Lawful tracking of visual motion in humans, macaques, and marmosets in a naturalistic, continuous, and untrained behavioral context

Jonas Knöll[a,b,c,d], Jonathan W. Pillow[e,f], and Alexander C. Huk[a,b,c,d,1]

[a]Department of Psychology, The University of Texas at Austin, Austin, TX 78712; [b]Department of Neuroscience, The University of Texas at Austin, Austin, TX 78712; [c]Center for Perceptual Systems, The University of Texas at Austin, Austin, TX 78712; [d]Institute for Neuroscience, The University of Texas at Austin, Austin, TX 78712; [e]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544; and [f]Department of Psychology, Princeton University, Princeton, NJ 08544

Edited by Tony Movshon, New York University, New York, NY, and approved August 31, 2018 (received for review April 26, 2018)

Much study of the visual system has focused on how humans and monkeys integrate moving stimuli over space and time. Such assessments of spatiotemporal integration provide fundamental grounding for the interpretation of neurophysiological data, as well as how the resulting neural signals support perceptual decisions and behavior. However, the insights supported by classical characterizations of integration performed in humans and rhesus monkeys are potentially limited with respect to both generality and detail: Standard tasks require extensive amounts of training, involve abstract stimulus–response mappings, and depend on combining data across many trials and/or sessions. It is thus of concern that the integration observed in classical tasks involves the recruitment of brain circuits that might not normally subsume natural behaviors, and that quantitative analyses have limited power for characterizing single-trial or single-session processes. Here we bridge these gaps by showing that three primate species (humans, macaques, and marmosets) track the focus of expansion of an optic flow field continuously and without substantial training. This flow-tracking behavior was volitional and reflected substantial temporal integration. Most strikingly, gaze patterns exhibited lawful and nuanced dependencies on random perturbations in the stimulus, such that repetitions of identical flow movies elicited remarkably similar eye movements over long and continuous time periods. These results demonstrate the generality of spatiotemporal integration in natural vision, and offer a means for studying integration outside of artificial tasks while maintaining lawful and highly reliable behavior.

motion | eye tracking | optic flow

For effective interactions with the world, animals have to perform some degree of spatiotemporal integration to form decisions and guide appropriate motor actions. This fundamental transition from sensory to cognitive processing has been studied extensively in the context of visual motion processing (1–3). Visual motion is a model system for this because of the apparent simplicity of the involved circuits in the primate brain, and also because motion is inherently defined as integrating information appropriately over space and time. This work has operated within the context of artificial tasks that involve dissociated (and often intentionally arbitrary) stimulus-to-response mappings, and thus, in turn, rely on extensive training. This classical approach is elegant for isolating various subprocesses, but also raises the specter of generalizability in several domains: Do similar degrees of integration also support naturally occurring behaviors? Can substantial integration occur volitionally but naturally, or is it the by-product of extensive training, perhaps only in higher primates with substantial cognitive capacities?

Although ecological validity and broad generalizability are attractive ideals for transcending what can be learned within a well-controlled model system, it is commonly assumed that leaving simplified domains involves substantial losses of experimental rigor and quantitative power. Here, we sidestep this apparent tradeoff with an experimental paradigm that lies at the transition between synthetic and naturalistic. By coopting a naturally occurring behavior controlled by simple aspects of visual stimuli, we demonstrate the capability to make detailed quantitative assessments of visual integration with strong experimental efficiency and statistical power—thereby establishing the general nature of spatiotemporal integration and facilitating more detailed and broader insights into this basic sensory–cognitive computation.

To test for naturally occurring spatiotemporal integration, we exploited the basic insight that primates look at important things: They naturally direct their gaze where they are going, and also look at things that are directly approaching them. Regardless of whether a dynamic visual array corresponds to self-motion or object motion, such naturally salient events can be defined by the spatiotemporal structure of movement across the visual field. This allowed us to leverage a naturally occurring but goal-directed behavior as the crux of a synthetic behavioral context that supports detailed quantitative characterizations analogous to those of classical tasks. We implemented this motion structure within a dynamic field of dots, and defined the direction of (either ego or object) motion as the focus of expansion (FOE) within this dynamic dot field. In the simplest ideal case, retinal velocity at the FOE is zero, and all velocity vectors emanate from this single point. In our implementation, we added random

Significance

We characterize spatiotemporal integration of naturalistic, continuous visual motion of three primate species (humans, macaques, and marmosets). All three species volitionally, but naturally, track the center of expansion of a dynamic optic flow field. Detailed analysis of this flow-tracking behavior reveals lawful and repeatable dependencies of the behavior on nuances in the stimulus, revealing that even unconstrained and continuous behavior can exhibit the sort of precise dependencies typically studied only in artificial and constrained tasks.

Author contributions: J.K., J.W.P., and A.C.H. designed research; J.K. performed research; J.K. and J.W.P. contributed new reagents/analytic tools; J.K. and A.C.H. analyzed data; and J.K., J.W.P., and A.C.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All data have been deposited in an Open Science Framework repository at https://osf.io/h5rxv/.

See Commentary on page 11112.

[1]To whom correspondence should be addressed. Email: huk@utexas.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1807192115/-/DCSupplemental.

Published online October 15, 2018.

perturbations over space and time to partially decorrelate this idealized spatiotemporal structure, supporting the characterization of how portions and periods of the stimulus guide gaze behavior over space and time.

When the FOE moved according to a random walk, we found that three primate species (humans, rhesus macaques, and common marmosets) tracked the meandering FOE with their freely occurring gaze. Such tracking was often continuous for long periods of time, thus providing a very high data efficiency: Characterizations of temporal integration that typically involve many hours worth of data were now attainable within a few minutes. The detailed oculomotor behavior was reliable over repeats of identical stimuli and contained large fractions of fixations followed by saccades toward the tracked FOE. Analyses of slow eye movements in between saccades and fixations were inconsistent with more reflexive behaviors, such as optokinetic nystagmus and ocular following (4–6). Taken together, these results offer evidence that these naturally occurring, goal-directed patterns of gaze allow for detailed spatiotemporal characterizations, provide insights via analysis of behavior across repetitions of identical trials, and allow for integration of studies across a range of species with widely varying cognitive abilities.

Results

We measured eye movements during continuous tracking of the meandering FOE of an optic flow pattern in three primate species (one human, subject, H; two rhesus monkeys, M1 and M2; and one common marmoset, C), in which the FOE moved according to a random walk defined by the global velocity pattern of a large dot field (Fig. 1, *SI Appendix*, and Movie S1). Subjects followed the FOE with their gaze with little instruction, training, or practice (see *Materials and Methods*). A single continuous trial could last up to 5 min. Trials were ended by the subjects by either blinking for longer than 250 ms or by looking near or beyond the edges of the screen. Continuous trial time varied by subject and comprised extended consecutive periods (median successful trial duration: H, 300 s; M1, 23 s; M2, 140 s; C, 300 s).

Fig. 2 shows the goal-directed and lawfully reliable nature of this flow-tracking behavior. It displays eye movements across "frozen repeats" of the same motion displays (i.e., repetitions with the same random seed, and thus identical stimulus characteristics). Eye position traces comprised periods of stable (or low-velocity) fixation, interrupted by recentering saccades. A majority of saccades were directed toward the FOE, even though the flow pattern switched from contracting to expanding multiple times (and hence the local motions alternated from pointing mostly toward to mostly away from the FOE). To evaluate the goal-directed nature of the saccades, we calculated the percentage of saccades in which the component parallel to the eye–FOE axis was reduced. Evidence for FOE-targeted saccades was similar in human and macaque subjects: For both expanding (H, 85%; M1, 80%) and contracting (H, 82%; M1, 75%) stimuli, saccades were more likely to be directed toward the FOE, which makes them distinct from more reflexive oculomotor behaviors such as optokinetic responses, for which fast phases are typically directed against the motion and thus would switch with changing flow patterns within our paradigm.

