

Optimal Multimodal Integration in Spatial Localization

Martina Poletti,¹ David C. Burr,^{3,4} and Michele Rucci^{1,2}

¹Department of Psychology and ²Program in Neuroscience, Boston University, Boston, Massachusetts 02215, ³University of Florence, Florence 50125, Italy, and ⁴Institute of Neuroscience, National Research Council, Pisa 56124, Italy

Saccadic eye movements facilitate rapid and efficient exploration of visual scenes, but also pose serious challenges to establishing reliable spatial representations. This process presumably depends on extraretinal information about eye position, but it is still unclear whether afferent or efferent signals are implicated and how these signals are combined with the visual input. Using a novel gaze-contingent search paradigm with highly controlled retinal stimulation, we examined the performance of human observers in locating a previously fixated target after a variable number of saccades, a task that generates contrasting predictions for different updating mechanisms. We show that while localization accuracy is unaffected by saccades, localization precision deteriorates nonlinearly, revealing a statistically optimal combination of retinal and extraretinal signals. These results provide direct evidence for optimal multimodal integration in the updating of spatial representations and elucidate the contributions of corollary discharge signals and eye proprioception.

Introduction

How do we keep track of the locations of objects in a scene? Imagine looking for a specific book in a bookcase. While going through the shelves, you see another potentially interesting book, but keep looking for the desired one until realizing that it is not there. Then, you confidently look back to reach the book you previously spotted. Although for most people this is an effortless operation, it requires complex computations. The projections of stationary objects shift on the retina every time a saccade occurs, and many visual functions progressively decline with eccentricity, making localization of non-fixated objects difficult and often impossible. Thus, spatial localization cannot rely solely on the image currently present on the retina, but needs to be based on a stable internal representation of the scene.

The retinal image is not the only source of spatial information. Important contributions to the updating of spatial representations come from efferent oculomotor signals (von Helmholtz, 1925) (known as corollary discharges; Sperry, 1950), as revealed by experiments in which stimuli are displayed immediately before or during saccades (Hallett and Lighthstone, 1976; Wurtz, 2008). It has long been argued that another extraretinal signal, extraocular muscle proprioception, may also play a key role (Sherrington, 1918). However, experimental evidence on the function of eye proprioception has remained controversial (Donaldson, 2000), leading to the idea that this signal is primarily used for oculomotor calibration and learning rather than for spatial representation (Lewis et al., 2001). Furthermore, it is unclear how

retinal and extraretinal signals are combined to yield the stable and unified representation of space that is necessary for accurately localizing objects across saccades.

Theoretically, the updating of spatial representations could rely on a single source, such as the corollary discharge. However, a more robust and effective strategy would be to use information from all the available sources. In tasks in which multiple sensory cues are available, human observers often follow a strategy of cue integration similar to maximum likelihood (Ernst and Bühlhoff, 2004; Knill and Pouget, 2004), an approach that minimizes the variance of the resulting perceptual estimate (Clark and Yuille, 1990). As exemplified in Figure 1, a similar strategy could in principle be used in spatial localization to optimally combine location estimates obtained in different modalities.

To distinguish among different possible mechanisms of spatial updating, we developed a new procedure of gaze-contingent display, which allowed us to test spatial localization after a variable number of spontaneous, unconstrained saccades. In this task, different mechanisms yield different predictions. If spatial representations are updated solely on the basis of the efferent corollary discharge (effectively a form of “dead reckoning”), each saccade would contribute its own error, so the variance of the localization error should increase with the number of saccades performed after fixation on a given target. On the other hand, if the updating process relies only on the retinal image and/or extraocular muscle proprioception, localization precision should not depend on the number of previous saccades. The statistically optimal integration of all the available sources gives yet a different scenario: it predicts the variance of the localization error to increase with the first few saccades following target fixation and then saturate, as the contribution of the corollary discharge progressively loses reliability and is eventually no longer considered. We show that localization of a previously fixated target is well predicted by the statistically optimal integration of retinal, efferent, and afferent signals.

Received Feb. 1, 2013; revised June 6, 2013; accepted July 16, 2013.

Author contributions: M.P. and M.R. designed research; M.P. performed research; M.P., D.C.B., and M.R. analyzed data; M.P., D.C.B., and M.R. wrote the paper.

This work was supported by National Institutes of Health Grant EY18363 and National Science Foundation Grant BCS 1127216 (M.R.) and European Research Council Grant STANIB (D.C.B.). We thank Harold E. Bedell, Dan Bullock, and Jonathan D. Victor for helpful comments.

Correspondence should be addressed to Michele Rucci, Department of Psychology, 2 Cummings Mall, Boston University, Boston, MA 02215. E-mail: mrucci@bu.edu.

DOI:10.1523/JNEUROSCI.0523-13.2013

Copyright © 2013 the authors 0270-6474/13/3314259-10\$15.00/0

Materials and Methods

Subjects. Four emmetropic subjects (two males and two females), all naive about the purpose of the study, participated in the experiments. All four took part in Experiment 1, and three of them participated in Experiments 2 and 3. Informed consent was obtained from all participants following the procedures approved by the Boston University Charles River Campus Institutional Review Board.

Apparatus. Horizontal and vertical eye positions were sampled at 1 kHz with a Generation 6 DPI eye tracker (Fourward Technologies) and recorded for subsequent analysis. A dental-imprint bite bar and a head-rest kept subjects at a fixed distance from the monitor (126 cm) and prevented head movements. Stimuli were observed monocularly with the right eye, while the left eye was patched. They were displayed on a fast-phosphor CRT monitor (Iiyama HM204DT) with a vertical refresh rate of 200 Hz and spatial resolution of 800×600 pixels. Stimuli were rendered and modified in real time by means of EyeRIS (Santini et al., 2007), a system for gaze-contingent display that enables precise synchronization between eye movements data and the refresh of the image on the monitor, as well as accurate spatial localization of the line of sight. Gaze-contingent procedures were implemented by custom-developed software based on OpenGL and real-time C++ routines running in parallel on the host CPU and on EyeRIS' digital signal processor, respectively.

Stimuli and procedure. Data were collected in separate experimental sessions. Every session started with preliminary setup operations that lasted a few minutes and included positioning the subject optimally and comfortably in the apparatus, tuning the eye tracker, and calibrating EyeRIS to accurately convert the eye position measurements given by the eyetracker into screen coordinates. Subjects were never constrained in the experimental setup for >15 min consecutively.

Accurate localization of the line of sight is important in this study. To this end, before each recording session, we conducted a gaze-contingent calibration, which provides more accurate localization than the standard procedure used in oculomotor research (Poletti et al., 2010). This calibration consisted of two phases. In the first phase, average eye positions were measured as subjects sequentially fixated on each of the dots of a three-by-three grid, as is customary in eye-tracking experiments. The initial mapping from eye position coordinates to degrees of visual angle was determined by bilinear interpolation over the mean eye positions measured at these nine points. In the second phase of the calibration procedure, subjects refined this gaze-to-pixel mapping. They fixated again on each of the points of the grid and, at each point, corrected the estimated location of the center of gaze, which was displayed in real time on the monitor as a retinally-stabilized red cross. Subjects used dedicated controls on the EyeRIS joystick to move the cross on the horizontal and vertical axes until it matched their perceived fixation location. These refinements were then incorporated into the offsets and gains of the bilinear interpolation. This method improved localization of the line of sight by approximately a factor of three on each axis.

Experiments were conducted in complete darkness, and special care was taken in ensuring the absence of spurious light sources. The monitor was set to minimum contrast and brightness settings, and stimuli were displayed for brief periods (50 ms) at low luminance (0.6 cd m^{-2}), conditions that excluded influences from phosphor persistence, as demonstrated in previous experiments (Poletti et al., 2010). Furthermore, to prevent dark adaptation, each trial ended with the presentation of a white noise mask at maximum intensity, and subjects took frequent breaks during which the room was illuminated normally.

