

The brain uses adaptive internal models of scene statistics for sensorimotor estimation and planning

Oh-Sang Kwon¹ and David C. Knill

Center for Visual Science, University of Rochester, Rochester, NY 14627-0270

Edited by Wilson S. Geisler, University of Texas, Austin, TX, and approved January 31, 2013 (received for review August 27, 2012)

Because of uncertainty and noise, the brain should use accurate internal models of the statistics of objects in scenes to interpret sensory signals. Moreover, the brain should adapt its internal models to the statistics within local stimulus contexts. Consider the problem of hitting a baseball. The impoverished nature of the visual information available makes it imperative that batters use knowledge of the temporal statistics and history of previous pitches to accurately estimate pitch speed. Using a laboratory analog of hitting a baseball, we tested the hypothesis that the brain uses adaptive internal models of the statistics of object speeds to plan hand movements to intercept moving objects. We fit Bayesian observer models to subjects' performance to estimate the statistical environments in which subjects' performance would be ideal and compared the estimated statistics with the true statistics of stimuli in an experiment. A first experiment showed that subjects accurately estimated and used the variance of object speeds in a stimulus set to time hitting behavior but also showed serial biases that are suboptimal for stimuli that were uncorrelated over time. A second experiment showed that the strength of the serial biases depended on the temporal correlations within a stimulus set, even when the biases were estimated from uncorrelated stimulus pairs subsampled from the larger set. Taken together, the results show that subjects adapted their internal models of the variance and covariance of object speeds within a stimulus set to plan interceptive movements but retained a bias to positive correlations.

Bayesian inference | motion perception | statistical learning | sequential effects

Many sensorimotor tasks involve active interactions with moving objects and therefore, require accurate estimates of object velocity. Because sensory motion signals are noisy, the visual system should use knowledge of the statistics of object velocities to integrate sensory motion information with the velocity predicted by the history of previously observed stimuli. The fact that stimulus history biases perceptual judgments is well-known. Broadly speaking, it appears in two ways—biases to (or away from) the mean of a stimulus distribution (1–9), referred to as central-tendency biases, and biases to (or away from) recently observed stimuli (10–16), referred to as $n - 1$ biases.

It has been proposed that central-tendency biases reflect the behavior of a perceptual system that optimally integrates noisy sensory measurements with prior knowledge of the distribution of stimulus values to make perceptual judgments (2, 4, 7–9). Evidence for this hypothesis comes from studies showing that central-tendency biases in perceptual judgments increase when the sensory uncertainty of a stimulus increases (4) or the variance of a stimulus set decreases (2, 7–9). It also has been suggested that observers learn the mean and variance of a stimulus distribution in a statistically optimal fashion (8) and continuously update internal estimates of mean and variance (7). However, $n - 1$ biases in perceptual judgments, perhaps because they are clearly suboptimal when stimulus sequences are independent, have typically been regarded as a sensory fusion between consecutive stimuli (16) (for example, as might be caused by short-term adaptation or priming mechanisms). Like central-tendency biases, however, $n - 1$ biases may also be a consequence of statistical inference, albeit

using an incorrect assumption that successive stimuli are correlated with one another.

Although one can always retrofit a statistical model to account for observed perceptual biases (as had been done for speed judgments) (17), the strong prediction of the probabilistic model of human perceptual inference is that observed biases (both to the mean and the preceding stimulus) will adapt to changes in the global statistics within a stimulus set. We tested this prediction for perceptual estimates of object speed used to plan hand movements to intercept a moving object (6, 14–16). Most previous studies of perceptual biases have used some form of perceptual report of the stimulus dimension of interest. Although simple and direct, this approach confounds response biases with perceptual biases, and it is unclear whether the reported biases generalize to the perceptual computations embedded in behavior. We, therefore, measured biases in subjects' estimates of object speed indirectly through their performance on a sensorimotor task requiring subjects to extrapolate the motion of a briefly viewed target object. Subjects viewed a target object moving at constant speed before it disappeared behind an occluder. They were asked to hit the target with their finger when it reached an impact zone located at variable distances from the edge of the occluder (see Fig. 1). We used the timing of subjects' hitting movements to infer the statistical properties of their speed estimates.

Although it complicates the data analysis and modeling, this approach has two advantages over previous studies that used direct perceptual reports of a stimulus parameter. First, it probes the perceptual computations embedded in a natural sensorimotor task. Second, it sidesteps the potential confound between perceptual and response bias that is inherent in studies that use direct perceptual report. In particular, the sensorimotor paradigm allows us to experimentally disassociate the statistics of the response variable (time that the object reaches the impact zone) from the statistics of the signal that we are studying (object speed). By manipulating the statistics of the distance to the target zone, we can keep the statistics of time-to-impact zone constant across

Significance

To test the hypothesis that the brain uses adaptive models of object statistics to interpret sensory information, we measured the statistical models used by subjects to estimate object speed when asked to hit a moving object. Subjects' behavior showed perceptual biases to the mean speed within a stimulus set that accurately adapted to changes in the variance of a stimulus set. More significantly, the results show that the way that stimuli on one trial influence following trials adapts appropriately to changes in trial-to-trial correlations in a stimulus set, although subjects' estimates of correlations retain erroneous positive biases.

Author contributions: O.-S.K. and D.C.K. designed research; O.-S.K. performed research; O.-S.K. and D.C.K. analyzed data; and O.-S.K. and D.C.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: oskwon@cvs.rochester.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1214869110/-DCSupplemental.

experimental conditions in which we vary the statistics of object speeds, ensuring that differences in subjects' behavior across conditions truly reflect perceptual adaptations to the statistics of object speed.

In a first experiment, we tested whether subjects adapted their central-tendency biases to stimulus statistics by testing subjects in stimulus contexts with different speed variances. In a second experiment, we tested whether $n - 1$ biases adapted to changes in the trial-to-trial correlations within a stimulus set. To measure how well subjects adapted to the statistics of the experimental stimulus sets, we fit subjects' data using a Bayesian observer model that incorporated an internal model of the temporal statistics of stimulus speeds. The internal model fit to subjects' data provides a characterization of the environment in which subjects' performance would be optimal. As such, it provides a quantitative characterization of the statistical environment to which subjects' estimations strategies are tuned in any given experimental condition and allows us to measure how well their estimation strategies adapted to the statistics within a stimulus set.

Results

Fig. 1 illustrates subjects' task. A virtual ball moved along a tabletop before disappearing behind an occluder. Subjects were told to hit the ball when it passed under an impact zone painted over the occluder (Fig. 1*B*). If successful, the ball reappeared and exploded. If not, it simply reappeared in its position at the time that the subjects' fingers made contact with the table (providing quantitative error feedback). In a first experiment, the speed of the ball and the position of the impact zone were drawn randomly and independently from fixed probability distributions across trials. We ran three groups of subjects in conditions with low, medium, and high variances in target object speeds. The distribution of distances from the front edge of the occluder to the center of the impact zone was set to correspondingly high, medium, and low variances to equate the variance in the time that it took the ball to reach the center of the impact zone across the three conditions. To maximize the variance of target speeds across trials while maintaining reasonable upper and lower bounds

on the speeds, we used uniform distributions in log space for both velocity and distance (Fig. 1*C*).

Twenty-four subjects (eight subjects in each variance condition) ran in one session of six blocks of 100 trials each. Another set of 24 subjects ran in two sessions on different days to explore possible long-term learning effects. We first analyzed data from only the first session for all 48 subjects. We treated the first two blocks of a session as training trials and discarded those blocks from the analysis, in accord with previous results showing that adaptation to changes in mean occurs very fast, whereas adaptation to changes in variance takes 100–200 trials (8).

Subjects' timing performance could have been influenced by separate biases in estimates of target speed and distance traveled behind the occluder or simple timing biases. To provide an initial insight into which of these biases affects subjects' performance, Fig. 2*A* shows subjects' average timing biases in each of the three experimental conditions as a function of target speed and distance to impact zone. The brightness of each pixel represents the log ratio of subjects' hitting time and the time-to-impact zone computed from the stimulus speed and distance. Dark pixels represent early hits, and light pixels represent late hits. Both axes and time values are given in log units, because the relationship between the time-to-impact zone, speed, and distance is linear in log space ($\log T = \log D - \log V$). The raw data show that subjects' behavior does not reflect a simple timing bias, which would appear as 45° iso-bias lines in Fig. 2 (because stimuli with equal times-to-impact zones are given by lines with a slope = 1 in Fig. 2). Rather, the apparent iso-bias lines appear slanted slightly away from vertical, suggesting that subjects' timing behavior is more heavily influenced by speed biases than distance biases. Moreover, the size of the biases for any particular stimulus speed depends on the distribution of speed in the stimulus. For instance, the biases are stronger for a log speed of 2.3 in the low-speed variance condition (gray scales are darker in that column) than the same speed in the high-speed variance condition.