A closer inspection of the eye movements during these frozen seed trials also revealed strong similarities across repetitions within observers (individual traces in Fig. 2, *Insets*). This is best seen in a video (*SI Appendix* and Movie S3) where the meandering FOE (black square) and the eye trace from each frozen trial (orange discs) are shown over time. Again, human and macaque subjects exhibited similar repeatability of gaze patterns, although the larger macaque dataset makes this more compelling and allows for richer quantification. Thus, although the experimental design and lack of explicit training impose very little direct behavioral constraint, the resulting behavior is very predictable down to single similar gaze patterns occurring repeatedly within windows as small as 300 ms.

Quantification of Flow-Tracking Performance. To quantify the temporal integration underlying this naturally occurring visually guided "flow-tracking" behavior, we characterized the relation between the gaze and the history of FOE positions (using ridge regression) on data with unique motion displays ("liquid seed"). Gaze was mostly influenced by the FOE of roughly the previous 100 ms to 500 ms, with a peak impact at a lag of ∼200 ms (see Fig. 3*A*, gray lines, for the shape of the full kernel). With this, we were able to predict the gaze patterns on left-out (frozen seed) data using the extracted temporal kernels (Fig. 2, blue lines). The gaze is generally predicted well from the kernels and the prediction resembles the gaze much more closely than it follows the raw FOE (*SI Appendix* and Movie S2) and captures systematic differences between the FOE and the gaze tracking it.

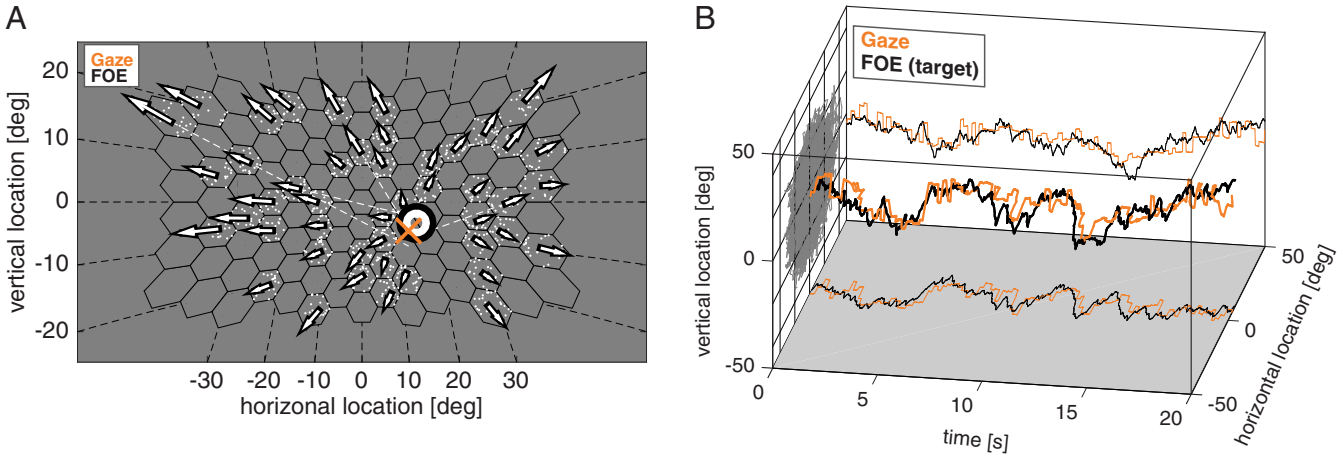

Fig. 1. Naturalistic tracking paradigm. (*A*) Dots were moving in 115 hexagonal subfields according to a randomly varying FOE with a small randomly varying offset for each subfield, covering the central 80° by 50° of the visual field. White arrows show the median direction and relative speed of dots within each subfield. Dashed white lines project from the arrows of selected subfields to the location of the FOE encoded by their dot's motion. (*B*) Example spatiotemporal course of an FOE (black) and gaze (orange) from a macaque over 20 s. Thin lines show projections onto the *x*-time and *y*-time planes. Gray lines on the *xy* plane show the distribution of FOE locations throughout one 300-s trial. See also *SI Appendix* and Movie S1.

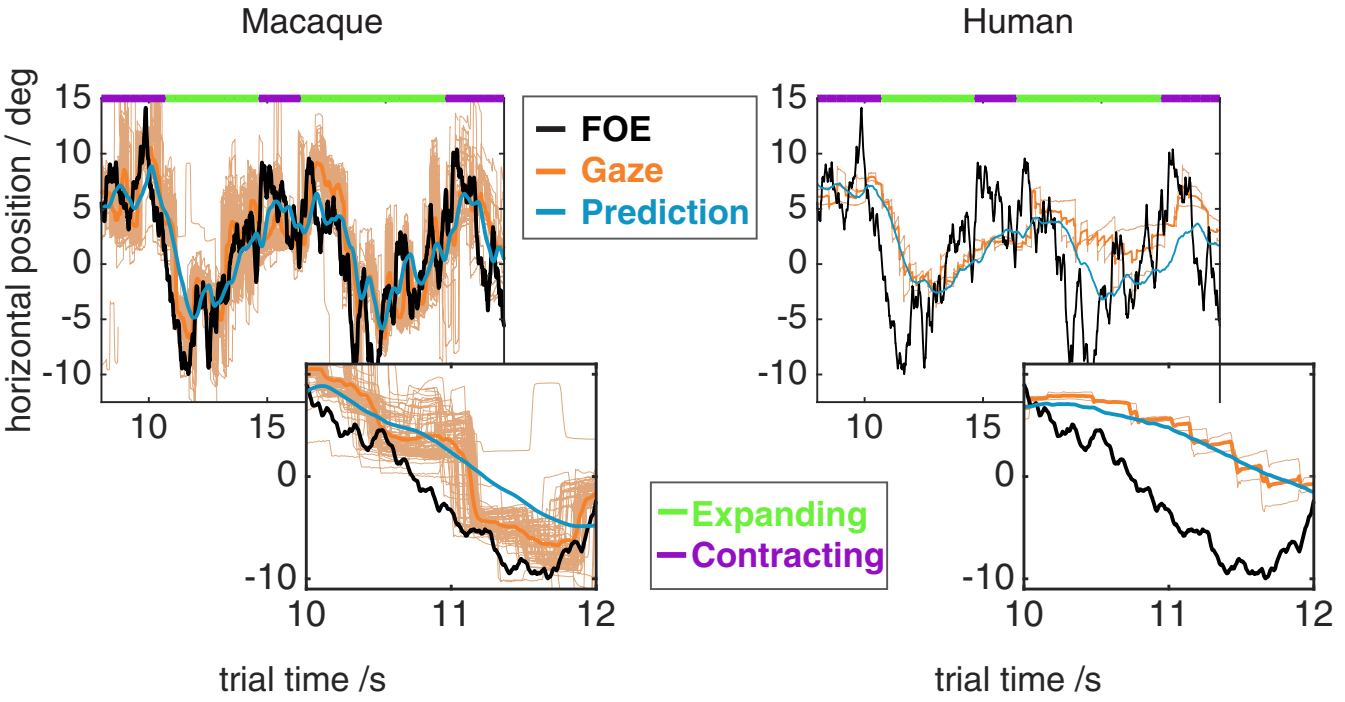

Fig. 2. Reliable goal-directed eye movements. Eye traces (thin orange lines, single repeat; thick orange lines, average) from a subset of trials which had identical stimulation (15% of all trials) for a macaque (*Left*) and a human (*Right*). Saccades appear to be very predictable and directed toward the FOE (black). See also *SI Appendix* and Movie S3. Gaze is well explained by the prediction (blue), which is the FOE filtered with the temporal kernel (Fig. 3*A*) obtained from all other trials (with random stimulation parameters). Time points from 1 s before to 1 s after a blink are removed. *Insets* show an enlarged view of 2 s of data.