Observers were instructed to search for two small ($20'$ radius) red circles, which, they were told, were hidden behind the black background,

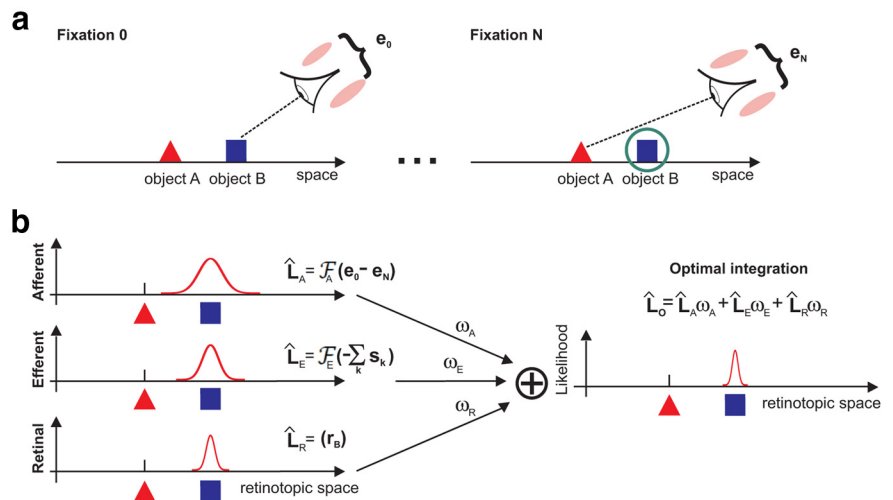


Figure 1. An example of biologically plausible integration of efferent, afferent, and retinal signals. **a**, While looking at object A (red triangle, right), an observer plans a saccade toward object B (blue square), which was at the center of gaze at fixation 0, n saccades before (left). **b**, Likelihoods of independent estimates of object B's location in retinotopic coordinates. The afferent estimate, \hat{L}_A , is proportional to the difference between the current eye position, e_N , and the position e_0 assumed by the eye during fixation on B. The efferent estimate, \hat{L}_E , is proportional to the sum of all the saccades, s_k , which intervened between the two fixations. The retinal estimate, \hat{L}_R , is determined by the position of object B on the retinal image. These three estimates can be combined to maximize the likelihood of localization. F_A and F_E represent mappings into retinotopic coordinates. This scheme is meant to provide an intuitive example of how independent estimates can be obtained and integrated in spatial localization. Several other plausible implementations of this strategy that do not necessarily rely on retinotopic representations are conceivable.

but would briefly appear when fixated. Their task was to report the location of the first object (the target) upon finding the second one (the response cue). In reality, each circle was displayed during a selected fixation, with the two fixations separated by a predetermined number of saccades (1–5, 9, or 10), which varied pseudorandomly across trials. Both circles were displayed at the center of gaze at fixation onset. In Experiments 1 (see Fig. 4) and 3 (see Fig. 7), subjects used a joystick to position a $5'$ cross initially displayed at the response cue location. These two experiments differed only for the presence or absence of a visual reference, a stationary $5'$ white dot at maximum contrast. In Experiment 2 (see Fig. 6), subjects pressed a button when fixating on the remembered target location.

Data analysis. Recorded eye-movement traces were segmented into separate periods of fixations and saccades on the basis of the velocity of the trajectory. Eye movements with minimal amplitude of $3'$ and peak speed higher than $3^\circ/\text{s}$ were selected as possible saccades. Consecutive events closer than 15 ms were then merged together, a method that automatically excluded possible postsaccadic artifacts. Saccade amplitude was defined as the modulus of the vector connecting the two locations at which eye speed became greater (saccade onset) and lower (saccade offset) than $3^\circ/\text{s}$. Trials in which eye tracking was not continuous were discarded. For each trial, we measured the localization error, i.e., the vector difference between the estimated and real positions of the target (see Fig. 4a). For each subject, trials with the same numbers of intervening saccades were pooled together, and the variability of localization error was quantified by means of its dispersion area, defined as the area of the 68th percentile confidence ellipse (Steinman, 1965). This quantity has been used extensively in the literature; since it estimates the area within 1 SD from the mean, it can be regarded as the extension of the concept of variance to two dimensions. For every subject, dispersion areas were estimated over an average of 64 trials for each considered number of intervening saccades. Since for each individual subject all the trials with the same saccades number were collapsed into one single measurement of the dispersion area, the variability of this estimate and within subject statistical comparisons were determined by means of bootstrap (see Figs. 4–7, error bars).

Bootstrap was also used to determine whether measurements after 9 and 10 saccades differed from the predictions of the standard corollary discharge model of error accumulation. At each bootstrap iteration k , we

first obtained an estimate of the dispersion area of each subject, $D_k(n)$, by random sampling with replacement from all of the n saccade trials ($n = 1-3$) available for the considered subject. We then used the bootstrapped areas from all subjects to estimate, by means of least-squares interpolation, the areas $E_k(n)$ predicted by a purely efferent mechanism of spatial updating. Under the assumption of independence of the localization error introduced by each saccade, the corollary discharge model predicts that each saccade enlarges the dispersion area by a fixed amount:

$$E(n) \propto \sigma_O^2 + n\sigma_E^2, \tag{1}$$

where σ_O^2 is an error associated with the reporting method (in Experiment 1, placing a cursor on the display), and σ_E is the corollary discharge error. This method enabled estimation of the distributions of the corollary discharge predictions at $n = 9$ and 10 saccades, which were compared with the dispersion areas measured experimentally by means of two-sample two-tailed t tests. The method was first tested by means of simulations with random data extracted from various distributions, which confirmed the validity of the test and the correctness of the resulting levels of significance.

Since, as shown by our data, the dispersion area increases at a progressively lower rate with saccades, estimation of the corollary discharge model by linear regression of the dispersion areas measured after $n = 1-3$ saccades actually underestimates the error predicted by a mechanism of spatial updating entirely based on corollary discharge. Results did not change if the corollary discharge model was estimated over the first four saccades, rather than the just the first three. These two numbers of saccades, three and four, have been used previously in the literature to measure the influence of extraretinal signals in multiple-step saccade tasks (Bock et al., 1995; Collins, 2010).

The ideal observer curves shown in Figures 4–7 were obtained by least-squares fitting the optimal integration model to the data. This model assumes that the localization error is isotropic and normally distributed. It has three parameters: the SDs of the efferent (σ_E , after one saccade), afferent (σ_A), and retinal signals (σ_R). After n saccades, the model predicts that the localization error will have variance $\sigma_M^2(n)$:

$$\sigma_M^2(n) = \omega_E^2(n) n\sigma_E^2 + \omega_A^2(n) \sigma_A^2 + \omega_R^2(n) \sigma_R^2, \tag{2}$$

where

$$\omega_E = \frac{1/(n\sigma_E^2)}{1/(n\sigma_E^2) + 1/\sigma_A^2 + 1/\sigma_R^2}, \tag{3}$$

$$\omega_A = \frac{1/\sigma_A^2}{1/(n\sigma_E^2) + 1/\sigma_A^2 + 1/\sigma_R^2}, \tag{4}$$

$$\omega_R = \frac{1/\sigma_R^2}{1/(n\sigma_E^2) + 1/\sigma_A^2 + 1/\sigma_R^2}. \tag{5}$$

The predicted dispersion area, the area of the 68th percentile confidence circle, is given by $A(n) = 2\pi \log[(1 - 0.68)^{-1}] \sigma_M^2(n)$. The assumption of radial symmetry is supported by the data in Figure 4, but the model can be easily extended to eliminate this assumption. In complete darkness (Experiment 1 and 2), σ_R was set to infinity, and only σ_A and σ_E were estimated. Identification of these two parameters is possible because the afferent signal gives a constant localization error independent of the number of saccades, whereas the localization error given by the efferent signal increases with the number of saccades. In the presence of a visual reference, σ_R was also estimated. In this case, the model cannot distinguish between σ_R and σ_A , as both terms give a constant localization error across saccades, and we therefore set σ_A to the value estimated previously in Experiment 1.

The optimal integration model was also compared to the corollary discharge linear model that best interpolated all of the available data points ($n = 1-5, 9$, and 10 saccades). Different models were compared by means of a corrected version of the Akaike information criterion (AIC; Akaike, 1974; Hurvich and Tsai, 1989). This criterion determines which model among those compared is more likely to be closer to the true

model in the sense of minimizing Kullback–Leibler discrepancy. ΔAIC_C values reported in the text are averages across subjects.