To quantify the influence of sensory signals on subjects' hitting behavior, we regressed subjects' hitting times against the speed and distance presented in the current trial and previous trials (in log space). The regression equation becomes:

$$L_t i = \left(w_0^d L d_i + \sum_{j=1}^N w_j^d L d_{i-j} \right) - \left(w_0^v L v_i + \sum_{j=1}^N w_j^v L v_{i-j} \right) + k, \quad [1]$$

where i is the trial number, k is a constant bias, and L_t , L_d , and L_v are log-hitting time, log distance, and log speed, respectively. We performed the regression in log space, because the relationship between true time, distance, and speed is linear in log space. By doing so, we implicitly assumed that perceptual uncertainty in speed and distance estimation as well as production noise in motor timing are constant in log space. This assumption is consistent with psychophysics showing that variance in these perceptual and motor quantities at least approximately follows Weber's law (18–20). Fig. 2*B* shows the results of the regression analysis. A cross-validation test (*SI Text*, section 1) showed that including only the immediately preceding target speed in the regression and none of the preceding target distances provided the best fit to subjects' data. Thus, subjects' behavior is influenced by the speed of the target on the previous trial but not the distance to the impact zone on the previous trial.

What is immediately apparent from Fig. 2*B* is that the weights to the speed and distance terms differ markedly from each other; in particular, speed biases are significantly larger than distance biases. To visualize the biases more directly, we can reexpress the regression equation to include an explicit term for subjects' biases,

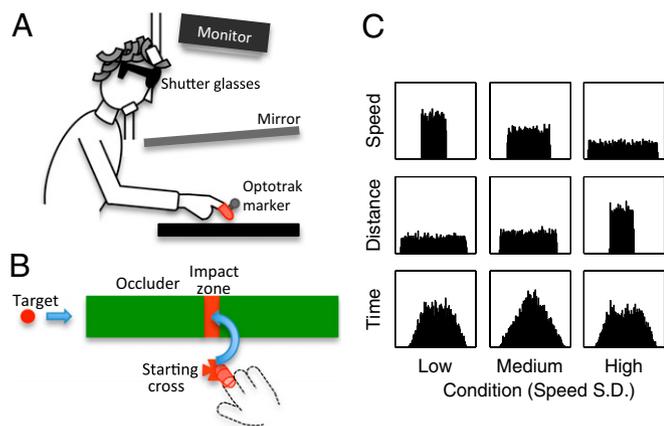


Fig. 1. The schematic illustration of experimental setting (A) and the arrangement of target, occluder, impact zone, and starting cross in the working space (B). The starting position of the target was chosen so that the time that it was visible before disappearing behind the occluder was 400–600 ms. A hit was recorded if a subject's finger touched the red impact zone at any position touching on the hidden target. (C) Histograms of target speed, distance to impact zone, and time to impact zone in each of the three conditions. The variances of time to impact zone were equivalent across conditions in log space and had exactly equivalent distributions in the low- and high-speed variance conditions.

across variance conditions is inconsistent with the behavior of a statistically optimal estimator of distance, because even if we treat the results of the t test as significant, it decreases with decreasing variance in stimulus distances. This inconsistency suggests that the biases in speed and distance derive from qualitatively different processes, a point that we return to in *Discussion*.

The results provide seemingly mixed evidence for the hypothesis that subjects' use an adaptive internal model of stimulus statistics to estimate target speed in this task. On the plus side, central-tendency biases decrease as speed variance increases within a stimulus set, as predicted. On the negative side, the observed $n - 1$ biases do not match the statistics of the stimuli in the experiment, which were uncorrelated from trial to trial. We hypothesized that subjects' $n - 1$ biases do, in fact, result from an adaptive statistical estimation process but one that is biased to a prior belief that stimuli are correlated from trial to trial. The strong prediction of the adaptive estimation hypothesis is that, even if their behavior is suboptimal in an absolute sense, subjects will adapt their $n - 1$ biases to the correlations within a stimulus set. Experiment 2 tested this prediction.

In experiment 2, one group of subjects was tested using target speeds with a strong positive trial-to-trial correlation ($\rho = 0.6$ —positive-correlation condition), and another group was tested using target speeds with a strong negative trial-to-trial correlation ($\rho = 0.6$ —negative-correlation condition). Target speeds were generated as samples from a stationary random walk (a discrete form of Ornstein–Uhlenbeck process) given by the state update equation

$$Lv_{i+1} = \mu_{Lv} + \alpha_v(Lv_i - \mu_{Lv}) + \sqrt{1 - \alpha_v^2} \omega_{Lv}, \quad [3]$$

where μ_{Lv} is the mean speed of targets, α_v is set equal to the desired correlation in target speeds from trial to trial, and ω_{Lv} is a white noise process with variance that is equal to the variance of target speeds. The variance of target speeds was equated across conditions. Fig. 3*A* shows example time courses of target speeds in each of the conditions. In both conditions, the same process was used to generate occluder distances but with trial-to-trial correlations set opposite to the correlations of target speeds. This manipulation guaranteed that stimuli contained no trial-to-trial correlations in the predicted times to impact zone for the targets.

To test whether subjects adapted to the different correlations in the two conditions, we applied the same regression analysis as in experiment 1 to the data from experiment 2. Fig. 4 shows the regression weights for the speed terms. The results for distance were similar to experiment 1 (little bias to the mean and no bias

to the previous stimulus). To avoid possible sampling artifacts created by the fact that the difference between successive stimulus speeds was very different in the two conditions, we used a Monte Carlo subsampling technique (*Methods* and Fig. 3*B*) to select successive pairs of trials for which the joint statistics (both in speed and distance) were the same in the two correlation conditions. For the subsampled pairs of trials used in this analysis, the mean and variance of trial-to-trial differences in both speed and distance were the same in both stimulus conditions. Similarly, the pairs of trials used in the analysis contained no correlations between target speeds or distances to the impact zone. Subsampling eliminated 60% of trials from the analysis. As shown in Fig. 4, the results of the regression analysis did not change when only the subsampled trial pairs were used. The regression applied to the subsampled data shows that subjects' biases to the mean speed did not change across correlation conditions [$F(1,29) = 0.30, P = 0.588$], but subjects' biases to the previous target speed were significantly lower in the negative-correlation group than the positive-correlation group [$F(1,29) = 5.31, P = 0.029$]. Results confirm the prediction of our hypothesis that subjects adapt their $n - 1$ biases to the correlations within a stimulus set.

Modeling

To understand the computational basis for subjects' performance, we fit an observer–actor model that combined a Bayesian estimator of target speed with a model that mapped the speed estimate and the occluder distance to a hitting time. The Bayesian estimator model treats subjects as ideal observers for a stimulus environment with well-specified statistics for trial-to-trial variations in target speed and the sensory noise on speed estimates. The parameters of the model fit to each subjects' data specify the statistical environment in which each subject's behavior, at least in regards to speed estimates, would be optimal. Unlike previous approaches using Bayesian modeling, the goal of our analysis is not to test whether subjects are Bayesian but rather, to assume that they are (a weak assumption, because Bayesian models are fairly general) and use modeling analysis to understand what kind of statistical model that subjects use to estimate speed, test whether subjects adapt appropriately to the statistics of stimuli, and elucidate where their performance is maladaptive.

Although we are most interested in the observer model for speed estimation, fitting that model to subjects' data requires the second component of the full observer–actor model—how speed estimates and occluder distance map to hitting time. Existing ev-

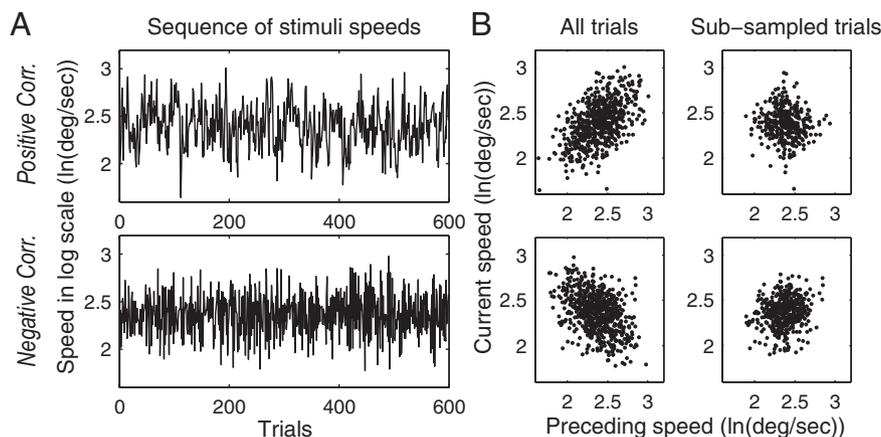


Fig. 3. Speeds of stimuli for experiment 2. (*A*) Example sequences of target speeds in each of the conditions. (*B*) Scatter plots of stimulus speeds for pairs of temporally adjacent trials (in the first case, including all pairs and in the second case, including pairs that had equivalent joint statistics in the two correlation conditions).

