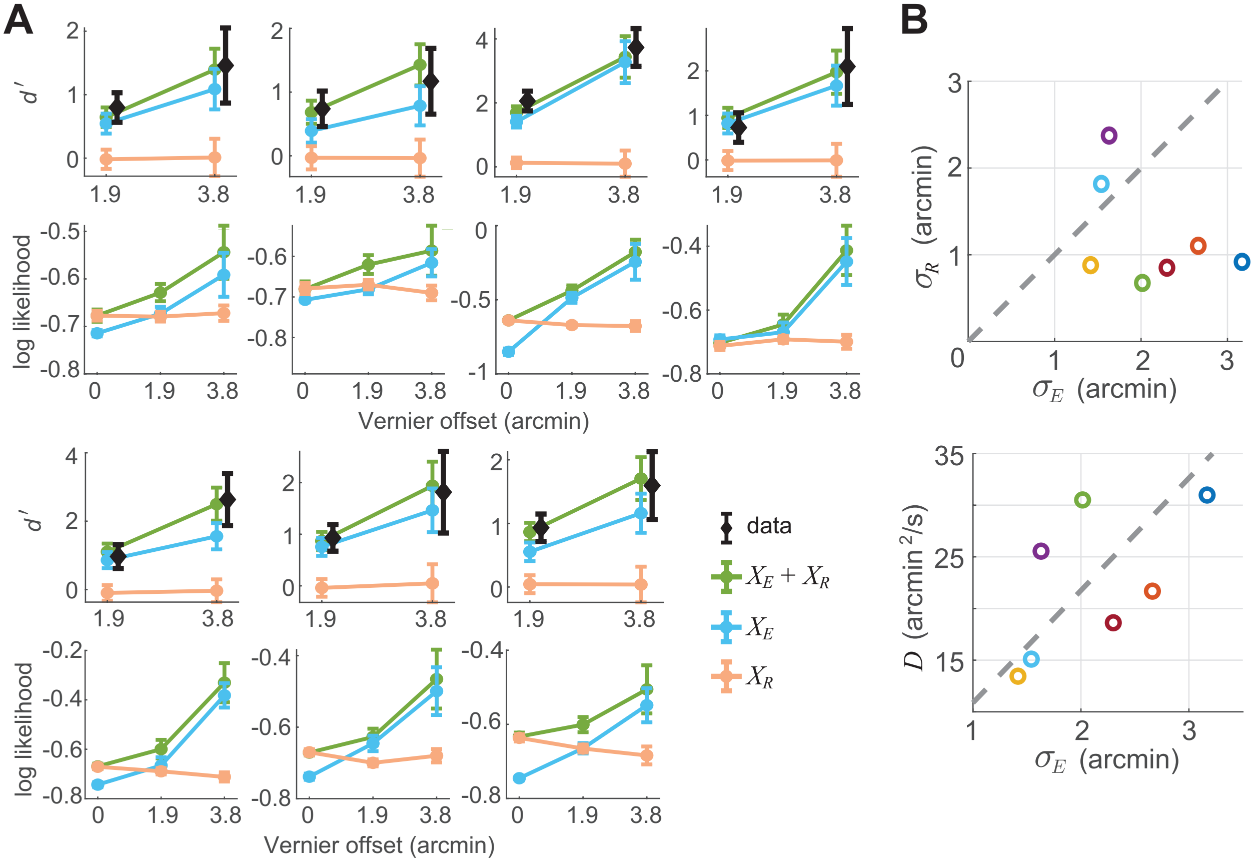

Fig. S4: **Model parameters and predictions for individual subjects.** **(A)** Three models are compared for each of the 7 subjects in Experiment 2: the full Bayesian model ($X_R + X_E$) and the two reduced models with a single cue (X_R or X_E). The performance of each model was evaluated by means of both predicted d′ (top row) and the mean log likelihood of all trials given the model (L in Eq. 4, see Methods. Bottom row). Errorbars are ± one SEM derived from bootstraps over an average of N=176 trials across subjects and gaps. **(B)** Model parameters fitted to empirical data in Experiment 2. Each data point corresponds to one subject: The s.d. of retinal noise σ_R and uncertainty in extraretinal displacement estimation σ_E in the top panel; The diffusion coefficient of ocular drift in the bottom panel. Source data are provided as a Source Data file.