Our corollary discharge model in Equation 1 assumes that the localization error increases linearly with the number of saccades. However, the predictions of a purely efferent mechanism of spatial updating may deviate from linearity because of two factors: (1) lack of independence in the errors introduced by separate saccades and (2) the finite size of the display. If correlations between successive saccades exist, inaccuracies in the corollary discharge could yield correlations in the localization errors given by separate saccades. In this case, the dispersion area would increase nonlinearly with rate determined by the extent of the correlation. Furthermore, in Experiment 1, subjects could report the position of the target by moving the cursor only up to the edges of the monitor. This limitation in the working area implies that the reported dispersion area cannot grow larger than the surface of the monitor. For this reason, even a perfect corollary discharge mechanism of spatial updating—one in which the internal copy of the saccade is veridical—would eventually deviate from linearity and saturate.

To examine the impact of both factors, for each subject, we conducted Monte Carlo simulations of the individual experimental trials. These simulations estimated the error distribution that may be expected from a linear model of the corollary discharge, given the sequences of saccades performed by the subject in the experiments. We assumed that, on each Cartesian axis k , the internal representation of the saccade displacement \hat{S}_k was linearly related to the actual saccade shift, S_k : $\hat{S}_k = \gamma_k S_k + \eta_k + \varepsilon_k$, where ε_k is a random error with normal distribution $N(0, \sigma_k)$. For each individual subject, γ_k and η_k were estimated on the basis of the localization errors measured in the one-saccade trials, $\mathbf{e}_m = (e_{1,m}, e_{2,m})$, by linear regression of the points $(e_{k,m}, S_{k,m})$, where $S_{k,m}$ represents the k th component ($k = 1, 2$) of the saccade performed in the m th trial. The remaining parameter σ_k and a possible offset at zero saccades, o_k , representing the variability associated with the reporting method (the process of placing the cursor on the display) were estimated by optimally fitting, in the least-squares sense, the SDs of the localization errors measured on the corresponding axis after one to three saccades.

In the Monte Carlo simulations, for each experimental trial (i.e., the combination of target position and recorded saccade sequence), we computed the distribution of the estimated target location resulting from this model and compared it to the experimental data. To take into account the effect of the monitor size, estimated positions outside of the display boundaries were adjusted by taking the closest point on the monitor edges. The results of these simulations are given in Figure 5. They show that correlations across saccades had little effect on the predictions of a corollary discharge model, even when combined with the bounding effect of the limited working area of the display. The data in Figure 5 were obtained by fitting the corollary discharge model in a coordinate system aligned with the saccade (as in Fig. 4*b*). Very similar results were obtained by fitting the model in the original reference system (as in Fig. 4*a*).

Results

Observers were instructed to search for two small circles, which, they were told, were hidden behind the black background but would briefly appear when fixated. Their task was to report the location of the first object (the target) upon finding the second one, either by placing a cursor (Experiment 1) or by looking back at its remembered location (Experiment 2). In reality, the two circles were displayed sequentially at the center of gaze, each during a separate fixation (Fig. 2*a*). By presenting visual stimuli at different locations in space but at the same position on the observer’s retina, this procedure always resulted in the same retinal stimulation independent of the eye movements performed by the observer.

We first examined performance in total darkness, a condition in which spatial localization can only occur on the basis of efferent and/or afferent signals. These two mechanisms yield different predictions: if spatial representations are updated from the efferent corollary discharge, each saccade will contribute its own in-

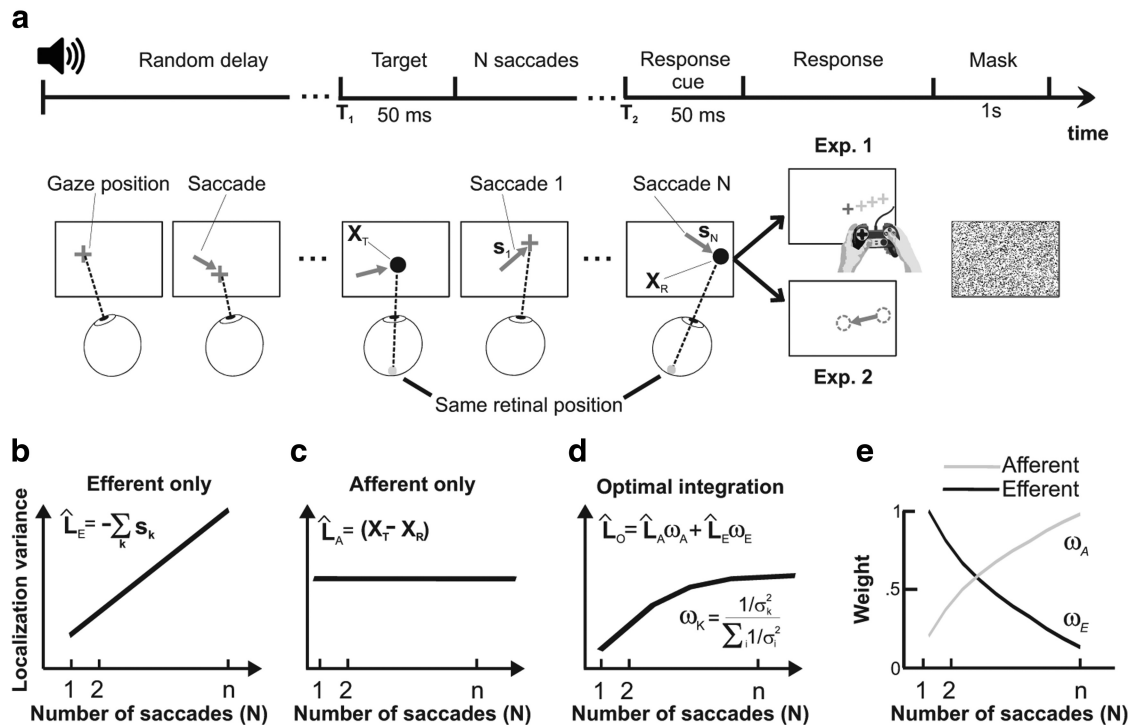


Figure 2. Experimental procedure and theoretical predictions. **a**, Two 20° radius circles, the target and the response cue, were sequentially displayed at the center of gaze after n saccades (s_1 – s_n). Observers were asked to search for the two cues and report the remembered location of the target upon appearance of the response cue, either by placing a cursor (Experiment 1, visual localization) or by looking back (Experiment 2, oculomotor localization). The spatial positions of both the target and the response cue (X_T and X_R) varied across trials depending on the subject's eye movements. **b–d**, Predicted precision of different localization strategies. The variance of the localization error is expected to increase proportionally to the number of saccades between the target and the cue with a purely efferent mechanism of spatial updating (**b**), and to remain constant with a purely afferent one (**c**). **d**, **e**, The optimal integration method is to combine both sources, each weighted inversely to its variance (σ_k^2). **d**, This strategy predicts that the localization error will first increase almost linearly and then saturate as saccades occur, since (**e**) weights are progressively reallocated from the corollary discharge (ω_E) to eye proprioception (ω_A). \hat{L}_E , \hat{L}_A , and \hat{L}_O represent location estimates in a gaze-centered frame of reference.

dependent error, so the variance of the localization error should increase linearly with the number of saccades performed after exposure to the target (Fig. 2*b*). On the other hand, an afferent signal from eye proprioception should be independent of the number of intervening saccades, so, apart from a possible memory decay with time, the resulting variance should remain constant (Fig. 2*c*).

Both strategies are, however, suboptimal. Use of the corollary discharge may work well for a recently fixated target, but after many saccades, the error will eventually become unacceptably large. In contrast, an estimate based on extraocular proprioception may be relatively imprecise (Donaldson, 2000), but will outperform the corollary discharge after a sufficient number of saccades. The probabilistically optimal strategy is to combine both signals based on their precision (Fig. 2*d*), an approach that yields consistent estimates with minimum variance (Clark and Yuille, 1990; Ernst and Bühlhoff, 2004; Knill and Pouget, 2004). This strategy predicts that the variance of the localization error will increase almost linearly with the first few saccades, as it initially relies heavily on the more precise efferent signal, and then progressively saturate as more weight is allocated to the afferent signal (Fig. 2*e*).