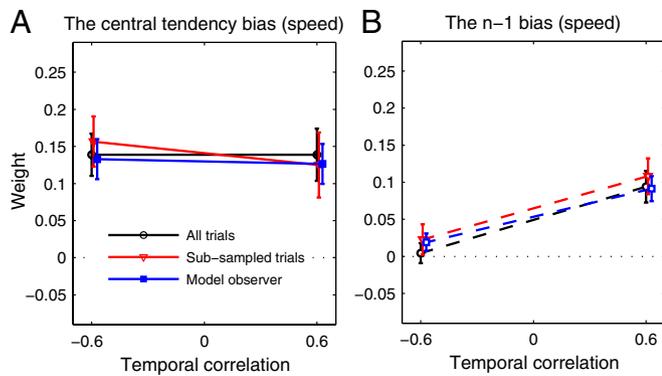


Fig. 4. Results of experiment 2. (A) Subjects' biases to the mean speed (derived by fitting Eq. 2 to subjects' data). The black line shows the regression results using all trials, the red line shows the regression results using subsampled pairs of successive trials, and the blue line shows the regression results derived by simulating the best-fitting Bayesian models for each subject on the task. (B) Subjects' biases to the target speed on the previous trial (same color coding).

idence suggests that, when performing motion prediction tasks like the one used here, humans attentively track objects after they disappear behind an occluder (21). A tracking model would account for the observed differences between the speed and distance bias terms derived from the regression analysis. Unfortunately, fitting a tracking model to subjects' data is computationally intractable. We, therefore, fit a simplified model to subjects' data that is based on the expected behavior of a suboptimal generalization of a perfect, noiseless tracking system.

A perfect tracker will hit the impact zone at the time that its internal estimate of target position reaches the center of the impact zone. Assuming a noiseless tracker, the equation for hitting time is given in log space by

$$L\hat{t}_i = Ld_i - L\hat{v}_i, \quad [4]$$

where $L\hat{t}_i$ is the log-hitting time on trial i , Ld_i is the true occluder log distance on that trial, and $L\hat{v}_i$ is the estimate of log-target speed derived from sensory speed signals on that trial and previous trials. Because it is unrealistic to assume that subjects are perfect trackers, we fit subjects' data using a generalized linear model that assumes that the tracking system induces both biases and noise in the timing of subjects' hitting movements. The generalized model is given by

$$L\hat{t}_i = w_d Ld_i + k - L\hat{v}_i + \eta_i, \quad [5]$$

where w_d and k are multiplicative and additive bias terms (in log space) and η_i is additive white Gaussian noise. Modeling the response (timing) noise as additive in log space is done here for computational convenience; however, in *SI Text, section 3*, we show that a linear tracker that integrates a noise-corrupted internal estimate of target speed over time produces hitting times with biases and noise characteristics that are well-approximated by this model.

We model subjects' estimates of log speed as the outputs of a Bayesian estimator that computes the mean of a posterior probability density $p(Lv_i | Lv_i^{sense}, Lv_{i-1}^{sense}, \dots)$ computed using a particular statistical model of target speeds. The model operates in log space and assumes that target speed variations from trial to trial are generated by the same model used to generate the stimuli in experiment 2—a discrete form of an Ornstein–Uhlenbeck process (Eq. 3). Unlike a simple random walk model, the Ornstein–Uhlenbeck process has the desirable property that it is stationary with finite variance (given by the variance of the noise process

ω_{Lv}). Sensory observations are modeled as noisy versions of the true target speed on each trial

$$Lv_i^{sense} = Lv_i + \eta_{Lv}, \quad [6]$$

where η_{Lv} is constant-variance Gaussian white noise. The sensory noise model, because it is constant variance in log space, displays Weber law behavior in speed discrimination, which has been shown experimentally (18).

The optimal estimator for target speed is given by the standard Kalman filter equations. The posterior mean speed for a given trial is given by a weighted sum of the mean speed in the stimulus set, the estimate of speed from the previous trial, and the noisy sensory speed signal measured on that trial:

$$L\hat{v}_i = w_\mu \mu_{Lv} + w_1 L\hat{v}_{i-1} + (1 - w_\mu - w_1) Lv_i^{sense}. \quad [7]$$

The weights in the update equations are determined by the variance of sensory noise, the assumed variance in the stimulus set, and the assumed trial-to-trial correlation between stimulus speeds (α_v). Given the long delays between the beginnings of each trial (~ 5 s), it is likely that the internal estimate of the previous target speed used to estimate the target speed on any given trial is corrupted by memory noise. We, therefore, included in the model a term for internal memory or state noise. The resulting recursive estimator equations are given by

$$\begin{aligned} L\hat{v}_i &= w_\mu \mu_{Lv} + w_1 L\hat{v}_{i-1}^{mem} + (1 - w_\mu - w_1) Lv_i^{sense}, \\ L\hat{v}_i^{mem} &= L\hat{v}_i + \omega_{mem}, \end{aligned} \quad [8]$$

where ω_{mem} is the internal Gaussian memory noise source. The optimal weights for an estimator with memory noise in the internal state estimates are given in *SI Text, section 2*. Eq. 8, combined with Eq. 5, describes the full observer–actor model for hitting times on each trial.

We used a hierarchical Bayesian model to estimate population means of the observer–actor model parameters that best characterize each subject's performance. The model assumes that the observer–actor parameters characterizing each subject are drawn from population distributions with unknown parameters (e.g., means and SDs). We used Markov Chain Monte Carlo technique to compute the posterior probability density functions on the observer–actor model parameters that characterize individual subjects' performance (*Methods*). The means of these posteriors can be thought of as the best-fitting observer–actor model parameters to each subject's data. We also used the results of Markov Chain Monte Carlo sampling to compute posterior probability density functions for the population means of observer–actor parameters. The means of these posteriors serve as best estimates of the population means of observer–actor parameters characterizing all potential subjects' performance in the task, whereas the ranges of values that contain 95% of the posterior density around the mean provide uncertainty bounds on these estimates, referred to in Bayesian data analysis as credible intervals. Like confidence intervals in classical statistics, credible intervals form the basis for testing for differences between populations in Bayesian data analysis.

We simulated the performance of observer–actor models fit to each subject's data in experiment 1 (using the actual sequence of stimulus values presented to each subject) and performed the same regression analysis on the model observers' timing behavior in the task. Fig. 2C plots the model observers' weights for comparison with subjects' weights. As can be seen in Fig. 2C, the model observers' regression weights fit subjects quite well. The model also captured the variability in subjects' performance. Fig. 5A shows scatter plots of the hitting time on each trial as a function of the stimulus-specified time to impact zone for both model and human observers. Fig. 5B plots the SD of the residual errors for the model

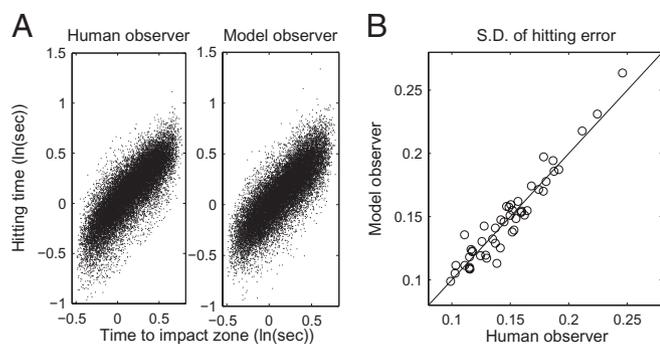


Fig. 5. Variability in the model and human observers' performance. (A) Scatter plots of the hitting time on each trial as a function of the stimulus-specified time to impact zone for both model and human observers. (B) SD of the timing errors for the model observers against the SD of the timing errors for the matched human subjects.

observers against the SD of the residual errors for the matched human subjects.

The best-fitting sensory noise SD was 0.14 in log units (95% credible interval = 0.12, 0.17), equivalent to the same Weber fraction. This value is higher than most psychophysical estimates of the Weber fraction for speed discrimination derived from experiments using fixation (between 0.05 and 0.1 for similar stimulus conditions) (18); however, it is very close to the Weber fraction estimated from one study (22) that had subjects track moving targets (mean Weber fraction = 0.15), which subjects are likely to have done in the current experiment. The SD of memory noise fit to subjects' data was significantly greater than zero (95% credible interval = 0.06, 0.16), justifying its inclusion in the model.

Fig. 6 shows estimates of the population means for the statistical model parameters effectively used by subjects to estimate target speed, with 95% credible intervals for all parameters. The results are consistent with what would be, on average, optimal adaptation to the variance of target speeds in a stimulus set (the true speed variances are within the 95% credible intervals for the population mean of that parameter fit to subjects' data) (Fig. 6A). However, the population mean for the assumed trial-to-trial correlation in target speeds was significantly greater than zero (95% credible interval for $\mu_\alpha = 0.40, 0.54$) (Fig. 6B). Thus, on average, subjects performed near-optimally with regard to their internalized models of speed variance but suboptimally with regard to their internalized models of trial-to-trial speed correlations.