Given the mostly discrete nature of the flow-tracking behavior (i.e., fixations plus saccades), we performed a saccade-triggered analysis in which we interpreted each saccade endpoint as a decision about the estimated FOE location that was formed over some amount of time before the saccade start. This temporal dissociation between the time the decision was made (relative to saccade start) and the time the estimate takes effect (saccade end) automatically shifts the kernel toward earlier times by the mean saccade duration. The measured saccade-triggered kernels (Fig. 3*A*, turquoise lines) also concentrate the weight over a narrower time window, indicating that this analysis captures a temporally precise aspect of the tracking behavior, but that it still involves substantial integration.

In further quantifications, we found that the fitted kernels provided better accounts of the gaze than a variety of alternatives (Table 1). For example, kernel-predicted gaze outperformed a variety of null models (e.g., from stable fixation without tracking to perfect zero-lag tracking) and models described by simpler mechanisms (e.g., perfect tracking with optimal lag). Together, these comparisons provide quantitative confirmation not only that subjects were tracking the stimulus but that the idiosyncrasies of their tracking behavior are well accounted for by the simple kernel estimates described earlier.

To characterize the spatial integration of motion, we had also perturbed the FOE represented by the motion in each hexagon by a random amount that was reassigned in each subfield at random intervals (two frames to 0.5 s). To determine the retinocentric spatial integration, we performed a regression analysis on each hexagon's FOE time course (in retinocentric coordinates) on the residuals from the temporal regression. Spatial integration was largely dominated by input from the central hexagon that included the fovea, with some contribution of the ring of hexagons surrounding the central hexagon (Fig. 3*B*). This pattern of spatial integration is in line with the current location of gaze playing a central role in determining future gaze position, as well as prior studies that have emphasized wide-field spatial integration of optic flow (7), as information correlated with the FOE can be sampled at any eccentricity.

Efficiency of Estimation and of Task Acquisition. Robust temporal kernels could be estimated from very little data (Fig. 4). Clear structure of temporal integration emerged after just 30 s of tracking, and captured most of the dynamics after 5 min to 10 min of data collection. Given that sessions lasted several times longer than that (median single FOE data per session in minutes: M1, 125; M2, 33; H, 47; C, 30), this efficiency provides surplus experimental power for addressing other (more complex) aspects of perceptual function in extensions of the paradigm.

We carefully monitored the training of one macaque and analyzed the behavior from the first time this monkey was ever subjected to an optic flow stimulus to about 360 min of training. To facilitate initial learning of the task, the reward contingency was formalized to dispense reward whenever the accumulated reward from a Gaussian spatial field around the FOE exceeded a set threshold. Initially, there were no empty hexagons and no subfield directional perturbations. Over the course of the next few hours, we increased the number of empty hexagons as well as the spatial distortions. Fig. 5 shows the performance over the course of this training. The median distance of the gaze to the FOE over 10-min slices of data (Fig. 5*A*) started low and further decreased and reaches asymptote after about 100 min. The temporal kernels showed the same windows of temporal integration from the start of the training and quickly stabilized after about 300 min (Fig. 5*B*, similarity of plotted curves for 240 min and 320 min). In summary, rhesus monkeys figured out the general stimulus–response mapping within minutes, and stabilized their performance within hours. Thus, this paradigm requires only a small fraction of training time relative to conventional behavioral paradigms, and involves refinement on time scales that are amenable to the study of

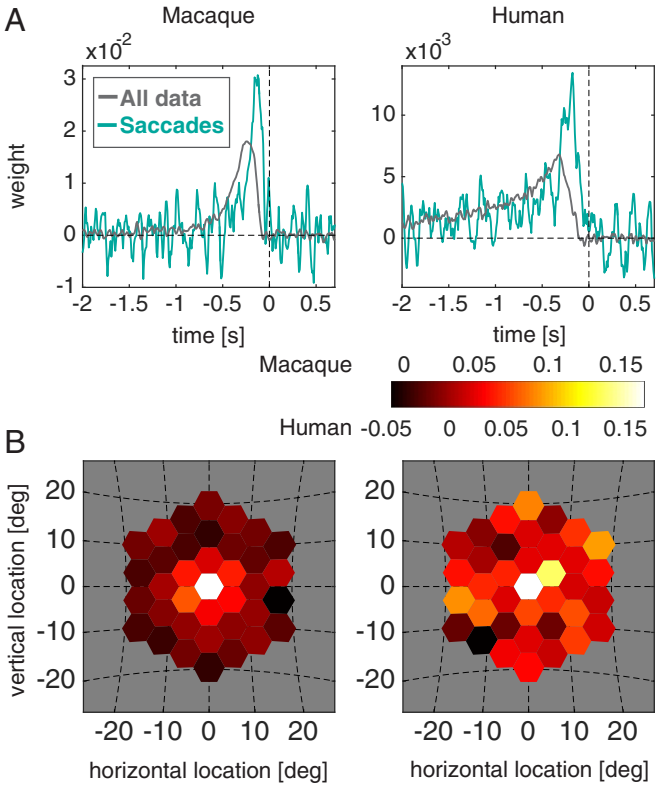

Fig. 3. Spatiotemporal motion integration during single FOE tracking for a macaque (*Left*) and a human (*Right*). (*A*) Temporal integration kernels for all data (gray) or triggered on saccades (turquoise). Kernels were estimated by a ridge regression of the horizontal component of the FOE with different lags to the horizontal component of the gaze. For saccade-triggered kernels, FOE data were limited to time points of the detected saccade onset and regressed against the detected location of the saccade end points. (*B*) Foveocentric integration weights for the central 40° × 40° of the visual field. The time course of the individual foci of expansions of each subfield reorganized in a retinocentric frame of reference was first filtered with the temporal kernels from *A* to then perform a ridge regression on the residuals from the temporal regression.

changes within a session or across a small number of back-to-back sessions.

Tracking by a Marmoset. The stimulus was designed with the goal of generalizing to animals with lower cognitive abilities compared with humans or rhesus monkeys. To test whether marmosets could also learn to perform flow tracking, we recorded behavior of a marmoset in this paradigm. Because marmosets have a reduced natural range of eye-in-head movements, we scaled down the FOE location range by a factor of 0.25 (8). Additionally, the data reported here were collected without any spatial distortions to the FOE and with 70% of the hexagons containing dots. The width of the hexagons was reduced to half to account for the reduced dynamic range of the FOE.

With these adjustments, the marmoset's gaze quickly reflected tracking of FOE (*SI Appendix* and Movie S4). Fig. 6*A* shows the histogram of performance for the recorded marmoset for 300 min of data as measured by the average distance of the marmoset's gaze to the FOE at any given frame. The performance exceeded that of a comparison dataset, where the eye position data were shuffled within each trial. Fig. 6*B* shows the temporal kernels computed from the recorded data. Results are comparable to human and macaque kernels; although slightly noisier, they exhibit a clearly similar temporal integration time course relative to the other species. This similarity demonstrates that flow tracking grants the potential to study visual integration in a range of species using similar tasks and stimuli.