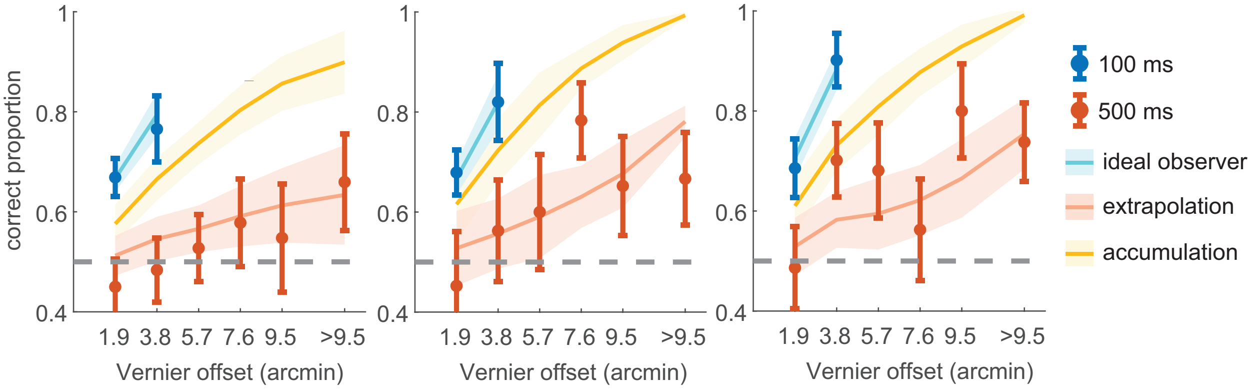

Fig. S5: **Model predictions for individual subjects in Experiment 3.** The graphical convention is the same as Fig. 4C. Error bars of the empirical data and the shaded region of model predictions are ± one SEM derived from bootstraps. Source data are provided as a Source Data file.

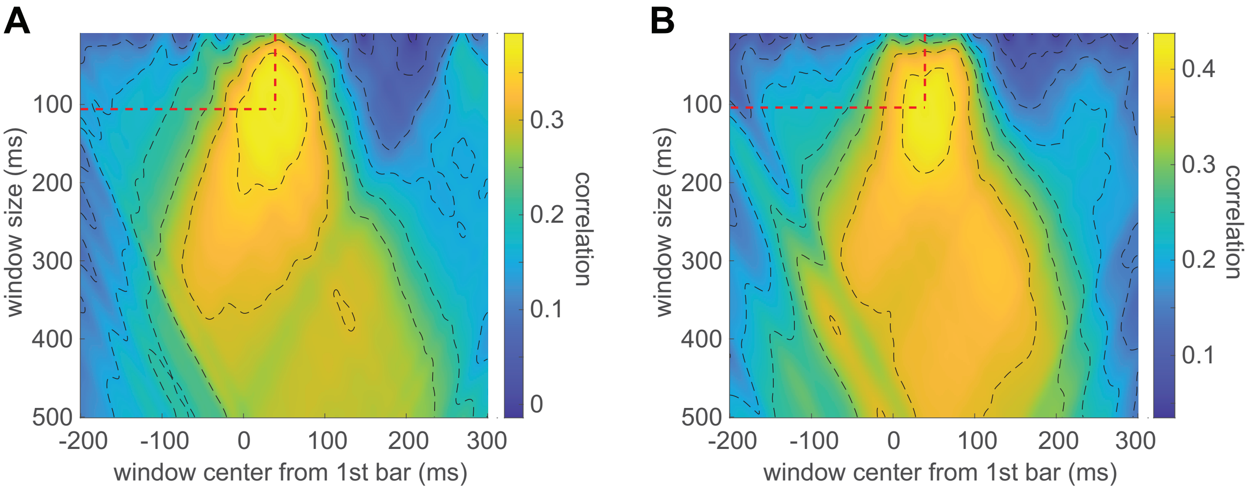

Fig. S6: **Extra-retinal strategies did not change over the course of training.** Correlation between gaze displacement and perceptual reports as a function of the position and duration of the window of observation, as in Fig. 4 (ISI=100). The two panels show averages obtained over the first (A) and last third (B) of the trials collected from each subject in Experiment 2.