Twenty-four subjects ran in the experiment for 2 d (eight subjects in each variance condition). To examine the effect of learning over the two sessions, we calculated the population mean of subjects' fitted observer-actor model parameters separately for the two sessions and estimated the population mean changes in the variance and correlation parameters for subjects from session 1 to session 2. As can be seen in Fig. 6C and D, the changes in subjects' internal models of speed variance were in the right direction (compare with the average changes needed to match the true speed variance in each condition) but not significantly different from zero (95% credible interval = $-0.19, 0.06$). The changes in subjects' internal models of trial-to-trial correlations were also in the right direction (decreasing from session 1 to session 2) and marginally significant (95% credible interval = $-0.31, 0.02$).

We applied the same analysis to the data in experiment 2. First, we simulated the performance of observer-actor models fit to each subject's data. Fig. 4 plots the model observers' weights for comparison with subjects' weights. The model observers' regression weights fit subjects quite well.

Fig. 7 shows the results of fitting the Bayesian model to subjects' data in experiment 2. Although subjects' internal models of speed variance were not significantly different between groups

(95% credible interval on the difference in speed variance = $-2.90, 0.73$), their internal models of trial-to-trial correlations were significantly different (95% credible interval on the difference in speed trial-to-trial correlation = $-0.68, -0.16$). Subjects' internal estimates of trial-to-trial correlations were lower in the negative-correlation condition than the positive-correlation condition, showing that subjects adapted their internal model of stimulus correlations in the right direction to match the stimulus statistics. The model estimates of subjects' Weber fractions for speed and their internal memory noise were very similar to the parameters estimated in experiment 1. The Weber fraction for speed fit to subjects' data in experiment 2 was 0.14 (95% credible interval = 0.11, 0.16), and the SD of the memory noise on internal estimates of speed was 0.16 (95% credible interval = 0.08, 0.20).

Discussion

The Bayesian model fit to subjects' data shows that their performance is consistent with the true statistics of target speeds in two aspects. First, the close match between the true variances within a stimulus set and the fitted models' estimates of speed variance shows that subjects internalize accurate estimates of speed variance. Second, the fitted levels of noise on sensory

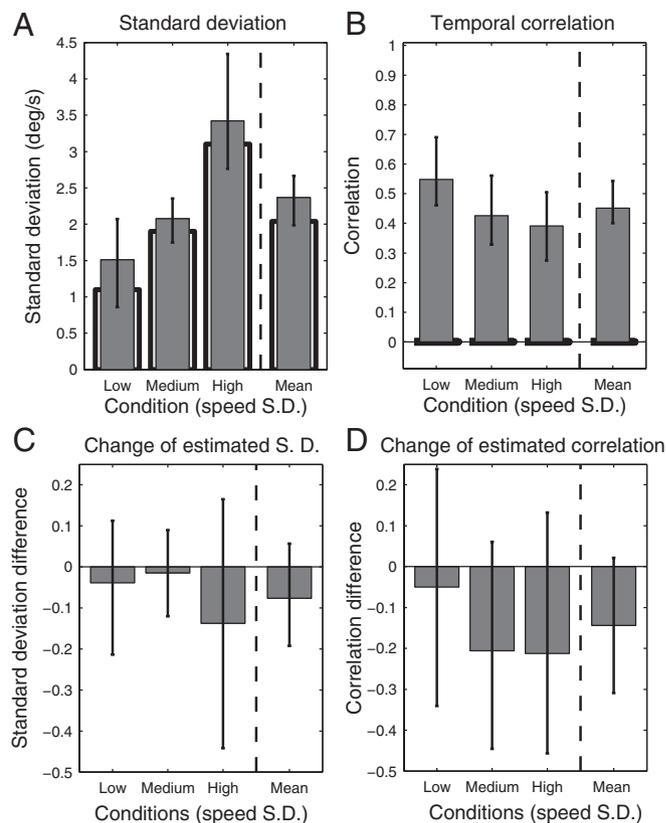


Fig. 6. Best estimates of population mean parameters for experiment 1. (A) Best estimates of the population means of speed SDs assumed by subjects in each of the three variance conditions in the experiment with 95% credible intervals (using data from the first session only). The light bars show the true SDs of speeds. The rightmost bar shows the best estimate of the population means averaged over the three variance conditions. The true average SD is near the boundary of the 95% credible interval, suggesting that subjects behave as if slightly overestimating the speed variance. (B) Best estimates of the population means of the trial-to-trial correlations in each of the three variance conditions, with the rightmost bar showing the average over the three conditions (with 95% credible intervals). (C) Best estimates of the population mean changes in observers' internal estimates of speed SD from session 1 to session 2. (D) Same for observers' internal estimates of intertrial correlations.

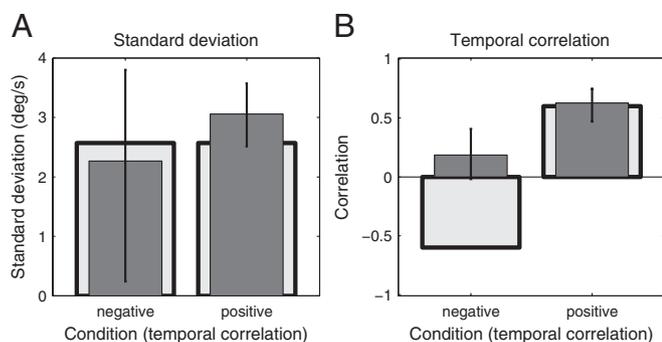


Fig. 7. Best estimates of population mean parameters for experiment 2. (A) Best estimates of the population means of speed SDs assumed by subjects in each of the two correlation conditions. (B) Best estimates of the population means of the trial-to-trial correlations. Error bars reflect 95% credible intervals on the estimates and light gray bars show the true stimulus parameters.

speed signals closely approximate the Weber fractions measured in an experiment using similar viewing conditions (in which subjects were asked to track targets) (22). This pair of observations lends credence to the hypothesis that subjects use adaptive internal models of speed statistics to estimate object speed to plan motor actions.

The finding that subjects accurately adapt their central-tendency biases to the statistics of the stimulus set mirrors the results of several recent studies showing similarly optimal central-tendency biases in a number of perceptual and sensorimotor domains—estimating time intervals between sequentially presented stimuli (4) and judging distance and direction traveled from optic flow (23), coincidence timing (9), pointing (7), and more cognitive judgments of hidden target location (8). Thus, the current results add to the growing body of evidence that the brain continuously adapts and uses internal models of the first-order statistics of scene parameters (mean and variance) to estimate those parameters from sensory data.

The current results on central-tendency biases differ from earlier studies in two important ways. First, previous studies used tasks that directly measured subjects' estimates of a stimulus parameter, potentially confounding response biases with perceptual biases. The current study, largely because subjects showed different biases in speed and distance, dissociated biases in perceptual estimates of speed from simple response biases on the time to hit the target. Second, the current study estimated subjects' central-tendency biases in the context of a model that also incorporated $n - 1$ biases, thus separating the effects of the two on subjects' responses. To the extent that subjects may have shown similar $n - 1$ biases in previous studies, the weights given to internal models of stimulus means would have been overestimated. To illustrate this point, consider the extreme case of an observer whose internal statistical model for a stimulus parameter was that it follows a simple random walk from trial to trial. Such an observer would show no central-tendency bias (the variance of a random walk is infinite) but rather, would use an exponential weighting of the previous sensory signals to estimate the value for a current stimulus. The observer, presented with a sequence of finite-variance, independent stimulus values, would seem to be biased to the mean of the stimulus in a regression that did not include the previous stimulus history. Moreover, the apparent central-tendency bias would be larger when sensory noise was increased, which has been shown in a number of the reported studies.

A final consideration that affects the current study as well as previous studies purporting to show optimal adaptation to first-order stimulus statistics is that one cannot estimate from a study like the current one the absolute mean value to which subjects' percepts are biased. Observers can use feedback to correct any differences between the internal standard to which their percep-

tual estimates are biased and the true mean by compensatory biases in their responses. In the current experiment, subjects could have corrected errors in their internal estimates of the mean by appropriately adding a timing bias in their response to minimize error. This consideration holds for any study that measures the mean to which subjects' responses are biased using an experimental paradigm in which error feedback is provided. Thus, the most significant feature of the results of the current and other studies of central-tendency biases is not that subjects' responses are biased to the true mean but that the strength of the bias, as measured by the weight given to the putative internal estimate of mean, changes with either the variance of stimulus values or the noise in sensory signals (or both). In terms of empirical measures, this weight is given by one minus the weight given to the observable stimulus value incorporated into a regression analysis. Having said that, it seems implausible that subjects would adapt their central-tendency biases to match the variance of target speeds within a stimulus set without similarly biasing those estimates to an estimate of the mean within the stimulus set.

The most notable failure of optimality in subjects' behavior is the influence of the previous target speed on subjects' estimates of the current target's speed on each trial. Although suboptimal, subjects' biases do adapt to the correlations in the stimulus sequences presented to them; $n - 1$ biases decrease almost to zero when target speeds are negatively correlated from trial to trial. This result suggests that $n - 1$ biases reflect subjects' estimates of the correlations in stimulus sequences but that subjects' estimates of stimulus correlations are positively biased. The positive-correlation bias observed in subjects' perceptual estimates of target speed is consistent with the literature on temporal dependencies in more cognitive binomial decision-making tasks (24–26). Kareev (24), for example, presented a sequence of binary items (Xs and Os) to subjects and asked them to predict the next item on each trial. As in the current study, subjects' performance was systematically altered by the temporal correlation of the sequences presented; however, subjects' predictions showed a consistent bias to positive correlations.