Extension to a Visual Selection Task. Subjects could track an FOE in a visual selection task comprising multiple alternatives. Even when we extended the flow-tracking paradigm to allow for spontaneous selection between two independent targets (statistically identical FOE patterns represented by dots of different colors in distinct hexagon subfields), subjects were able to track an FOE without notable influence of the other FOE. The task can be seen as an extension of traditional attention tasks to a more dynamic selection of task-dependent information. The use of two statistically identical stimuli provides the additional benefit of allowing the combination of data from either flow to an untracked FOE and a tracked FOE, independent of which FOE was being tracked, and thus providing data for the tracked and untracked FOE in parallel. In the core single-FOE task, 45% of the hexagons were carrying information for the FOE, with 55% of the fields remaining empty at any given time. For this parallel FOE selection task, another 45% of the fields were showing information for a second, statistically independent, FOE (Fig. 7*A*). Subjects were free to track either FOE and were allowed to switch between tracking the one or the other FOE (*SI Appendix* and Movie S5). We used the temporal kernels obtained from the single-FOE conditions to estimate which target was being tracked (Fig. 7*B*) and then rearranged the regression design matrices for both FOEs into a tracked and untracked FOE design matrix. We then performed the same analyses as in the single-FOE task for the tracked and untracked FOE data separately.

The structure of temporal integration for the tracked FOE mimics that from the single-FOE conditions (Fig. 8*A*), albeit with a slightly reduced amplitude, while those from the untracked FOE are at or close to the noise level. The spatial kernels of the tracked FOE (Fig. 8*B*) show the same central peak as observed in the single-FOE condition. This demonstration affirms that subjects can easily select information to track one FOE while discarding most (if not all) information from the other, untracked, FOE.

Discussion

We investigated flow-tracking behavior in three species for a continuously changing FOE. The results show a highly reliable tracking behavior for repeats of the same stimulus, down to occurrences of similar saccades at similar times, and were well predicted from temporal kernels obtained from nonrepeating stimuli. Visual information was integrated over roughly 100 ms to 250 ms, and spatial kernels obtained through perturbations in the flow shown in different subfields of the visual array show a strong reliance on the parafoveal information. In addition, the tracking paradigm allowed for easy training across all three primate species investigated, yielding comparable results with a high data efficiency.

Our characterizations of integration align with classical measures of these properties. The temporal kernels match the integration windows derived in classic trial-based experiments studying the integration of visual motion, which specify ∼100 ms of integration evident in the sensory representation in MT (9) with further temporal integration attributed to cognitive/decision processes (10, 11). The spatial weights also qualitatively align with results from two alternative forced choice tasks and optimal observer analyses on the sensitivity to changes in FOE (7).

The flow-tracking behavior is also distinct from well-described purely reflexive optokinetic eye movements which are typically associated with slow phases in the direction of the dot motions and fast phases in the opposite direction. In this task, saccades

Table 1. Various quantifications of reliability and variability of gaze

Dataset parameters			Residual SD for difference models, deg					
			Reference values		Alternate models			Regression
Subject	No. of repeats	Dimension	Fixation	Frozen average	Frozen single	0 lag	Opt. lag	kernel
M1	68	*x*	6.5	2.6	3.7	4.6	3.9	3.2
		y	5.1	2.5	3.4	4.7	4.0	3.1
		xy	4.6	2.6	3.7	3.6	3.3	2.6
M2	12	*x*	6.6	4.5	7.6	7.4	7.0	5.2
		y	5.2	3.5	5.5	7.3	7.1	4.2
		xy	5.3	3.2	5.7	5.4	5.4	3.6
H	2	*x*	6.2	1.3	2.6	5.3	4.0	2.2
		y	6.4	1.7	3.4	6.3	5.0	2.6
		xy	5.5	1.2	2.5	3.9	3.2	1.8
C	4	*x*	1.7	1.0	2.0	1.9	1.9	1.5
		y	2.2	1.4	2.6	2.5	2.3	2.1
		xy	1.5	0.9	1.8	1.7	1.6	1.4

Performance of different models in predicting the gaze time series during repeats of the same FOE time course (frozen repetitions) for all four subjects (M1, M2, H, C). All values (except # repeats) are the residual SD calculated by subtracting the model's prediction from the actual eye trace, over all samples in a given repeat, and combining over all repeats. Samples from 1 s before to 1 s after a blink or the end of trial were removed. Additionally, samples of time points with less than two repeats for a given subject were discarded. The columns "Fixation" and "Frozen average" provide reference points for the minimum and best a model can be expected to perform. Fixation denotes a model of steady fixation. This reflects the SD of the eye trace itself and is the SD to be explained by other models. Frozen average denotes predicting a given repetition's trace using the mean trace from all repetitions (including its own). This is a measure of a lower bound for the explainable SD, as this metric uses all data to directly predict the time series. "Frozen single" denotes prediction of any one repetition's trace for all other repetitions. This provides a complementary characterization of how well a single repetition can explain another one. The "0 lag" column is the model of perfect tracking. This null model attempts to explain the gaze with the instantaneous FOE, corresponding to perfect tracking with no lag. "Opt. lag" denotes the model of delayed perfect tracking, using a delta function kernel (i.e., with all weight localized to one delay), with the delay optimized to minimize residual SD during liquid trials. "Kernel" denotes residual SD applying the kernel obtained during nonfrozen trials. This characterizes the variability of gaze across repeated frozen trials relative to the predicted gaze, and thus can be compared with the other metrics in earlier columns; note that these values are universally lower (with exception to the lower bound frozen average column), indicating that estimating the full kernel does a better job in accounting for the gaze time series than other, simpler, models.

were often directed toward the FOE, for both expanding and contracting motion patterns. Such eye movements have been described before (12) in the context of optic flow stimuli in free viewing conditions, and we propose that they represent a nonreflexive, goal-directed response to the stimulus. Together, this highlights that the results found in these and future experiments using this paradigm relate back to previous work, supporting the study under naturalistic conditions without abandoning comparability to more-classic paradigms.

Relation to Other Paradigms. The flow-tracking paradigm is highly efficient, yielding temporal integration kernels in as little as 4 min, and requiring minimal training. In some regards, our approach is conceptually similar to those recently introduced by other members of our research group, which involve explicit verbal instruction of human subjects to manually track a small, randomly moving stimulus by controlling a cursor (13). Our approach here does not require verbal instruction, by working within an even more intuitive (and essentially naturally occurring) context that is evident in the quick engagement in the task by both rhesus monkeys and common marmosets. Furthermore, the flow-tracking paradigm adds the ability to determine spatial integration parameters by virtue of using widely distributed visual patterns.

Steering paradigms have been used before to study optic flow, although such tasks have typically required training and have focused on a specific experimental manipulation, as opposed to offering a general platform for assaying spatiotemporal integration (14, 15). Likewise, smooth pursuit eye movement and ocular following paradigms have been expanded to assess spatial integration (16–18), but are typically conceived of within a trial-based format (as opposed to being continuous over long time scales) and do not offer as clear a route to generalization and extension. Thus, while we wish to emphasize that our paradigm is, in fact, richly contextualized within a number of existing approaches, it is uniquely suited to general quantitative analysis, and is broadly generalizable (as described further in *Application to Neurophysiological Data*) while exploiting naturalistic behavior.