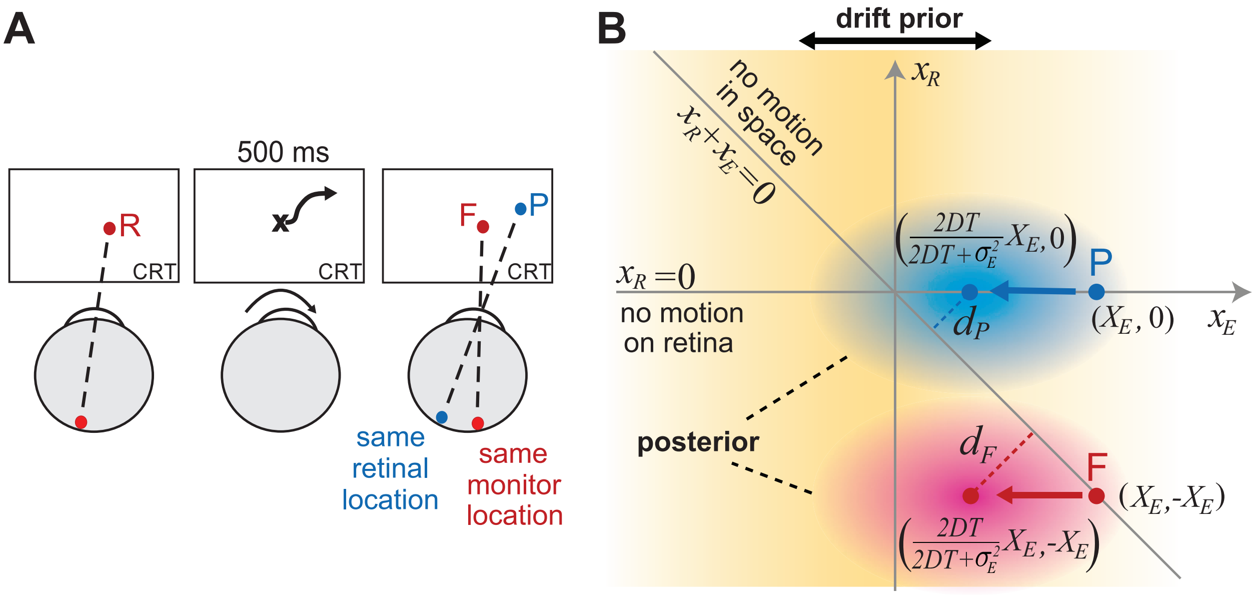

Fig. S7: **Predicted errors in spatial localization.** Our ideal observer model accounts for seemingly contradictory previous findings. (**A**) In a 2AFC task, subjects report the position of a previously displayed reference (R) by selecting between two probes, one at the same reference's location on the display (F) and one at its same position on the retina (P). The more the eye drifts in complete darkness, the less likely subjects are to correctly select F (see [29]). (**B**) The model predicts this paradoxical behavior as a consequence of the specific choice presented to the observer. The oculomotor prior weights identically both probes, causing both posterior distributions to shift towards smaller estimated displacements. The posterior distribution of the retinotopic probe P will be closer to the no-motion line (the line $X_R + X_E = 0$) if the motor uncertainty in measuring the displacement, σ_E, is larger than the variance of the prior ($\sigma_E^2 > 2DT$, where D is the drift diffusion constant and T the ISI). The data in Fig. 4 confirm that this will occur for sufficiently long ISI. Under these conditions, the model predicts that the retinotopic probe P will have higher probability to be mistaken for the reference than the spatiotopic probe F, despite having access to an extraretinal drift signal.