It is not clear why the human sensorimotor system erroneously assumes positive temporal correlations when none exist in stimulus sequences. One proposal is that it reflects a strong prior on the statistical structure of the world, in which strong positive temporal correlations between events may be ubiquitous (26). The current study has shown that the brain can adapt its internal model of temporal correlations to more closely match the correlations of stimulus sequences in the proximal environment, although only partially on the short time scale used here (1 h).

Another intriguing possibility is that the brain may be able to switch between different models, given cues about the generative process creating the stimuli in the environment. In support of this idea, Green et al. (26) have shown that subjects' betting behaviors on the outcomes of independent sequences of random binary events reflected an assumption of positive correlations when the outcomes were generated by a hidden random process in the environment. When subjects actively generated the outcomes themselves using motor behaviors that necessarily led to independence of outcomes, subjects correctly assumed independence.

An alternative account for the $n - 1$ bias is that it results from a process that continuously adapts its internal estimate of the mean of the stimulus set, assuming that the mean continuously drifts over time and that trial-to-trial deviations from the current mean are independent (7). We will refer to this model as an adaptive mean model. Such an account is similar to Kalman filter models of sensorimotor adaptation, which rely on error feedback from hand movements to adapt an internal estimate of calibration parameters mapping motor commands to the end results of hand movements—typically simple shifts in the mapping (27). In such models, errors in the endpoints of pointing movements are created by motor noise (assumed to be independent from trial to trial) superimposed on a randomly drifting shift in the mapping from efferent signals to

hand endpoints. Sensorimotor adaptation in such models is conceived of as estimating this shift—akin to estimating a drifting mean in the adaptive mean model. Just as strong temporal correlations in the parameters determining sensorimotor mappings induce positive correlations in movement errors (28), strong serial correlations in a drifting mean process would induce positive correlations in estimated target speeds from trial to trial; thus, an adaptive mean model will seem to show $n - 1$ biases, resulting from tracking the drifting mean.

To test the hypothesis that subjects' $n - 1$ biases result from an adaptive mean process, we fit a second-order Bayesian model that assumed that the mean speed follows a simple random walk and that trial-to-trial deviations from the drifting mean are independent. All aspects of the observer-actor model and fitting procedure were the same as used for the correlated speed model, with the exception of the generative model assumed for target speeds. The model had the same number of free parameters as the correlated speed model, with the variance of the random walk on the mean replacing the correlation term in the correlated speed model (details in *SI Text, section 4*). We simulated the best-fitting observer-actor model for each subject on the stimuli used in the experiment and performed the same regression on the model observer's data as we did for the human subjects. The best-fitting adaptive mean model emulated the observed central-tendency bias and the $n - 1$ bias fairly closely; however, Bayesian model comparison revealed that the correlated speed model fits the empirical data better by a large margin in both the first and second experiments (*SI Text, section 4*). The difference in model fits arises from the fact that the adaptive mean model gives more weight to target speeds farther back from the current trial than the correlated speed model. To examine the weights on more than one-back trials, we regressed the hitting time on the current to eight previous speeds and the current distance. The regression was applied to human subject data and the simulated behavior of the best-fitting versions of the two candidate models. As shown in Fig. 8, the weights decrease sharply after the n -first trial for both the human observers and the correlated speed model, whereas they decrease much more slowly for the adaptive mean model. The results of experiment 2 further argue against the hypothesis that the $n - 1$ bias results primarily from an adaptive mean process. For such an account to fit the results of experiment 2, one would have to assume that subjects adapt to the change in stimulus correlations by decreasing the rate with which they adapt their internal mean estimate.

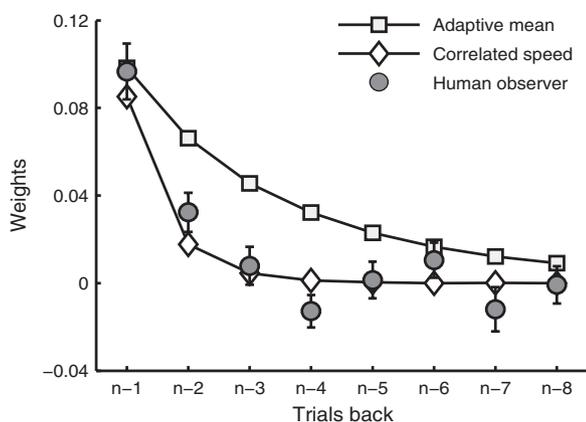


Fig. 8. Weights on the speeds of trials $n - 1$ to $n - 8$. Human observers' weights on the past trials (circles) are consistent with the performance of correlated speed model (diamonds) in that the weights sharply drop after $n - 1$ trial. The adaptive mean model (squares) gives relatively large weights to the past trials. Human observers' weights are computed from the first session of all 48 subjects' data. Error bars represent SEs.

Our data do not permit fine tracking of subjects' learning process in the current experiment from trial to trial, partly because of the large number of free parameters in the regression needed to account for subjects' performance (four). For a simple localization task, a recent study has shown that subjects learn the value of the mean of a prior distribution within 10–20 trials and adapt the central-tendency bias (the weight to the mean) to a nearly asymptotic value within 100–150 trials (8). Assuming similarly fast learning in the current experiment, subjects' performance should have been relatively stable over the four blocks that we used to fit model data (throwing out the first two blocks, or 200 trials, as learning trials). Berniker et al. (8) showed that subjects' performance in both the early trials of an experiment and after a discrete change in variance are well-fit by a Bayesian model that learns the mean and variance from noisy stimuli but assumes that those parameters are fixed in the stimulus set (unlike the adaptive mean model proposed above as a possible account for $n - 1$ biases).

Recent work has suggested that the visual system uses a prior to slow speeds to interpret motion patterns, providing an explanation for a number of illusory 2D motion percepts (29, 30). Stocker and Simoncelli (17) have even measured the shape of this purported prior psychophysically using simple speed discrimination data for patterns with different contrast levels (i.e., different effective encoding noise levels). This work implicitly assumes that the visual system uses a fixed prior for interpreting the speeds of noisy sensory motion signals. It is not clear whether the prior measured by Stocker and Simoncelli (17) and purportedly underpinning our percepts of simple 2D motion patterns reflects neural processing at the same level as the priors used to plan movements in the current task. One might argue that a prior to slow speeds is implemented in the motion-pooling networks in early visual motion processing areas like macaque middle temporal and medial superior temporal areas and that the prior statistical model on object speeds affecting subjects' motor planning is implemented in later stages of neural processing that integrate motion information to make decisions and guide motor behavior. Psychophysical results like the current ones cannot resolve this question; however, it is intriguing to note that, after viewing stimuli with relatively high mean speeds, biases in subjects' percepts of the directions of moving, oriented texture patterns shift away from the direction predicted by a prior peaked at zero to a direction predicted by a prior with a peak at higher speeds (31). This result suggests that even supposedly low-level motion percepts are affected by adaptation to the statistical context of stimuli.

In summary, subjects in our experiments adapted their biases in object speed estimates to the statistics of object speeds within a local temporal context. Central-tendency biases seem to be near optimal, in that they are consistent with an optimal Bayesian observer tuned to the speed variances in a stimulus set and the sensory noise within an observer. Biases to the speed of preceding stimuli are suboptimal but nevertheless, adapt to the correlational structure within a stimulus set. These results support the hypothesis that the brain uses adaptive internal models of scene statistics to estimate object speed from noisy visual motion signals.

Methods

Experiment 1. Subjects. Subjects volunteered to take part in the experiment for payment (\$10/h). They were naïve with respect to the purpose of the experiment and had never participated in similar experiments before. The present study is part of an ongoing project that had been approved by the local ethics committee.

Apparatus. Subjects performed the task in a virtual environment with their head in a head-and-chin rest as illustrated in Fig. 1A. They viewed the experimental environment displayed on a monitor through a mirror. The mirror obscured the table and the moving hand. The distance between the eyes and the table was ~55 cm. A virtual finger was rendered at the position of the subject's active index finger, over which they wore a steel tube with three infrared markers attached to it. The markers were tracked by an

optotrak 3020 system at 120 Hz. Subjects viewed the display stereoscopically through liquid crystal display shutter glasses. A steel plate was on the right side of the table where the impact zone was projected. The plate was connected to a 5-V source, and the steel tube worn over the index finger acted as a ground; therefore, by measuring the voltage of the plate, we acquired a precise measurement of the timing of hitting.