Application to Neurophysiological Data. The ability to efficiently characterize motion-selective neural responses during more natural behaviors was one of the key aims for designing the paradigm. Specifically, the stimulus needs to provide information to simultaneously determine the receptive fields and their direction tuning over a wide range of the visual field. The motion of a conventional (fixed-FOE) optic flow stimulus is perfectly correlated, making a reverse correlation of neural responses under stable fixation impossible. Here the tracking behavior already comes as a benefit, as it causes variations of the retinal stimulus parameters. Furthermore, naturally occurring variation of the gaze position relative to the array of subfields composing the stimulus allows for mapping of RFs at finer-than-subfield resolution, although analytic tools for smoothing are likely desirable to maintain analytic tractability (19).

Another source of information for the mapping of receptive fields is the changing pattern of hexagons showing dots and the spatially localized distortions of the motion between hexagons. Similarly, for estimating the direction tuning, the course information of the alternating motion directions during expanding and contracting phases is complemented by variations due to said distortions and the behavior. Together, this allows for a multilinear regression of neural parameters, by iteratively and repeatedly estimating one parameter set (e.g., direction and space) after another. The paradigm may also help bring approaches to characterize neural responses using naturalistic movies (20) during passive fixation to continuous naturalistic behavior.

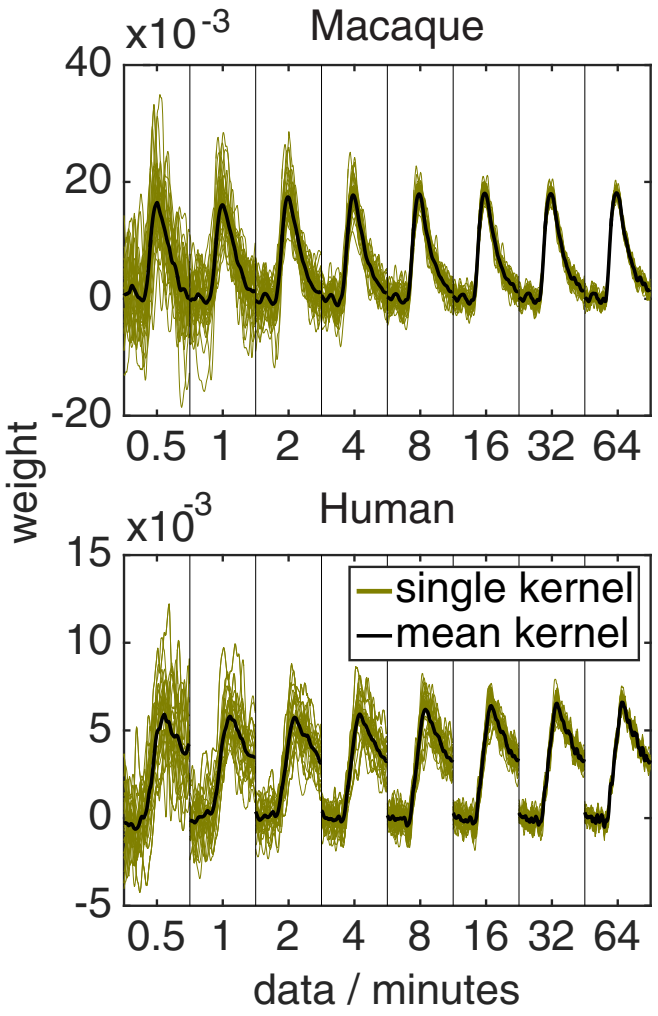

Fig. 4. The flow-tracking paradigm is data-efficient. Shown are 1-s portions of integration kernels obtained from increasing amounts of data. Green lines show the kernels from each of 32 different datasets. Black lines show the average of these 32 estimates. Datasets were created by randomly reordering the trials of the collected data and analyzing the first 0.5 min to 65 min of data.

Broader Applications Within Primates and Across Other Species. Training of the task was also successful in a marmoset and yielded comparable temporal kernels. However, the limited range of eye-in-head movements required the motion of the FOE random walk to be scaled down, thereby reducing the statistical power of the paradigm. This could be improved by using head-free measurements or by placing the animal in a virtual reality (VR) environment, allowing it to move in the direction of the FOE.

Similarly, the paradigm is a tractable one for experiments in rodents, for whom VR environments have recently taken a prominent role (21). In such experiments, visual cues are presented under computer control, often used to probe the neural mechanisms of spatial memory. Versions of the statistical perturbations we have introduced could provide greater leverage on how sensory information is integrated and ultimately translated into memories in such paradigms. Similarly, this approach may also generalize studies of behavioral and neural sensory integration in invertebrates, where navigational patterns are often studied but may benefit from the additional efficiency and statistical power when it comes to understanding the integration of information over space and time, especially considering the emerging technical abilities to study large portions of neural circuits simultaneously.

The efficiency of the FOE-tracking paradigm also makes it a potential tool for other applications within primates, such as the study of perception where long experiments are unsuitable, and/or with populations with reduced abilities to learn complicated new tasks. As an example, scotomas could potentially be detected in the lack of spatial integration of certain subregions of the visual field. Because tracking behavior will correlate with the ability to see the stimulus, another potential option would be to embed secondary visual perception questions into the framework of FOE tracking. For example, to study color perception, one could vary the color of the dots and/or background and analyze tracking performance as a function of chromatic contrast. Likewise, studies of children, the elderly, or patients with neurological dysfunction are likely amenable to this task framework. Relatedly, although we observed strong consistency of gaze patterns within observers, it will be intriguing to test for cross-subject consistency. Such commonalities have been observed in more-natural tasks (22), but gaze patterns for more synthetic stimuli often show less consistency, instead carrying signatures of individual differences in sensitivity across the visual field (23).

We wish to highlight one last critical point regarding generalization: Although we have focused on motion per se in our initial demonstrations, other stimulus dimensions can be assessed within this framework. As described above, one class of generalizations would maintain flow tracking per se but vary nonmotion aspects of the visual scene to focus on the processing of these other cues. Another class of generalization could maintain the spatiotemporal structure of the target but replace it with cues from other senses, such as tracking auditory, somatosensory, or even olfactory targets. Thus, the essence of this paradigm simply involves finding a continuous behavior driven by a stimulus dimension of interest. While we have focused on space and time, any dimensions can be probed given that one can add time-varying statistical fluctuations, central to the estimation of kernels that map stimulus to behavior.

Summary. A variety of primate species exhibited lawful tracking behavior within a continuous, dynamic stimulus. The temporal and spatial kernels obtained with this paradigm align with results obtained with more traditional tasks, demonstrating the generality of spatiotemporal integration in natural vision. This allows the study of mechanisms of spatial and temporal visual

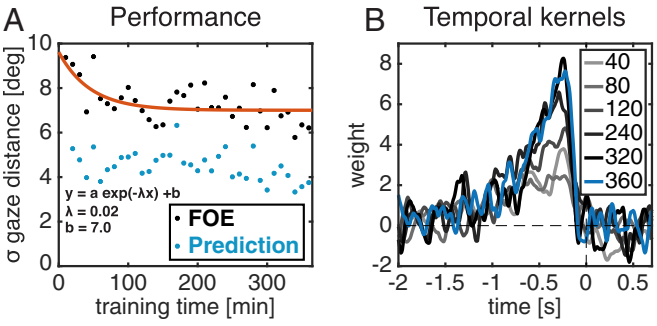

Fig. 5. The task is easily learned by macaques. (*A*) SD in 10-min bins over the first 360 min of collected data of Euclidean distance of gaze to either the FOE (black) or the prediction (blue) obtained using the kernel from the previous bin (excluding frozen seed data), with time points from 1 s before to 1 s after a blink removed. Performance quickly improves and reaches asymptote after about 100 min. (*B*) Evolving temporal kernels during the course of training for a selection of independent 40-min slices of data. Weights of the temporal kernel increase and become stable after about 300 min of training.