Even though observers were asked to search in complete darkness, sequences of saccades were very similar to those measured when the same task was conducted in the presence of a visual reference and, as shown in Figure 3, also similar to those occurring during free viewing of pictures of natural scenes.

Interestingly, localization accuracy was little affected by saccades (Fig. 4*a,b*). For all subjects, on both the horizontal and

vertical axes, the mean localization error after 10 saccades was very close to zero (averages across observers, x -axis, $0.14 \pm 0.97^\circ$; $p = 0.79$; y -axis, $-0.03 \pm 0.39^\circ$) and statistically indistinguishable from the error measured after just one saccade ($p > 0.43$, paired two-tailed t test). However, the variability of the estimated location increased with saccades so that, for all subjects and on both Cartesian axes, variances after 10 saccades were significantly larger than those measured after just one saccade ($p < 0.002$, two-tailed F test of equality of variances). Thus, spatial localization remained accurate on average, but lost precision as the number of saccades increased.

We quantified the precision of localization after a given number of intervening saccades by the dispersion area, defined as the area of the 68th percentile confidence ellipse (the direct extension of the variance to two dimensions; Steinman, 1965). With only a few saccades, the dispersion area appeared to increase almost linearly, a finding that, without additional data points, could have been easily mistaken as evidence for a corollary discharge mechanism of spatial updating. However, it deviated from linearity with additional saccades (Fig. 4*c*; see Fig. 5 for individual subject data). In all subjects, the increment in variability occurring with the first five saccades ($13.8 \pm 1.7 \text{ deg}^2$) was significantly larger than the change caused by the following five saccades ($6.8 \pm 2.3 \text{ deg}^2$; $p < 0.05$, paired two-tailed t test). As a consequence, the dispersion areas after 9 and 10 saccades were significantly smaller than the predictions of the standard corollary discharge model of error accumulation based on the rate of error increment with the first three saccades ($p < 0.001$, two-tailed t test). This deviation from the prediction of constantly increasing variance cannot be

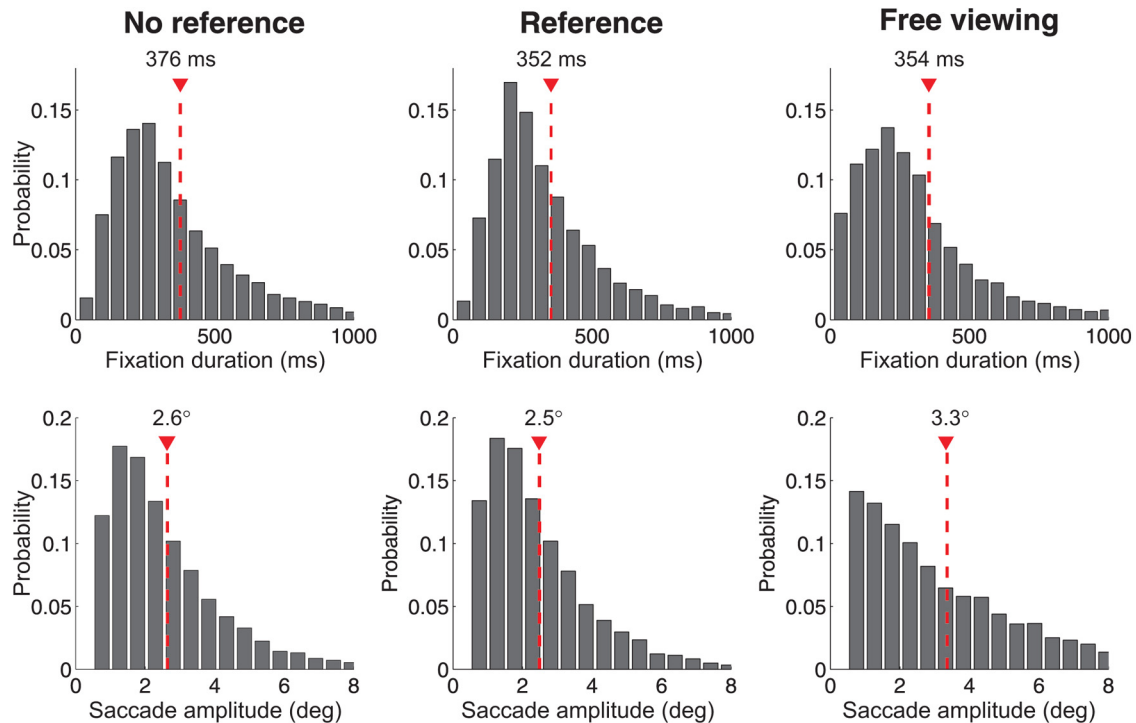


Figure 3. Characteristics of eye movements. Probability distributions of fixation durations (top) and saccade amplitudes (bottom) in the absence (left; data from Experiments 1 and 2 combined) and presence (center) of a visual reference. For comparison, fixation durations and saccade amplitudes measured during normal examination of a scene are also shown (right). In this condition, the same observers freely viewed pictures of natural scenes, each presented for 10 s. Data from all subjects were pooled together. The red line and number in each panel represent the mean of the distribution.

ascribed to differences in scan path for different numbers of saccades. The screen area covered by three saccades was, on average, $56 \pm 5 \text{ deg}^2$, compared with $52 \pm 8 \text{ deg}^2$ for nine saccades. Similarly, there was no difference in the average distance between the target and the response cue, on average $4.9 \pm 0.8^\circ$ and $4.1 \pm 0.6^\circ$ after three and nine saccades, respectively. These results are not compatible with the widely held assumption that spatial representations are updated exclusively on the basis of the corollary discharge.

As shown in Figure 4*b*, the data closely followed the predictions of a strategy of optimal integration of both afferent and efferent signals. The relationship of dispersion area to saccade number was best fit (least-squares) by a maximum likelihood model in which the SD of the corollary discharge (σ_E , the variability after one saccade) was approximately half the SD σ_A of the proprioceptive signal (mean across subjects, $\sigma_E = 1.31 \pm 0.13^\circ$, $\sigma_A = 2.44 \pm 0.79^\circ$). This model explained almost all the variance of the data (adjusted $R^2 = 0.95 \pm 0.03$) and provided a much better fit than the best (least-squares) discharge model of error accumulation estimated on the dispersion areas measured with the first three saccades, a model that yielded a negative coefficient of determination. The optimal integration model was also, on average, 801 times more likely to be close to the true model, in the sense of minimizing the Kullback–Leibler distance, than the standard corollary discharge model used in the literature, the best-fitting first-order linear model of all measured dispersion areas (mean ΔAIC_C , 11.36 ± 3.58). This latter model explained a smaller amount of variance (adjusted $R^2 = 0.73 \pm 0.18$) and predicted an error with no saccades ($8.53 \pm 1.57 \text{ deg}^2$) that was implausibly as large as the error after one saccade.

According to the optimal integration model, eye proprioception contributed $\sim 20\%$ after one single saccade, a value similar to

the estimates obtained at fixation by previous studies (Gauthier et al., 1990; Bridgeman and Stark, 1991). However, the contribution of proprioception increased with saccades so that this signal quickly became the predominant source of information (Fig. 4*d*).

These findings cannot be explained by possible memory decays, which would only further increase localization errors in the trials with larger numbers of saccades: regardless of possible memory contributions, subjects were far more precise than expected in these trials. Results were also not influenced by factors that could make the predictions of a purely efferent model of spatial updating deviate from linearity, such as possible correlations in saccade directions and the limited area of the display (Fig. 5).