Stimuli. Fig. 1*B* shows the schematic arrangement of target, occluder, impact zone, and starting cross in the working space. The target was a red circle of 1.6 cm (1 cm \sim 1° visual angle) in diameter presented against a black background. The position at which the target first appeared was randomly chosen on each trial so that the visible duration of the target was distributed uniformly between 400 and 600 ms. The occluder was a red rectangle 5.0 cm in height. The impact zone was a green rectangle 5.0 cm in height and 3.2 cm in width. On the right side of the impact zone was another occluder that had the same color and height as the first occluder. The left side of the occluder, where the target disappeared, was fixed, and the center of the impact zone within the red occluder varied randomly from trial to trial following a log-uniform distribution. The starting cross was positioned 6.0 cm below the center of the impact zone on each trial. The starting cross was 3.0 cm in width and height.

Procedure. Each trial started when a subject positioned his or her index finger on the starting cross. At the moment that the index finger contacted the starting cross, the occluder and impact zone were presented; 1,000 ms later, a target appeared on the left side of working space, moved horizontally rightward, and disappeared behind the occluder. Subjects were instructed to hit the target when the target was behind the impact zone. A trial was counted as a hit if the distance between the finger at the time of impact and the center of target was less than 1.2 cm and the finger hit the table within the impact zone. When subjects hit the target, the target visually exploded. When subjects missed the target, the position of target at the moment of impact was displayed. Each experimental session consisted of six blocks of 100 trials each. At the end of each block, subjects were informed of the total number of hits in the block.

Fifty subjects were in the experiment. Two subjects' data were excluded from additional analysis because of the high error rate (more than 2.5 SD above the group mean). Consequently, 16 subjects ran in one of three stimulus conditions (low-, medium-, and high-speed variance conditions). Target speeds had a log-uniform distribution between 9.6 and 13.5 cm/s in the low-speed variance condition, between 8.6 and 15.2 cm/s in the medium-speed variance condition, and between 7.2 and 18.2 cm/s in the high-speed variance condition. Distributions for the distance to the impact zone were also log uniform and set to equate the variance of times to the target zone across conditions. The distributions of distances were 8.7, 19.7; 9.2, 18.6; and 11.2, 15.1 in the low-, medium-, and high-speed variance conditions, respectively.

Eight subjects in each condition ran for one session. The other eight subjects ran for two sessions on separate days.

Experiment 2. The apparatus, stimuli, and procedure used in experiment 2 were identical to the settings of experiment 1. Only the statistics of the target speeds and distances to the impact zone differed. Subjects were run in one of two conditions—a positive-correlation condition and a negative-correlation condition. Target speeds on each trial were generated using a discrete form of an Ornstein–Uhlenbeck process in log space (Eq. 3). The autocorrelation function for the trial-to-trial sequence of target speeds (in log space) is given by $\rho(\tau) = \alpha_v^\tau$, where τ is the delay between trials. In the positive-correlation condition, α_v was set equal to 0.6; in the negative-correlation condition, it was set equal to -0.6 . The SD of log speeds in the stimulus set was 0.23 in both experimental conditions. Distances to the impact zone were generated using the same stochastic process but with α set to have a sign opposite to the value used for target speeds.

Thirty-two subjects (sixteen subjects in each condition) ran in the experiment. One subject's data were excluded from the analysis because of the high error rate (more than 2.5 SD above the group mean).

Modeling. To model subjects' hitting times, we assumed that observers computed a Bayesian estimate of target speed based on noisy sensory signals and an internal model of the statistics of real target speeds. Based on previous studies (21), we assume that subjects attentively track the object behind the occluder based on their initial estimate of target speed. This strategy would require subjects to apply some form of decision rule (unknown to us) to select a time to initiate a hitting movement. Assuming that both the tracking and motor output are noisy, these processes would lead to a hitting time that depends on the initial estimate of target speed and the distance to the impact zone. We assumed that the output of this process can be described by a generalization of an ideal tracker that includes bias terms

on distance and simple additive noise in log space (Eq. 5). In *SI Text*, section 3, we show that a noisy tracker that integrates an internal model of target position and speed and applies a simple decision rule to initiate a noisy hitting movement shows near-equivalent statistics to this model (both in its mean and variance).

The Bayesian speed estimator assumed that the logs-target speeds follow an Ornstein–Uhlenbeck process (Eq. 3) and that the measurement of true speed in each trial was corrupted by additive Gaussian sensory noise (Eq. 6). It was also assumed that, because of the long time delays between trials, the internal estimate of the previous speed used to estimate the current speed would be corrupted by memory noise. Including a memory noise term, the recursive update equations for internal estimates of speed become

$$\begin{aligned} L\hat{v}_{ij} &= L\hat{v}_{ij-1} + K_i(Lv_i^{sense} - L\hat{v}_{ij-1}) \\ L\hat{v}_{i+1j} &= \alpha_v(L\hat{v}_{ij} + \omega_v^{mem} - \mu_{Lv}) + \mu_{Lv}, \end{aligned} \quad [9]$$

where $L\hat{v}_{ij}$ is the estimate of the speed on trial i given all of the sensory measurements up through trial i . $L\hat{v}_{i+1j}$ is the predictive estimate of the speed on the next trial given the sensory information up through trial i . ω_v^{mem} is white Gaussian noise that represents additive memory noise. The choice to model memory noise as additive in the log domain was purely for computational convenience. The Kalman gain (K_i) is calculated on each trial taking into account the uncertainty in $L\hat{v}_{i+1j}$ added by the memory noise (a derivation is in *SI Text*, section 2).

The observer–actor model has four free parameters for the generative model for speed, representing the variance of speed in the stimulus set, the trial-to-trial correlations between target speeds (α_v), the variance of sensory noise, and the variance of memory noise. It also has free parameters for the additive and multiplicative biases and the variance of the additive noise in the mapping from speed estimates to hitting time. To fit the parameters, we need to compute the posterior density function

$$p(\vec{P}|\vec{L}\vec{t}) \propto p(\vec{L}\vec{t}|\vec{P}; L\vec{v}, L\vec{d})p(\vec{P}), \quad [10]$$

where \vec{P} is a vector of model parameters, $\vec{L}\vec{t}$ is a vector containing subject's hitting times in an experimental session (in log space), and $L\vec{v}$ and $L\vec{d}$ are vectors containing the true stimulus velocities and distances in a session in log space.

Because the model is linear and Gaussian, the likelihood function $p(\vec{L}\vec{t}|\vec{P}; L\vec{v}, L\vec{d})$ is a multidimensional Gaussian distribution with a mean vector given by $\vec{\mu}_{L\vec{t}} = W_d L\vec{d} - \vec{\mu}_{L\vec{v}} + k$ and covariance matrix $\Sigma_{L\vec{t}} = \Sigma_{L\vec{v}} + \Sigma_{output}$. $L\vec{d}$ is a vector containing the log of the true occluder distances on each trial, and $\mu_{L\vec{v}}$ is a vector containing the mean estimates of log target speed on each trial conditioned on the true stimulus speeds. $\Sigma_{L\vec{v}}$ is the error covariance of the log-speed estimates, and Σ_{output} is the covariance of the noise in hitting time induced by attentive tracking and the motor response (a diagonal matrix containing the variance of the noise in Eq. 5 in every entry along the diagonal). The model parameters determine both the model's mean log hitting time on each trial, $\mu_{L\vec{t}}$, and the covariance of log hitting times across trials, $\Sigma_{L\vec{t}}$ (both conditioned on the true stimulus conditions).

By merging the two lines of Eq. 9 into one equation and simplifying the notation $L\hat{v}_{ij}$ to $L\hat{v}_i$, we obtain

$$\begin{aligned} L\hat{v}_i &= \alpha_v(L\hat{v}_{i-1} + \omega_v^{mem} - \mu_{Lv}) + \mu_{Lv} \\ &+ K_i(Lv_i^{sense} - \alpha_v(L\hat{v}_{i-1} + \omega_v^{mem} - \mu_{Lv}) + \mu_{Lv}). \end{aligned} \quad [11]$$

This equation can be divided into two independent-state update equations representing a deterministic component ($L\hat{v}_i^{det}$) and a random component ($L\hat{v}_i^{random}$),

$$L\hat{v}_i = L\hat{v}_i^{det} + L\hat{v}_i^{random}, \quad [12]$$

where

$$\begin{aligned} L\hat{v}_i^{det} &= \alpha_v(L\hat{v}_{i-1}^{det} - \mu_{Lv}) + \mu_{Lv} + K_i(Lv_i^{sense} - (\alpha_v(L\hat{v}_{i-1}^{det} - \mu_{Lv}) + \mu_{Lv})) \\ L\hat{v}_i^{random} &= \alpha_v(L\hat{v}_{i-1}^{random} + \omega_i^{mem}) + K_i(\omega_i^{sense} - \alpha_v(L\hat{v}_{i-1}^{random} + \omega_i^{mem})). \end{aligned} \quad [13]$$

Lv_i^{sense} is the true log speed, and ω_i^{mem} and ω_i^{sense} are the memory and sensory noise on log speed, respectively. The random component is a zero-mean Gaussian process; therefore, the deterministic part is the mean

$[\bar{L}_V = (L_{V_1}^{\text{det}}, L_{V_2}^{\text{det}}, \dots, L_{V_n}^{\text{det}})]$ of the speed estimates, and the covariance of $L_{V_i}^{\text{random}}$ is the error covariance of the estimates. The variance of $L_{V_i}^{\text{random}}$ is computed recursively as follows:

$$\text{var}(L_{V_i}^{\text{random}}) = (\alpha_v - \alpha_v K_i)^2 \left(\text{var}(L_{V_{i-1}}^{\text{random}}) + \sigma_{\text{mem}}^2 \right) + K_i^2 \sigma_{\text{sense}}^2, \quad [14]$$

where σ_{mem}^2 is the variance of memory noise and σ_{sense}^2 is the variance of sensory noise. The covariance between $L_{V_i}^{\text{random}}$ and $L_{V_{i+1}}^{\text{random}}$ is given by

$$\text{cov}(L_{V_i}^{\text{random}}, L_{V_{i+1}}^{\text{random}}) = (\alpha_v - \alpha_v K_i)^i \text{var}(L_{V_{i+1}}^{\text{random}}). \quad [15]$$

Using Eqs. 14 and 15, we can compute the full covariance matrix for the speed estimator, Σ_{L_V} .