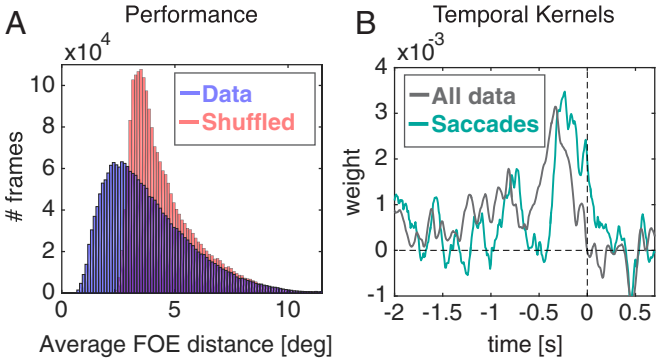

Fig. 6. Marmoset performance and temporal kernels. (*A*) Histogram of 16-s averages of the Euclidean distance of the eye to the FOE for the data (blue) and with the eye position shuffled within each trial (red). (*B*) Temporal kernel for tracking a single FOE.

integration in a more naturalistic behavioral setting by doing away with single decisions after exposure to a stimulus and, instead, moving to the continuous integration of (and reaction to) ever-changing stimuli; these conditions likely dominate in natural behavior. The flow-tracking paradigm we described also supports tractable studies of learning, with a few-hour time course of acquisition to capture the initial learning phase and the ability to potentially guide or manipulate temporal integration by adjusting reward using quickly assessed temporal kernels. Furthermore, the paradigm extends to dual stimulus selection tasks, which are conceptually similar to classical psychophysical tasks that involve binary choices. By reducing the requisite training times while providing high data efficiency per unit time, we hope that this framework will assist the investigation of scientific questions that have previously been challenged by the need for extensive training time, and/or the requirement of averaging over separate measurement periods to have sufficient data to test hypotheses or refine characterizations.

Materials and Methods

Research Subjects. Two rhesus macaques (*Maccaca mulatta*, M1 and M2), one common marmoset (*Callithrix jacchus*, C), and one naïve human (H) took part in this study. The nonhuman primates were kept and handled in accordance with National Institutes of Health guidelines and the Institutional Animal Care and Use Committee at The University of Texas at Austin. Standard surgery procedures were performed to place a head-stabilizing post in both the macaques (24) and marmoset (25). The human observer (male, aged 24 y) had good stereopsis and corrected-to-normal vision. Experiments were undertaken with the written consent of the human observer, and all human experiments were approved by the University of Texas at Austin Institutional Review Board. Across the macaques, marmosets, and human, a total of 39 sessions (M1, 8; M2, 15; C, 9; H, 8) and 1,675 min (M1, 539 min; M2, 601 min; C, 294 min; H, 240 min) of data were analyzed for this paper. Preceding the recording of analyzed data, H was preexposed to the stimulus for two sessions (66 min). M2 had no stimulus preexposure. Optimization of the stimulus parameters to each species (as done for macaques and marmosets, M1 and C) precluded quantification of the task acquisition process for those subjects, but it was similarly on the order of a small number of sessions of exposure, with no explicit training beyond the standard reward contingencies. We collected a relatively modest amount of data in a human subject to establish ground truth for task performance, and then focused on acquiring larger amounts of data in the macaques to confirm the generalizability of the paradigm to nonverbal species and to perform additional quantification.

Experimental Apparatus. Macaque and human participants were seated 57 cm away from a 150 cm × 86 cm rear-projection screen (IRUS; Draper Inc.) covering the central 106° × 73° of visual angle. Images were projected onto the screen by a PROPixx projector (VPixx Technologies Inc.) driven at a resolution of 1,920 × 1,080 pixels at 120 Hz. Marmoset viewing distance was 28.5 cm from a 52 cm × 29 cm ViewPixx display (VPixx Technologies Inc.) covering the central 85° × 54° of visual angle. Eye movements in all species were recorded with head fixed (M1, M2, C; head post) or head stabilized (H; chin and forehead rest) at 1 kHz using an Eyelink 1000 eye tracker (SR Research Ltd.).

Visual stimuli were generated by a Mac Pro-6.1 (H, M1, M2) or 5.1 (C) (Apple Inc.) using Matlab (MathWorks), Psychtoolbox (26), and version 4 of PLDAPS (27). Stimulus code is stored in an online repository, and data are stored on a local server.

Stimuli. The core stimulus comprised a moving cloud of dots, representing a large optic flow field within the central 80° × 50° of visual angle (M2: 60° × 40°), each having individually assigned *x*, *y*, and *z* coordinates in 3D space. Initially, all dots were positioned randomly to uniformly fill the screen, and virtual (*z*) distances to the observer were assigned from a uniform distribution (1 m to 3 m). To characterize the integration of the visual motion across the visual field, we divided the field into hidden hexagonal subfields with a common length of the sides of 3.75° (C: 1.875°) and an area of 36.5 deg^2 (C: 9.16 deg^2). On each frame, the 3D locations of the dots were updated according to the dot cloud's 3D velocity and then drawn at the resulting location projected onto the screen. When the depth of a dot fell below or

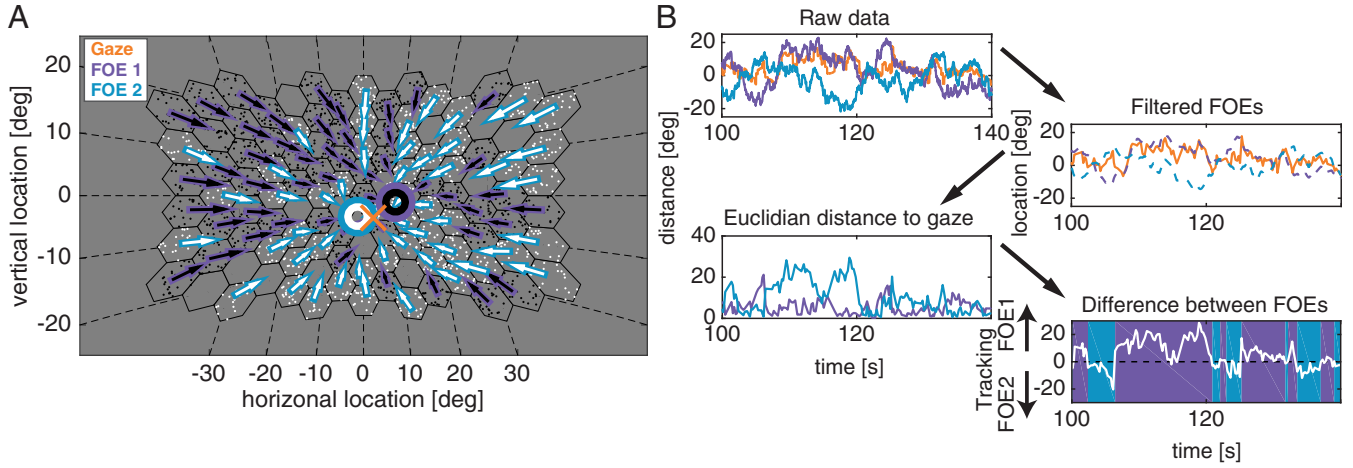

Fig. 7. Parallel FOE selection task. (*A*) Dots in the subfields are moving according to either of two statistically independent FOEs drawn in separate colors (black or white). The shown FOE of each subfield is randomly reassigned, with 10% of subfields remaining empty. Subjects were free to track either FOE and switch from tracking the one FOE to the other. (*B*) Tracked FOE is estimated by using integration kernels from single-FOE tracking conditions (Fig. 3) to predict gaze given the FOE being tracked (filtered FOEs). The FOE with the smallest prediction error (Euclidean distance to the gaze) at any given time is considered to be tracked.