If saccades within a trial are correlated, systematic inaccuracies in the efferent signals would cause statistical dependencies in the localization errors resulting from separate saccades. In this case, the model of Equation 1 no longer holds, and the predicted dispersion area would increase nonlinearly with saccades. Furthermore, limitations in the working area of the display—i.e., the fact that subjects could report the position of the target by moving the cursor only up to the edges of the monitor—imply that the dispersion area cannot grow larger than the surface of the monitor, so that even a perfect corollary discharge mechanism of spatial updating will eventually deviate from linearity and saturate. These effects, however, did not alter our conclusions, as demonstrated by the results of dedicated Monte Carlo simulations in Figure 5. The estimated dispersion areas at 9 and 10 saccades differed significantly from the predictions of a corollary discharge model of spatial localization even when (a) this model was individually fit for each subject, (b) it was applied to the sequences of eye movements recorded in the experiments, and

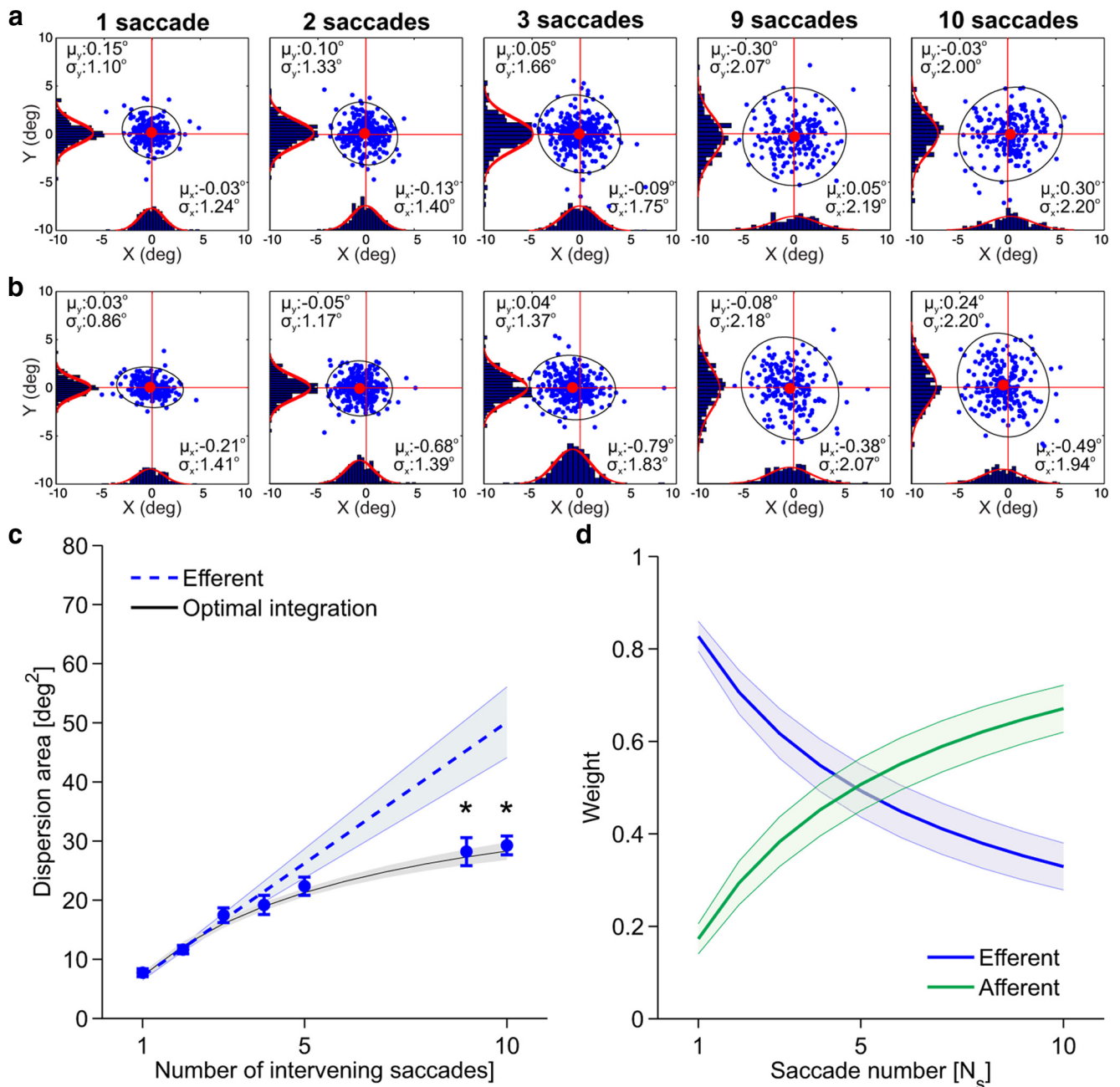


Figure 4. Visual localization (Experiment 1). **a**, Summary of all trials. Each dot represents the localization error in an individual trial. Different panels show trials with different numbers of saccades between the target and the response cue. The mean error (red dot) and the 95% confidence ellipse are shown in each panel together with the marginal probability distributions and their best Gaussian fits, $\mathcal{N}(\mu, \sigma)$ (red curves). Data from all subjects ($N = 4$) were pooled together. **b**, Same data as in **a** after rotating the axes to align the abscissa with the cue-target direction. **c**, Mean dispersion area across subjects as a function of the number of saccades. Asterisks mark significant deviations ($p < 0.001$, two-tailed paired t tests), from the predictions of a purely efferent estimate, as given by the linear regression of the measurements obtained with the first three saccades (blue line). The black curve represents the least-squares fit of the ideal observer model. **d**, Optimal weighting of afferent and efferent estimates. As the number of saccades increases, proprioception is weighted more strongly and eventually becomes the predominant source of information. Error bars and shaded regions in **c** and **d** represent SEM.

(c) localization errors were corrected to take into account the finite size of the display (see Materials and Methods).

Furthermore, very similar results were also obtained when observers reported the target location by looking back at the remembered position rather than by placing a cursor (Experiment 2; Fig. 6), a condition in which the finite size of the monitor was not an issue. These results reveal that the information provided by extraocular muscle proprioception is taken into account in evaluating the location of a previously fixated target and that the

contributions of afferent and efferent signals are weighted in a way that conforms to the rules of optimal cue integration.

The results of Figures 4–6 were obtained in complete darkness. To determine whether a scheme of optimal cue integration continues to hold in the presence of visual landmarks, we repeated Experiment 1 while maintaining a single visual reference (a 5' dot) fixed at the center of the display for the entire duration of the trial (Experiment 3). In this way, the target and the reference were simultaneously visible, and the task could in principle