- Hollingworth HL (1910) The central tendency of judgment. *J Philos Psychol Sci Met* 7 (17):461–469.
- Huttenlocher J, Hedges LV, Vevea JL (2000) Why do categories affect stimulus judgment? *J Exp Psychol Gen* 129(2):220–241.
- Helson H (1964) *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior* (Harper and Row, New York).
- Jazayeri M, Shadlen MN (2010) Temporal context calibrates interval timing. *Nat Neurosci* 13(8):1020–1026.
- Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol* 54(5):358–368.
- Makin A, Stewart A, Poliakoff E (2009) Typical object velocity influences motion extrapolation. *Exp Brain Res* 193(1):137–142.
- Verstynen T, Sabes PN (2011) How each movement changes the next: An experimental and theoretical study of fast adaptive priors in reaching. *J Neurosci* 31(27):10050–10059.
- Berniker M, Voss M, Kording K (2010) Learning priors for Bayesian computations in the nervous system. *PLoS One* 5(9):e12686.
- Miyazaki M, Nozaki D, Nakajima Y (2005) Testing Bayesian models of human coincidence timing. *J Neurophysiol* 94(1):395–399.
- Kowler E, McKee SP (1987) Sensitivity of smooth eye movement to small differences in target velocity. *Vision Res* 27(6):993–1015.
- Lockhead GR (1992) Psychophysical scaling: Judgments of attributes or objects? *Behav Brain Sci* 15(3):543–558.
- Poliakoff E, Collins CJS, Barnes GR (2005) Attention and selection for predictive smooth pursuit eye movements. *Brain Res Cogn Brain Res* 25(3):688–700.
- Sailor K, Antoine M (2005) Is memory for stimulus magnitude Bayesian? *Mem Cognit* 33(5):840–851.
- de Lussanet MHE, Smeets JBJ, Brenner E (2001) The effect of expectations on hitting moving targets: Influence of the preceding target's speed. *Exp Brain Res* 137(2):246–248.
- Lyon DR, Waag WL (1995) Time course of visual extrapolation accuracy. *Acta Psychol (Amst)* 89(3):239–260.
- Makin AD, Poliakoff E, Chen J, Stewart AJ (2008) The effect of previously viewed velocities on motion extrapolation. *Vision Res* 48(18):1884–1893.
- Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci* 9(4):578–585.
- de Bruyn B, Orban GA (1988) Human velocity and direction discrimination measured with random dot patterns. *Vision Res* 28(12):1323–1335.
- Teghtsoonian R (1971) On the exponents in Stevens' law and the constant in Ekman's law. *Psychol Rev* 78(1):71–80.
- Zarco W, Merchant H, Prado L, Mendez JC (2009) Subsecond timing in primates: Comparison of interval production between human subjects and rhesus monkeys. *J Neurophysiol* 102(6):3191–3202.
- DeLucia PR, Liddell GW (1998) Cognitive motion extrapolation and cognitive clocking in prediction motion tasks. *J Exp Psychol Hum Percept Perform* 24(3):901–914.
- Rasche C, Gegenfurtner KR (2009) Precision of speed discrimination and smooth pursuit eye movements. *Vision Res* 49(5):514–523.
- Petzschner FH, Glasauer S (2011) Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *J Neurosci* 31(47):17220–17229.
- Kareev Y (1995) Positive bias in the perception of covariation. *Psychol Rev* 102(3):490–502.
- Kahneman D, Tversky A (1972) Subjective probability: A judgment of representativeness. *Cogn Psychol* 3(3):430–454.
- Green CS, Benson C, Kersten D, Schrater P (2010) Alterations in choice behavior by manipulations of world model. *Proc Natl Acad Sci USA* 107(37):16401–16406.
- Burge J, Ernst MO, Banks MS (2008) The statistical determinants of adaptation rate in human reaching. *J Vis* 8(4):20.1–20.19.
- van Beers RJ (2009) Motor learning is optimally tuned to the properties of motor noise. *Neuron* 63(3):406–417.
- Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. *Nat Neurosci* 5(6):598–604.
- Hedges JH, Stocker AA, Simoncelli EP (2011) Optimal inference explains the perceptual coherence of visual motion stimuli. *J Vis* 11(6):14.
- Chalk M, Seitz AR, Series P (2010) Rapidly learned stimulus expectations alter perception of motion. *J Vis* 10(8):2.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grant R01EY017939 (to D.C.K.).

Supporting Information

Kwon and Knill 10.1073/pnas.1214869110

SI Text

1. Regression Model Selection. We used K-fold matched cross-validation (1) to determine the number of previous stimulus terms in the regression model. Data for each subject were randomly divided into 25 subsets; 24 subsets were used to find the best-fitting (least square) parameters for the regression models, each of which had different numbers of previous stimuli as predictors (from zero to four). The process was repeated 25 times for each of the subjects, producing 25×48 results (mean square errors) for each model. The mean of the mean square errors across subjects was minimized when the regression model includes the current speed and distance terms and the immediately preceding speed term. Fig. S1 shows the mean of mean square errors of the regression models. The error bar represents the SD caused by the random division of dataset. Although the SD is relatively large, the order of the mean square error values stays the same across repeated applications of analysis.

2. The Optimal Weights for an Estimator Formulated in Eq. 8. The optimal updates of the internal speed estimates can be achieved by applying a modified Kalman filter that takes into account the presence of memory noise. The filter update equations are given by

$$L\hat{v}_{i|i-1} = \alpha_v (L\hat{v}_{i-1|i-1} + \alpha_v^{mem} - \mu_{L_v}) + \mu_{L_v}, \quad [S1]$$

$$\text{var}(L\hat{v}_{i|i-1}) = \alpha_v^2 (\text{var}(L\hat{v}_{i-1|i-1}) + \sigma_{mem}^2) + (1 - \alpha_v^2) \sigma_{L_v}^2, \quad [S2]$$

$$L\hat{v}_{i|i} = L\hat{v}_{i|i-1} + K_i (L v_i^{sense} - L\hat{v}_{i|i-1}), \quad [S3]$$

and

$$\text{var}(L\hat{v}_{i|i}) = \text{var}(L\hat{v}_{i|i-1}) \sigma_{sense}^2 / (\text{var}(L\hat{v}_{i|i-1}) + \sigma_{sense}^2), \quad [S4]$$

where $K_i = \text{var}(L\hat{v}_{i|i-1}) / (\text{var}(L\hat{v}_{i|i-1}) + \sigma_{sense}^2)$ and $\sigma_{L_v}^2$ is the variance of stimuli speeds. By plugging Eq. S1 into Eq. S3 and rearranging it, we get

$$L\hat{v}_{i|i} = (1 - \alpha)(1 - K_i) \mu_{L_v} + \alpha(1 - K_i) (L\hat{v}_{i-1|i-1} + \alpha_v^{mem}) + K_i L v_i^{sense}. \quad [S5]$$

Eq. 8 in the text is

$$L\hat{v}_i = w_\mu \mu_{L_v} + w_1 L v_{i-1}^{mem} + (1 - w_\mu - w_1) L v_i^{sense}, \quad [S6]$$

where $L v_{i-1}^{mem}$ is the estimate of speed on trial $i - 1$ corrupted by memory noise. Comparing Eq. S5 with Eq. S6, we have the optimal weight:

$$\begin{aligned} w_\mu &= (1 - \alpha)(1 - K_i) \\ w_1 &= \alpha(1 - K_i) \\ (1 - w_\mu - w_1) &= 1 - (1 - \alpha)(1 - K_i) - \alpha(1 - K_i) = K_i. \end{aligned} \quad [S7]$$

3. Comparing a Tracking Model with the Bayesian Estimation Model. When making judgments of a moving object's position after it disappears, subjects appear to track the target behind the occluder

using a combination of eye movements and attentional tracking (2). Because fitting a noisy tracking model to subjects' data is computationally intractable, we fit an estimator model to subjects' data in the text. This model computes a best estimate of the time that the target will reach the impact zone and uses the estimated time to plan a noisy hitting movement. Here, we compare the performance of a noisy tracking model with the estimator model. To do this comparison, we fit the parameters of a noisy tracking model to best fit the behavior of the estimator model that we fit to subjects' data in experiment 1 and show that the performance of the tracking model fit was nearly equivalent to the performance of the estimator model.