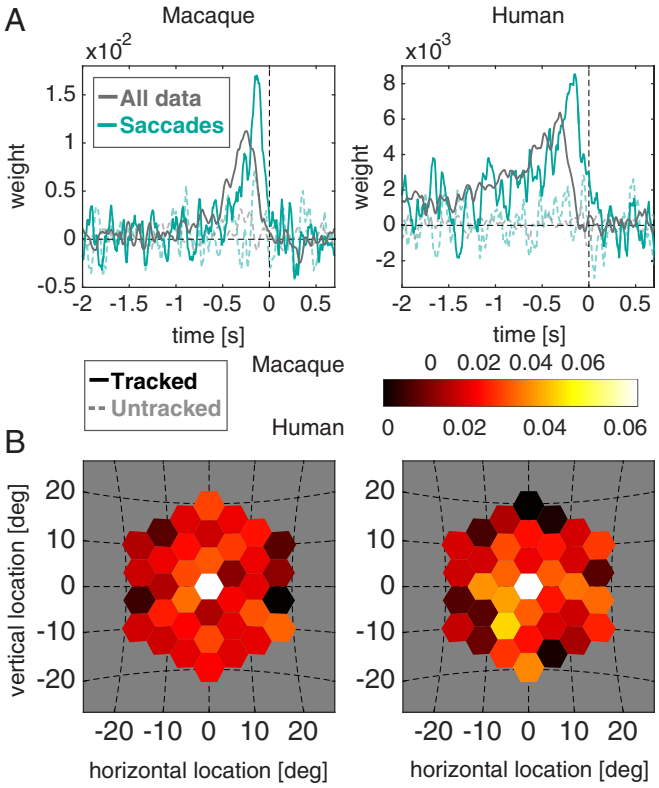

Fig. 8. Spatiotemporal integration during parallel FOE selection. The tracked FOE was first determined based on the prediction errors for each FOE using the temporal kernel estimated during single-flow conditions. Data of the parallel FOEs stimulus were reorganized to parallel tracked and untracked FOE datasets and analyzed analogously to the single-flow data. (*A*) Temporal kernels for all data (gray) or saccade-triggered (turquoise) for the tracked FOE. Kernels for the untracked FOE are shown as opaque and dashed lines. (*B*) Foveocentric integration weights for the central $40^\circ \times 40^\circ$ of the visual field for the tracked FOE.

above 0.5 m or 3.5 m, respectively, it was replaced to a random depth and a random location within its subfield. When the calculated new position of a dot fell outside its subfield, it wrapped to the other side, with a new location given by its speed and direction.

The FOE of the flow field is defined as the location at which the motion on the screen is zero. At each point in time, the FOE defined the velocities present in each subfield (but each subfield's velocity was then perturbed, as explained later). The FOE can be interpreted as the direction of heading in a static environment, or as indicative of an object's motion direction relative to the observer (ignoring eye movements; the current experiments do not attempt to distinguish between these perceptual interpretations). The FOE is given by

$$x^s_{foe} = \frac{\dot{x}}{\dot{z}} \cdot z^s, \qquad [1]$$

where x^s_{foe} is the location of the FOE on the screen, $\dot{x}$ and $\dot{z}$ are velocities of the dot in 3D, and z^s is the distance of projection plane (the screen). In degrees of visual angle, the FOE is described as

$$\tan(\theta_{x,foe}) = \frac{\dot{x}}{\sqrt{\dot{y}^2 + \dot{z}^2}}. \qquad [2]$$

As such, the velocity of the dot cloud directly relates to the location of the FOE on the screen.

The location of the FOE on the screen was not constant, but moved continuously according to a random walk on a damped spring (28). This time-varying walk of the FOE is the key manipulation, serving both as an intuitive way for subjects to engage in a continuous tracking behavior and as a means for temporal decorrelation of the FOE and the behavior.

We chose FOE movement parameters that resulted in Gaussian distributions of the FOE with respect to both the location and velocity. The horizontal and vertical FOE trajectories were independent of one another, each with an SD of the location within a trial of about 7.5° (C: 2.5°) and an SD of the velocity of about $33^\circ \cdot s^{-1}$ (C: $8.3^\circ \cdot s^{-1}$). A small trial by trial offset in the center of the FOE's motion increased the SD of the position to 10°. In addition, the direction of the dot cloud through depth alternated randomly, resulting in expanding or contracting optic flow patterns, with an average switch time of 3.6 s and an SD of 1.3 s.

Because the dots within each subfield were drawn according to the projection of the dots' motion through (virtual) 3D space, the motion within each subfield by itself contained sufficient information about its distance relative to the FOE. To allow for a partial decorrelation of the visual information across the subfields, each subfield either had a small, random perturbation relative to the motion associated with the "true" FOE or was blank (i.e., with no dots rendered within it). Perturbations to the FOE location across subfields were implemented by adding Gaussian noise $N(0, \sigma)$ to the 3D velocity of the dots in a given subfield,

$$x^s_{foe} = \frac{\dot{x} + N(0, \sigma)}{\dot{z}} \cdot z^s. \qquad [3]$$

The random offset of each subfield remained constant for 16 ms to 500 ms, at which point all parameters of the subfield were randomly reassigned. This resulted in an offset of the FOE in the subfields with $\sigma_x = \sigma_y = 2.8^\circ$ (C, 0°; M2, increasing from 0° to 5°). Because this manipulation occurred in the physical (i.e., virtual 3D) velocity space, the effect of the noise in degrees of visual angle depended on the location of the FOE, with larger offsets when the FOE was close to the center compared with the eccentricity. The noise was, however, constant across subfields. Note that, because the information showed in each subfield pointed toward a (slightly) different FOE that was unique from the FOEs shown in all other subfields at that same time, we were able to pick up on these differences to infer which of the subfields' information (and to what amount) was taken into account to direct the gaze.

Core Task. In the simplest conditions, the motions within 45% of all subfields were noisy representations of a single FOE, and the task was to direct the gaze at the FOE. Dots in the remaining 55% of subfields remained invisible. Visibility of subfields as well as their perturbations from the FOE were randomly reassigned every 16 ms to 500 ms. In each trial, dots could either be all black or all white. To maintain interest and solidify flow tracking, for M1 and H, we dispensed rewards periodically whenever the gaze was within 7.5° to 11° of the FOE for typically 1.7 s. The human subject (H) was instructed that a beep was to be interpreted as a reward, and received verbal instructions simply to look at the screen and figure out the task. For M2 and C, the animal was rewarded whenever an accumulated reward from a Gaussian reward field around the FOE exceeded a set threshold. For all subjects, trials ended when the gaze left the screen for more than 250 ms (e.g., a blink) or after 300 s of uninterrupted visual stimulation.

Visual Selection Task. In the visual selection variant of the task, different subfields could show motion for either of two independently moving FOEs, with the dot luminance (black or white) indicating subfields showing information for one FOE or the other. In these cases, each 45% of the subfields contained dots corresponding to one or the other FOE, with the remaining 10% of the subfields empty. Subjects were free to switch between the two FOEs, and were rewarded for tracking either FOE.