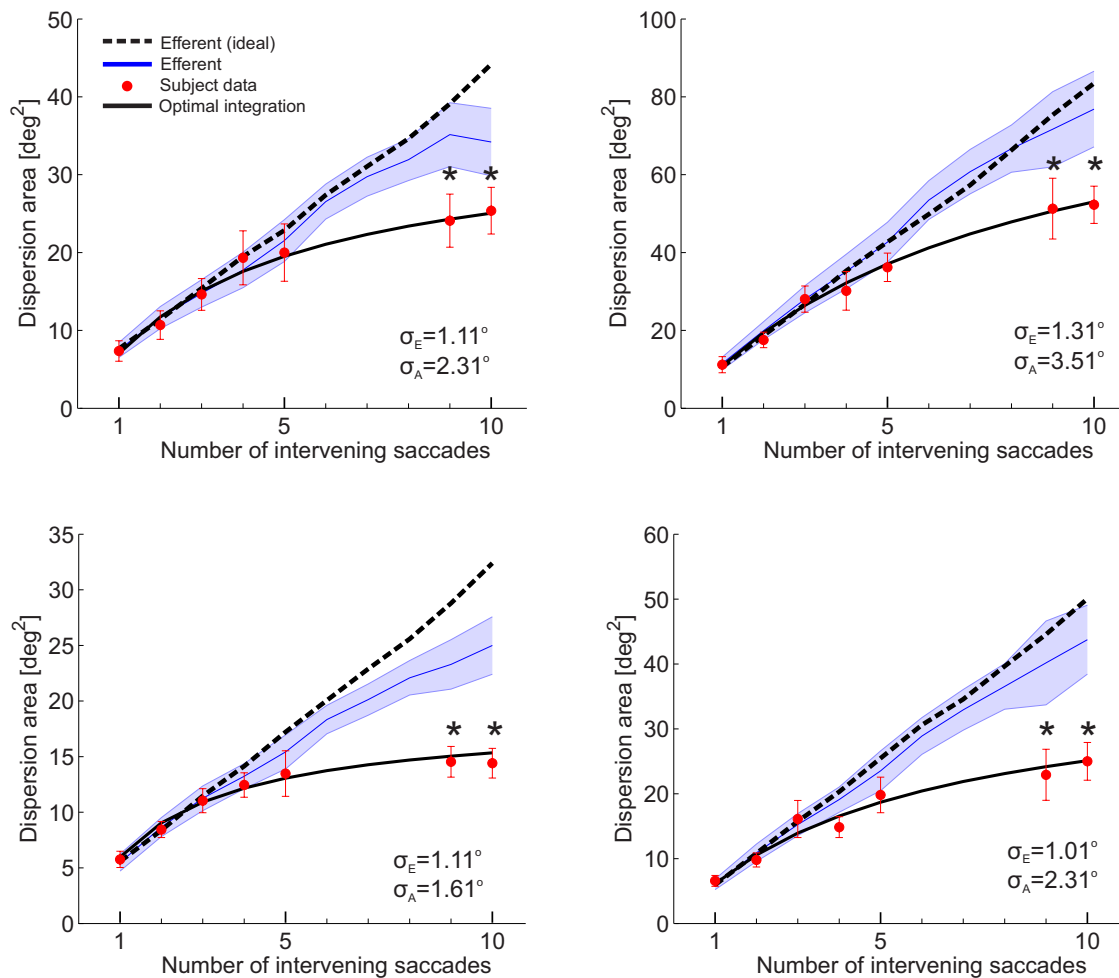


Figure 5. Individual subject data. The dispersion areas measured in the experiments are compared to the predictions of a corollary discharge model of spatial localization adjusted to take into account the finite size of the display and possible biases in the internal representation of saccades (blue line). The adjustment was obtained by means of Monte Carlo simulations of the individual experimental trials, in which the distribution of possible target positions, bounded by the monitor edges, was estimated by applying an individually fitted corollary discharge model to the recorded sequence of eye movements. Asterisks mark statistically significant differences between predicted and measured dispersion areas ($p < 0.02$, two-tailed t test). For comparison, the prediction of the corollary discharge model estimated on a random concatenation of saccades from different trials and without consideration of the monitor boundaries is also shown (dashed line). The least-squares parameters of the optimal integration model are shown in each panel (σ_E and σ_A). Error bars indicate SEM.

be accomplished just on the basis of the visual input, by remembering the target’s location relative to the reference. Like proprioception, this retinal cue does not depend on the number of saccades occurring after presentation of the target.

The visual reference greatly improved the precision of localization (Fig. 7). For each individual subject, the dispersion area was now smaller by a factor of 2 after 1 saccade ($p < 0.05$, paired two-tailed t test) and by a factor of 4 after 10 saccades ($p < 0.001$; confront with Fig. 4c). Yet, apart from this scale change, the shape of the error function remained strikingly similar to that observed in complete darkness. The localization error sharply increased with the first three saccades and settled on a constant value after four saccades, so that the dispersion areas measured after 9 and 10 saccades were significantly different from those predicted by a linear model of error accumulation based on the data points measured with the first three (or four) saccades ($p < 0.04$; two-tailed t test).

As in complete darkness, experimental data followed closely the predictions of an optimal cue integration strategy (Fig. 7a). When efferent and afferent parameters were constrained to be equal to those obtained in Experiment 1, the maximum likelihood model that best fitted the data was 3316 times more likely to be correct than the best-fitting corollary discharge model of error

accumulation (mean ΔAIC_C 15.44 ± 2.55). Furthermore, when both σ_E and σ_R were left free to vary, the optimal integration model yielded an SD of the corollary discharge signal that was very similar to that measured in the previous experiments ($\sigma_E = 1.08 \pm 0.13^\circ$).

The transition from progressive loss of precision with the first few intervening saccades to saturation with the following saccades is incompatible with a mechanism of spatial updating that relies on one single signal, but emerges naturally from the optimal integration of multiple signals. Again, this transition cannot be explained by memory influences. The number of saccades is inevitably correlated with the duration of the trial, which may lead to loss of precision due to memory restraints. However, any memory-driven effects should result in increased variability for larger numbers of saccades, whereas our results show that the rate of increase in dispersion errors decreases with saccade number, contrary to predictions from decay of memory traces. The optimal integration model attributes the initial loss in precision to the decline in reliability of the corollary discharge estimate, which is as precise as the retinal signal after one single saccade ($\sigma_R = 1.27 \pm 0.20^\circ$), but becomes less reliable as more saccades occur. Interestingly, proprioception continued to be considered, and its

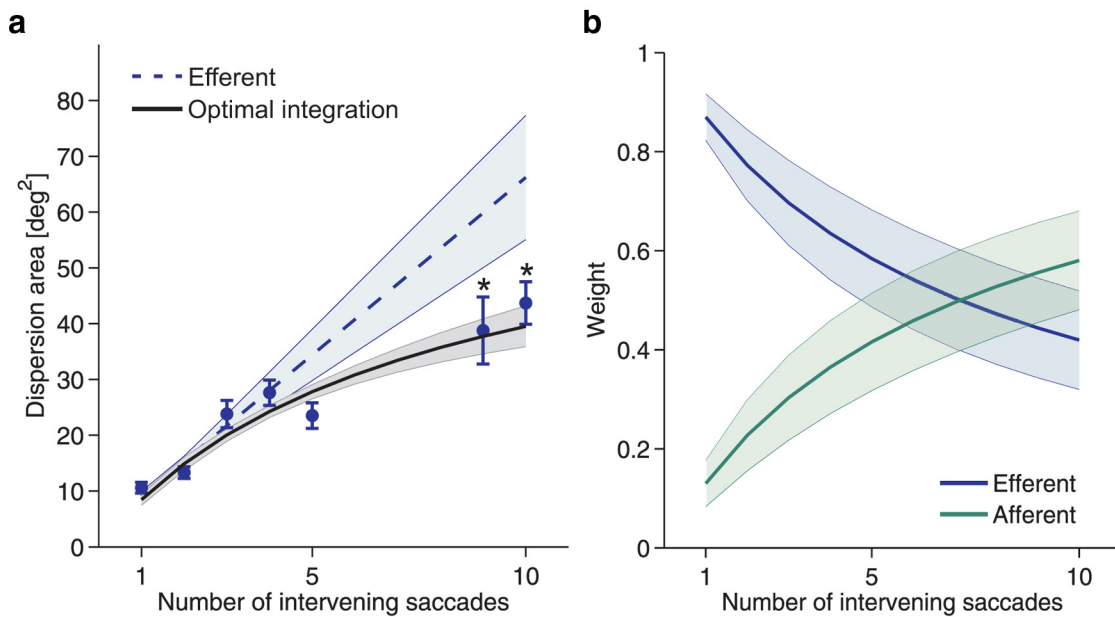


Figure 6. Saccadic localization (Experiment 2). *a*, Mean dispersion area across subjects as a function of the number of saccades. Asterisks mark significant deviations ($p < 0.05$, two-tailed paired t tests) from the predictions of a purely efferent estimate, as given by the linear regression of the measurements obtained with the first three saccades (blue line). The black curve represents the least-squares fit of the ideal observer model. *b*, Optimal weighting of afferent and efferent estimates. Symbols and graphic conventions are the same as in Figure 4.