The tracking model that we simulated propagates an internal estimate of the target state as follows after the target disappears behind occlude:

$$\begin{bmatrix} position_{t+1} \\ speed_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & \beta \end{bmatrix} \begin{bmatrix} position_t \\ speed_t \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \omega,$$

where β represents systematic drift of estimated velocity over time, ω represents random noise on the internal speed estimate, and Δt is the duration of the time step used in Monte Carlo simulations (0.01 s). The tracker uses the estimator described in the text to initialize the speed estimate and simply propagates the state update equation forward in time until the predicted time to the impact zone computed from the tracker's position and velocity estimates reaches 500 ms, at which point it initiates a hitting movement. The final hitting time is computed by adding 500 ms plus additive motor noise to the time at which the movement was initiated.

Because we cannot easily fit the parameters of the tracking model directly to subjects' data, we used the following procedure to fit it. For a given set of tracking model parameters, we ran Monte Carlo simulations of the tracking model to generate data akin to real subjects' data. We then fit the estimator model to the simulated data derived from the tracker. The hitting times of the resulting estimator are multivariate Gaussian (in log space), with mean and covariance determined by the model parameters. We used the Kullback-Leibler divergence between this distribution and the distribution derived from the estimator model fit to subjects' data in experiment 1 (using the population mean parameters) as a measure of the fit between the tracking model and the estimator model fit to subjects' data. We used the `fminsearch` command in MATLAB (Mathworks) to find the tracker model parameters (and associated speed estimator parameters) that minimized this KL divergence.

Fitting results show that the best-fitting tracking model can closely emulate the performance of the estimator model when the performances are evaluated by the weights on the current and preceding stimuli (Fig. S2 A and B) and the variance of hitting times as a function of true hitting time (Fig. S2 C and D). Table S1 shows that the parameters of the speed estimator used to initialize the tracker are nearly equivalent to the parameters used in the matched estimator model. The facts that the statistics of model performance are the same for matched tracker and estimator models and that the speed estimator parameters derived from both are essentially identical justify using the estimator model as a computationally tractable stand-in for a noisy tracking model.

4. Adaptive Mean Model. The adaptive mean model assumes that the mean speed follows a simple random walk and that trial-to-trial deviations from the drifting mean are independent. All aspects of the observer-actor model and fitting procedure were the same as used for the correlated speed model, with the ex-

ception of the generative model assumed for target speeds. The generative model can be formalized as a second-order system as follows:

$$\begin{bmatrix} Lv_{i+1} \\ \mu_{Lv_{i+1}} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Lv_i \\ \mu_{Lv_i} \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_v \\ \omega_{\mu_v} \end{bmatrix},$$

where μ_{Lv_i} is mean of speed in log space, and ω_{μ_v} is a zero mean Gaussian noise representing the drift rate of the mean.

As in the correlated speed model, the model observers' measurements of true speed were corrupted by Gaussian sensory noise, and the internal estimates of the previous mean speed used to estimate the current mean speed were corrupted by memory noise. Thus, the parameters of the drifting mean model were analogous to the parameters of the correlated speed model described in the text (and equal in number), except that the drifting mean model has a drift rate parameter instead of the correlation parameter of the correlated speed model. We fit the adaptive mean model in the same way as the correlated speed model using a similar hierarchical prior on the mean and SD of the population's parameters and the same Markov chain Monte

Carlo technique to sample from the posterior densities of the model parameters. The central-tendency bias and the $n - 1$ bias predicted by the best-fitting adaptive mean model closely match the performance of human observer; however, regressing the data derived from simulating each subject's best-fitting model against the speeds on the previous eight trials (rather than just the previous trial) resulted in much higher weights to trials more than one back from the current trial than subjects showed. As shown in Fig. 8 in the text, subjects' performances were much better fit by the correlated speed model than the adaptive mean model.

To further test whether the correlated speed model or the adaptive mean model best fits the data, we computed the marginal likelihood of each model using the Gelfand–Day method. The marginal likelihood of the correlated speed model was higher than the marginal likelihood of the adaptive mean model by a large margin given the data of the first (log Bayes factor = 182) and second experiments (log Bayes factor = 87). The better fit of the correlated speed model is because of the fact that the adaptive mean model gives higher weights to the more than $n - 1$ back trials than human observers.

1. Baddeley RJ, Ingram HA, Miall RC (2003) System identification applied to a visuomotor task: Near-optimal human performance in a noisy changing task. *J Neurosci* 23(7): 3066–3075.

2. DeLucia PR, Liddell GW (1998) Cognitive motion extrapolation and cognitive clocking in prediction motion tasks. *J Exp Psychol Human* 24(3):901–914.

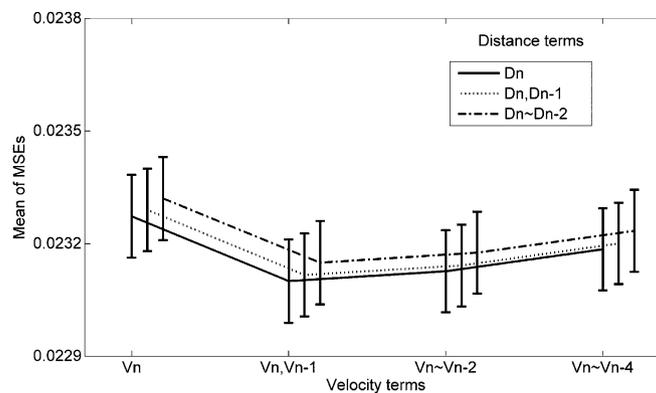


Fig. S1. Results of cross-validation test. The regression model with the current speed, current distance, and immediately preceding speed terms shows the best performance in cross-validation test.

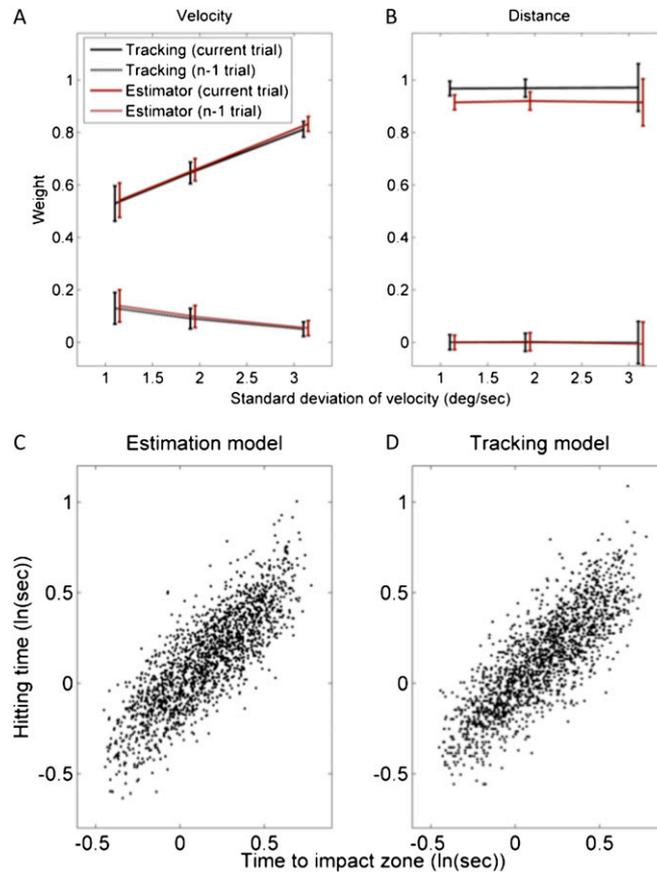


Fig. S2. Comparison between tracking model and estimator model. (A and B) Results of regression analysis applied to the simulation data generated from the tracking model and the estimator model. The parameters of the estimator model were the population mean parameters estimated from experiment 1. The parameters of the tracking model were chosen to minimize the Kullback-Leibler divergence between the simulation performances of the tracking model and the estimator model. (C and D) Simulation data generated from the tracking model and the estimator model. The parameters of the estimator model were the population mean parameters estimated from experiment 1. The parameters of the tracking model were chosen to match the performance of the tracking model to the estimator model.

Table S1. The parameters for the velocity estimation

Parameters	Tracking model	Estimator model (95% credible interval)
Weber fraction	0.142	0.146 (0.122, 0.169)
Temporal correlation (small variance)	0.547	0.548 (0.458, 0.687)
Temporal correlation (medium variance)	0.419	0.426 (0.330, 0.563)
Temporal correlation (large variance)	0.341	0.391 (0.272, 0.503)
SD of prior (small variance)	0.158	0.151 (0.086, 0.207)
SD of prior (medium variance)	0.223	0.208 (0.175, 0.235)
SD of prior (large variance)	0.367	0.342 (0.273, 0.431)