Eye Movements. Eye velocities were obtained by discrete differentiation of the unfiltered eye positions. Saccades were detected using a variable velocity criterion in four stages. Whenever the speed exceeded 100°/s of the average speed of the previous 20 samples, a potential saccade was detected. For each saccade, a first estimate of the onset and offset was obtained by finding the last and first samples, respectively, with a speed of 5°/s below that running average speed. The 2D velocity was then projected onto the estimated saccade trajectory for a final estimate of the saccade onset and offset, when the velocity along the trajectory was last below 5°/s of the velocity of the preceding 20 samples and when it first fell below 5°/s relative to the following 20 samples again. Finally, overlapping detected saccades were combined, a minimum saccade duration of 5 samples and a minimal latency of 50 samples between two saccades were assumed, and data around 50 ms of a blink were ignored. For the gaze time series shown in Fig. 2, and for quantifications in Table 1 and Fig. 5*A*, time points 1 s before to 1 s after a blink were removed from analysis; for all other figures and reports, the full time series was considered, although censoring blinks would have only small effects.

Bilinear Regression to Estimate Temporal and Spatial Parameters. We used a bilinear regression model which assumes that the temporal and spatial integration are separable. In the bilinear regression, we first estimate the temporal integration parameters by initially assuming an equal weighing of the information in all subfields, and next use the resulting temporal kernel to aid the extraction of the spatial integration weights. These weights could then be used to iteratively improve the estimates by using the estimate of the spatial weights to better estimate the temporal weights, and so on. While this iterative component is critical for estimates of neural integration, we found it to yield little improvement for the behavioral results focused on here, and thus only iterated once. The bilinear regression allowed us to determine the combination of temporal and spatial weights that best predicted the subject's gaze. To determine the temporal kernel, we regressed the location of the average FOE represented in all visible analyzed subfields for the same underlying FOE over time against the current eye position.

Foveocentric spatial integration parameters were then computed by regressing the subfields' unique FOE time courses to the residuals from the temporal regression. Specifically, we first determined the time course of FOEs that were shown in each subfield relative to the subfield that was fovealized at any given frame. We next filtered each subfield's time course of FOE locations with the causal portion of the temporal kernel ($t \leq 0$), filling times where no dots were shown with the mean FOE location of all visible subfields at that time. We also filtered the mean FOE location of all subfields in the same manner and subtracted this time course from each subfield's filtered time course, and from the gaze (residual error).

To offset scaling effects due to variations in the number of subfields supplying data to the regression, the gaze was multiplied by the filtered sum of subfields with data at each point in time. Finally, both the gaze and the subfield data were multiplied by that same value to perform a weighted regression with more weight for times with more stimulus-carrying subfields. To add robustness against the large autocorrelation of the stimulus, we used ridge regression, which is equivalent to placing a zero-mean Gaussian prior on the regression weights. The ridge parameter was determined using evidence maximization (29, 30). Data from the horizontal (x) and vertical (y) position of the FOE and gaze were fit independently. Unless specified otherwise, all data shown here are from the x component, but the y component was typically comparable.

ACKNOWLEDGMENTS. This work was supported by National Eye Institute Grant R01-EY017366 (to A.C.H. and J.W.P.) and a Postdoctoral Fellowship (to J.K.) from the German Research Foundation.

1. Born RT, Bradley DC (2005) Structure and function of visual area MT. *Annual Rev Neurosci* 28:157–189.
2. Maunsell JH, Newsome WT (1987) Visual processing in monkey extrastriate cortex. *Annual Rev Neurosci* 10:363–401.
3. Gold J, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535–574.
4. Mulligan JB, Stevenson SB, Cormack LK (2013) Reflexive and voluntary control of smooth eye movements. *IS&T/SPIE Electronic Imaging*, eds Rogowitz BE, Pappas TN, de Ridder H (Int Soc Optics Photonics, Bellingham, WA), p 86510Z.
5. Hoffmann KP (1988) Neural basis for changes of the optokinetic reflex in animals and men with strabismus and amblyopia. *Strabismus and Amblyopia* (Palgrave Macmillan UK, London), pp 89–98.
6. Miles FA, Kawano K, Optican LM (1986) Short-latency ocular following responses of monkey. I. Dependence on temporospatial properties of visual input. *J Neurophysiol* 56:1321–1354.
7. Crowell JA, Banks MS (1996) Ideal observer for heading judgments. *Vis Res* 36:471–490.
8. Mitchell JF, Reynolds JH, Miller CT (2014) Active vision in marmosets: A model system for visual neuroscience. *J Neurosci* 34:1183–1194.
9. Bair W, Movshon JA (2004) Adaptive temporal integration of motion in direction-selective neurons in macaque visual cortex. *J Neurosci* 24:7305–7323.
10. Huk AC, Shadlen MN (2005) Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J Neurosci* 25:10420–10436.
11. Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J Vis* 5:376–404.
12. Lappe M, Pekel M, Hoffmann KP (1998) Optokinetic eye movements elicited by radial optic flow in the macaque monkey. *J Neurophysiol* 79:1461–1480.
13. Bonnen K, Burge J, Yates J, Pillow J, Cormack LK (2015) Continuous psychophysics: Target-tracking to measure visual sensitivity. *J Vis* 15:14–14.
14. Page WK, Duffy CJ (2008) Cortical neuronal responses to optic flow are shaped by visual strategies for steering. *Cereb Cortex* 18:727–739.
15. Kountouriotis GK, Mole CD, Merat N, Wilkie RM (2016) The need for speed: Global optic flow speed influences steering. *R Soc Open Sci* 3:160096.
16. Leon PS, Vanzetta I, Masson GS, Perrinet LU (2012) Motion clouds: Model-based stimulus synthesis of natural-like random textures for the study of motion perception. *J Neurophysiol* 107:3217–3226.
17. Spering M, Gegenfurtner KR (2008) Contextual effects on motion perception and smooth pursuit eye movements. *Brain Res* 1225:76–85.
18. Ferrera VP, Lisberger SG (1997) Neuronal responses in visual areas MT and MST during smooth pursuit target selection. *J Neurophysiol* 78:1433–1446.
19. Sahani M, Linden JF (2003) Evidence optimization techniques for estimating stimulus-response functions. *Advances in Neural Information Processing Systems 15*, eds Becker S, Thrun S, Obermayer K (MIT Press, Cambridge, MA), pp 317–324.
20. Nishimoto S, Gallant JL (2011) A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J Neurosci* 31:14551–14564.
21. Dombeck DA, Khabbaz AN, Collman F, Adelman TL, Tank DW (2007) Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* 56: 43–57.
22. Franchak JM, Heeger DJ, Hasson U, Adolph KE (2016) Free viewing gaze behavior in infants and adults. *Infancy Off J Int Soc Infant Stud* 21:262–287.
23. Najemnik J, Geisler WS (2005) Optimal eye movement strategies in visual search. *Nature* 434:387–391.
24. Adams DL, Economides JR, Jocson CM, Horton JC (2007) A biocompatible titanium headpost for stabilizing behaving monkeys. *J Neurophysiol* 98:993–1001.
25. Mitchell JF, Priebe NJ, Miller CT (2015) Motion dependence of smooth pursuit eye movements in the marmoset. *J Neurophysiol* 113:3954–3960.
26. Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in psychtoolbox-3. *Perception* 36:1–16.
27. Eastman KM, Huk AC (2012) PLDAPS: A hardware architecture and software toolbox for neurophysiology requiring complex visual stimuli and online behavioral control. *Front Neuroinf* 6:1.
28. Vul E, Alvarez G, Tenenbaum JB, Black MJ (2009) Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems 22*, eds Bengio Y, Schuurmans D, Lafferty JD, Williams CKI, Culotta A (Curran Associates, Red Hook, NY), pp 1955–1963.
29. Bishop CM (2006) *Pattern Recognition and Machine Learning* (Springer, New York).
30. Park M, Pillow JW (2011) Receptive field inference with localized priors. *PLoS Comput Biol* 7:e1002219.