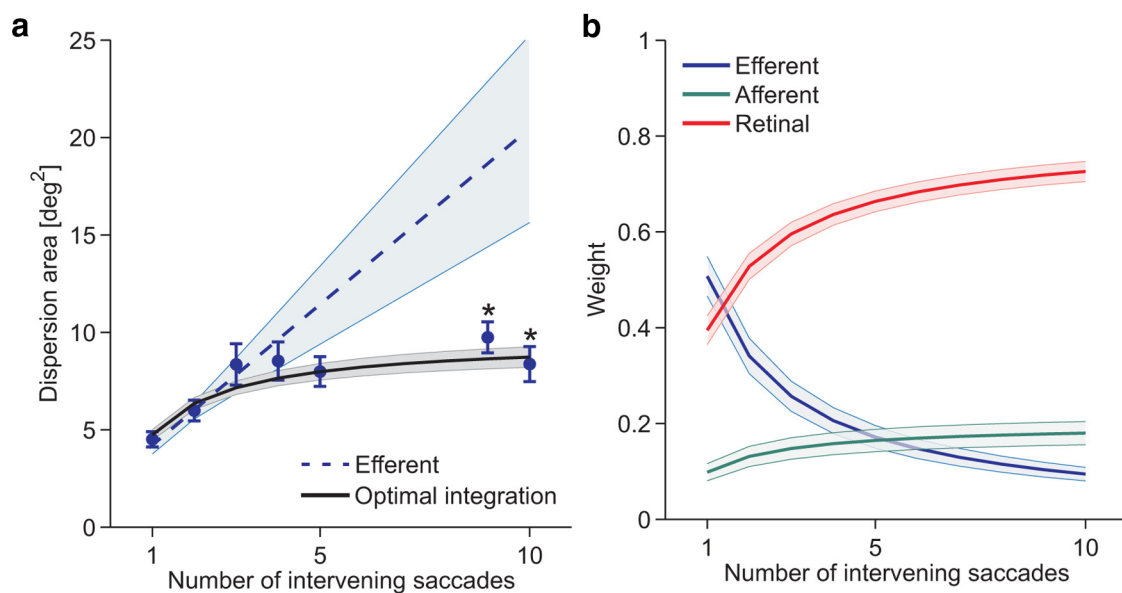


Figure 7. Influence of a visual reference. *a*, Mean dispersion area across subjects as a function of the number of intervening saccades. The black curve represents the least-squares fit of the ideal observer model that integrates afferent, efferent, and retinal signals with the optimal weight combination (*b*). Conditions were identical to those of Experiment 1, except for the presence of a 5' dot at the center of the display throughout each trial. Symbols and graphic conventions are the same as in Figure 4.

contribution grew from $\sim 10\%$ after one saccade to $\sim 20\%$ after 10 saccades (Fig. 7*b*). These data strongly support a scheme of spatial updating based on the optimal integration of different cues, each weighted by its current reliability.

Discussion

Stable spatial representation is essential for visual exploration, motor control, and ultimately for survival. Therefore, the visual system must have developed strategies to optimize the maintenance of stable representations under evolutionary pressure. The results of this study show that the accuracy and precision by which human observers locate a previously fixated target closely

follow the predictions of the optimal integration of efferent, afferent, and retinal signals. These findings are not compatible with an updating mechanism based on a single signal, either the corollary discharge or the retinal input.

In this study, stimuli were modified in real time according to the observer's eye movements. This gaze-contingent procedure provides a departure from the standard methods used to examine the influence of saccades on spatial localization (Hallett and Lighthstone, 1976). Traditional methods only allow analysis of the effects of a few saccades, as their extensions to longer saccadic sequences yield confounding factors in the interpretation of experimental data. For example, the extension of the dual-step sac-

cade task to multiple saccades (Bock et al., 1995; Collins, 2010) implies that both memory requirements (the number of locations the subject needs to remember) and visual input characteristics (the stimulated retinal locations) change with the number of saccades. Our approach circumvents these problems. In our experiments, the memory load was equalized, as subjects only needed to remember one spatial location, no matter how many saccades performed. Furthermore, the spatiotemporal stimulus on the retina was always the same, because all stimuli were displayed at the center of gaze. Since observers performed normal exploratory saccades, our method also eliminated the possibility of planning more than one saccade at the same time (McPeck et al., 2000). To our knowledge, only one previous pioneering study used a somewhat similar approach (Karn et al., 1997), but this study only examined performance after two and five saccades, making it impossible to determine whether localization precision deteriorated in a linear or nonlinear fashion.

The mechanisms responsible for establishing and maintaining stable representations during eye movements have been the subject of fierce unsolved controversy (Sherrington, 1918; von Helmholtz, 1925; for reviews, see Medendorp, 2011; Tatler and Land, 2011; Hamker et al., 2011). Our findings contribute to reconciling two vast bodies of conflicting experimental observations. The first set of data consists of results often taken as evidence that spatial representations are exclusively updated on the basis of the corollary discharge. These observations include (a) the retained capability of compensating for presaccadic ocular displacements after deafferentation of the extraocular muscles (Guthrie et al., 1983), a procedure that is supposed to eliminate eye proprioception; (b) the results of double- and multiple-step saccades (Bock et al., 1995; Collins, 2010) and the corresponding impairments measured after inactivation of neural pathways involved in signaling corollary discharges (Sommer and Wurtz, 2002); and (c) the illusory motion experienced during passive eye rotations (von Helmholtz, 1925) and ocular paralysis (Matin et al., 1982). All of these observations are, however, fully compatible with the optimal integration scheme proposed here, as they were obtained by means of experimental paradigms that only required a few saccades. As shown by our data, the optimal combination strategy weights heavily the corollary discharge under these conditions.

In contrast with this literature, a second set of experimental observations has provided support to a role for proprioception in the updating of spatial representations. These results include accurate localization of targets after very large numbers of saccades (Skavenski and Steinman, 1970), correction for passive displacement of the line of sight in darkness (Skavenski, 1972), as well as the localization errors occurring after passive eye rotations (Gauthier et al., 1990; Bridgeman and Stark, 1991) or following alterations of proprioceptive signals (Allin et al., 1996; Lennerstrand et al., 1997; Balslev and Miall, 2008). These results are also well explained by a model of optimal integration that uses eye proprioception, in addition to other signals, to localize objects in space. In this regard, it is worth observing that the typical durations of natural fixations are sufficiently long for proprioceptive signals to reach the cortex (Xu et al., 2011). Even though stimuli were displayed very briefly in our experiments, the fixations in which targets appeared were significantly longer than average, allowing sufficient time for proprioception to contribute to the process of spatial updating (Fig. 8).

Our finding that spatial localization follows the predictions of an optimal integration strategy is consistent with previous results (Niemeier et al., 2003; Munuera et al., 2009; Ziesche and Hamker, 2011) and adds to a considerable body of evidence showing that

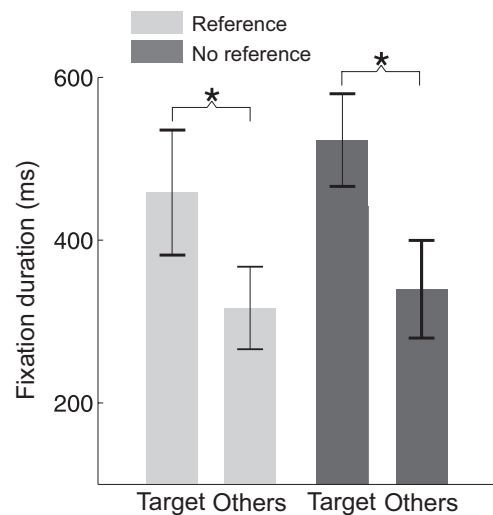


Figure 8. Fixation duration. Comparison between the mean duration of the fixations in which the target was displayed (Target) and the mean duration of the other fixations in the search task (Others). Values represent averages across subjects. Data from Experiments 1 and 2 were combined. * $p < 0.047$ (two-tailed paired t test).

humans tend to perform in a statistically optimal manner in tasks where multiple cues are available (Ernst and Banks, 2002; Stocker and Simoncelli, 2006; Freeman et al., 2010; Geisler, 2011). Several independent signals convey information about the location of a previously fixated object. These signals can be easily transformed into a common coordinate system, such as saccadic vectors or positions in retinotopic or head-centered coordinates. Unlike many tasks, however, the relative precision of these estimates depends not only on the characteristics of each individual signal, but also on the observer's recent behavior, i.e., on the sequence of saccades performed. This important feature distinguishes a maximum likelihood model from previously proposed multimodal models in spatial localization. Previous experiments that attempted to quantify the relative contribution of afferent and efferent signals have yielded different outcomes (Gauthier et al., 1990; Bridgeman and Stark, 1991; Li and Matin, 1992). Our study shows that variable contributions are to be expected depending on the cues available in each specific task and on the observer's oculomotor activity.

Interestingly, in our experiments with a visual reference, observers weighted the efferent signal by approximately the same amount as the retinal signal, a finding that appears to contrast with the general notion that the retinal input prevails over extraretinal contributions (Matin et al., 1982). In these experiments, subjects were presented with an isolated stationary dot for the entire duration of the trial. A stationary dot provides, in principle, an optimal reference for establishing spatial relationships, but also differs greatly from the visually rich scenes that humans normally encounter. Future studies will need to examine the contribution of the retinal input with more natural stimulation and test the specific prediction raised by our model that also this signal is weighted according to its reliability.

Under natural viewing conditions, the need emerges to keep track of multiple locations, each observed with one or more dedicated fixations. Our findings suggest that the visual system updates the representation of each location on the basis of how many saccades have occurred since the corresponding fixation. There are various plausible ways in which this spatial updating scheme may be implemented in neuronal populations. Several

mechanisms have been proposed for statistically optimal combination of neural signals, with supporting evidence (Ma et al., 2008; Fetsch et al., 2011). An interesting possibility, compatible with current theories, is that the same populations of neurons that convey each individual estimate also signal its reliability. For example, the precision of the location estimated by the corollary discharge could be inferred by the spread of activity in a population of neurons with shifting receptive fields. If each remapping operation causes an enlargement of the focus of activity, the spread in the pattern of activity representing each target will be proportional to number of elapsed saccades, therefore effectively signaling the reliability of the estimate. A similar mechanism may also occur in the other modalities. For example, in the presence of a crowded scene, the activity of a visual map signaling the location of a target may be more distributed, therefore implicitly signaling a lower reliability of the cue. Further work is needed to elucidate the neural mechanisms by which the visual system optimally combines afferent, efferent, and retinal signals to efficiently localize objects in space.

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Control* 19:716–723. [CrossRef](#)
- Allin F, Velay JL, Bouquerel A (1996) Shift in saccadic direction induced in humans by proprioceptive manipulation: a comparison between memory-guided and visually guided saccades. *Exp Brain Res* 110:473–481. [Medline](#)
- Balslev D, Miall RC (2008) Eye position representation in human anterior parietal cortex. *J Neurosci* 28:8968–8972. [CrossRef Medline](#)
- Bock O, Goltz H, Bélanger S, Steinbach M (1995) On the role of extraretinal signals for saccade generation. *Exp Brain Res* 104:349–350. [CrossRef Medline](#)
- Bridgeman B, Stark L (1991) Ocular proprioception and efference copy in registering visual direction. *Vision Res* 31:1903–1913. [CrossRef Medline](#)
- Clark JJ, Yuille AL (1990) Data fusion for sensory information processing systems. Boston: Kluwer.
- Collins T (2010) Extraretinal signal metrics in multiple-saccade sequences. *J Vis* 10(14):7 1–14. [CrossRef Medline](#)
- Donaldson I (2000) The functions of the proprioceptors of the eye muscles. *Phil Trans R Soc Lond B Biol Sci* 335:1685–1754. [Medline](#)
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433. [CrossRef Medline](#)
- Ernst MO, Bühlhoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8:162–169. [CrossRef Medline](#)
- Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE (2011) Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci* 15:146–154. [CrossRef Medline](#)
- Freeman TC, Champion RA, Warren PA (2010) A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Curr Biol* 20:757–762. [CrossRef Medline](#)
- Gauthier GM, Nommay D, Vercher JL (1990) The role of ocular muscle proprioception in visual localization of targets. *Science* 249:58–61. [CrossRef Medline](#)
- Geisler WS (2011) Contributions of ideal observer theory to vision research. *Vision Res* 51:771–781. [CrossRef Medline](#)
- Guthrie BL, Porter JD, Sparks DL (1983) Corollary discharge provides accurate eye position information to the oculomotor system. *Science* 221:1193–1195. [CrossRef Medline](#)
- Hallett PE, Lightstone AD (1976) Saccadic eye movements towards stimuli triggered by prior saccades. *Vision Res* 16:99–106. [CrossRef Medline](#)
- Hamker FH, Zirnsak M, Ziesche A, Lappe M (2011) Computational models of spatial updating in peri-saccadic perception. *Phil Trans R Soc Lond B Biol Sci* 366:554–571. [CrossRef](#)
- Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. *Biometrika* 76:297–307. [CrossRef](#)
- Karn KS, Møller P, Hayhoe MM (1997) Reference frames in saccadic targeting. *Exp Brain Res* 115:267–282. [CrossRef Medline](#)
- Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27:712–719. [CrossRef Medline](#)
- Lennerstrand G, Tian S, Han Y (1997) Effects of eye muscle proprioceptive activation on eye position in normal and exotropic subjects. *Graefes Arch Clin Exp Ophthalmol* 235:63–69. [CrossRef Medline](#)
- Lewis RF, Zee DS, Hayman MR, Tamargo RJ (2001) Oculomotor function in the rhesus monkey after deafferentation of the extraocular muscles. *Exp Brain Res* 141:349–358. [CrossRef Medline](#)
- Li W, Matin L (1992) Visual direction is corrected by a hybrid extraretinal eye position signal. *Ann N Y Acad Sci* 656:865–867. [CrossRef](#)
- Ma WJ, Beck JM, Pouget A (2008) Spiking networks for Bayesian inference and choice. *Curr Opin Neurobiol* 18:217–222. [CrossRef Medline](#)
- Matin L, Picoult E, Stevens JK, Edwards MW Jr, Young D, MacArthur R (1982) Oculoparalytic illusion: Visual-field dependent spatial mislocalizations by humans partially paralyzed with curare. *Science* 216:198–201. [CrossRef Medline](#)
- McPeck RM, Skavenski AA, Nakayama K (2000) Concurrent processing of saccades in visual search. *Vision Res* 40:2499–2516. [CrossRef Medline](#)
- Medendorp WP (2011) Spatial constancy mechanisms in motor control. *Phil Trans R Soc Lond B Biol Sci* 366:476–491. [CrossRef](#)
- Munuera J, Morel P, Duhamel JR, Deneve S (2009) Optimal sensorimotor control in eye movement sequences. *J Neurosci* 29:3026–3035. [CrossRef Medline](#)
- Niemeier M, Crawford JD, Tweed DB (2003) Optimal transsaccadic integration explains distorted spatial perception. *Nature* 422:76–80. [CrossRef Medline](#)
- Poletti M, Listorti C, Rucci M (2010) Stability of the visual world during eye drift. *J Neurosci* 30:11143–11150. [CrossRef Medline](#)
- Santini F, Redner G, Iovin R, Rucci M (2007) EyeRIS: a general-purpose system for eye movement contingent display control. *Behav Res Methods* 39:350–364. [CrossRef Medline](#)
- Sherrington CS (1918) Observation on the sensual role of the proprioceptive nerve supply of the extrinsic ocular muscles. *Brain* 41:332–343. [CrossRef](#)
- Skavenski AA (1972) Inflow as a source of extraretinal eye position information. *Vision Res* 12:221–229. [CrossRef Medline](#)
- Skavenski AA, Steinman RM (1970) Control of eye position in the dark. *Vision Res* 10:193–203. [CrossRef Medline](#)
- Sommer MA, Wurtz RH (2002) A pathway in primate brain for internal monitoring of movements. *Science* 296:1480–1482. [CrossRef Medline](#)
- Sperry RW (1950) Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol* 43:482–489. [CrossRef Medline](#)
- Steinman RM (1965) Effect of target size, luminance, and color on monocular fixation. *J Opt Soc Am* 55:1158–1165. [CrossRef](#)
- Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci* 9:578–585. [CrossRef Medline](#)
- Tatler BW, Land MF (2011) Vision and the representation of the surroundings in spatial memory. *Phil Trans R Soc Lond B Biol Sci* 366:596–610. [CrossRef](#)
- von Helmholtz H (1925) *Treatise on physiological optics*. New York: Dover.
- Wurtz RH (2008) Neuronal mechanisms of visual stability. *Vision Res* 48:2070–2089. [CrossRef Medline](#)
- Xu Y, Wang X, Peck C, Goldberg ME (2011) The time course of the tonic oculomotor proprioceptive signal in area 3a of somatosensory cortex. *J Neurophysiol* 106:71–77. [CrossRef Medline](#)
- Ziesche A, Hamker FH (2011) A computational model for the influence of corollary discharge and proprioception on the perisaccadic mislocalization of briefly presented stimuli in complete darkness. *J Neurosci* 31:17392–17405. [CrossRef Medline](#)