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

$$L\hat{v}_{i|i-1} = \alpha_v \left(L\hat{v}_{i-1|i-1} + \omega_v^{mem} - \mu_{Lv} \right) + \mu_{Lv}, \qquad \textbf{[S1]}$$

$$\mathrm{var}\left(L\hat{v}_{i|i-1}\right) = \alpha_v^2 \left(\mathrm{var}\left(L\hat{v}_{i-1|i-1}\right) + \sigma_{mem}^2 \right) + \left(1-\alpha_v^2\right)\sigma_{Lv}^2, \qquad \textbf{[S2]}$$

$$L\hat{v}_{i|i} = L\hat{v}_{i|i-1} + K_i \left(Lv_i^{sense} - L\hat{v}_{i|i-1} \right), \qquad \textbf{[S3]}$$

and

$$\mathrm{var}\left(L\hat{v}_{i|i}\right) = \mathrm{var}\left(L\hat{v}_{i|i-1}\right)\sigma_{sense}^2 / \left(\mathrm{var}\left(L\hat{v}_{i|i-1}\right) + \sigma_{sense}^2 \right), \qquad \textbf{[S4]}$$

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

$$L\hat{v}_{i|i} = (1-\alpha)(1-K_i)\mu_{Lv} + \alpha(1-K_i)\left(L\hat{v}_{i-1|i-1} + \omega_v^{mem}\right) + K_i Lv_i^{sense}. \qquad \textbf{[S5]}$$

Eq. **8** in the text is

$$L\hat{v}_i = w_\mu \mu_{Lv} + w_1 L\hat{v}_{i-1}^{mem} + (1 - w_\mu - w_1) Lv_i^{sense}, \qquad \textbf{[S6]}$$

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

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

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, ωrepresents random noise on the internal speed estimate, and $\Delta t$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 m 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 exception 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.

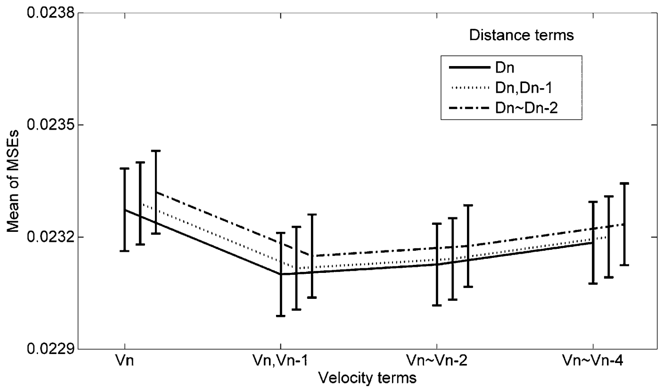

Fig. S1. Results of cross-validation test. The regression model with the current speed, current distance, and immediately preceding speed terms shows the best performance in cross-validation test.

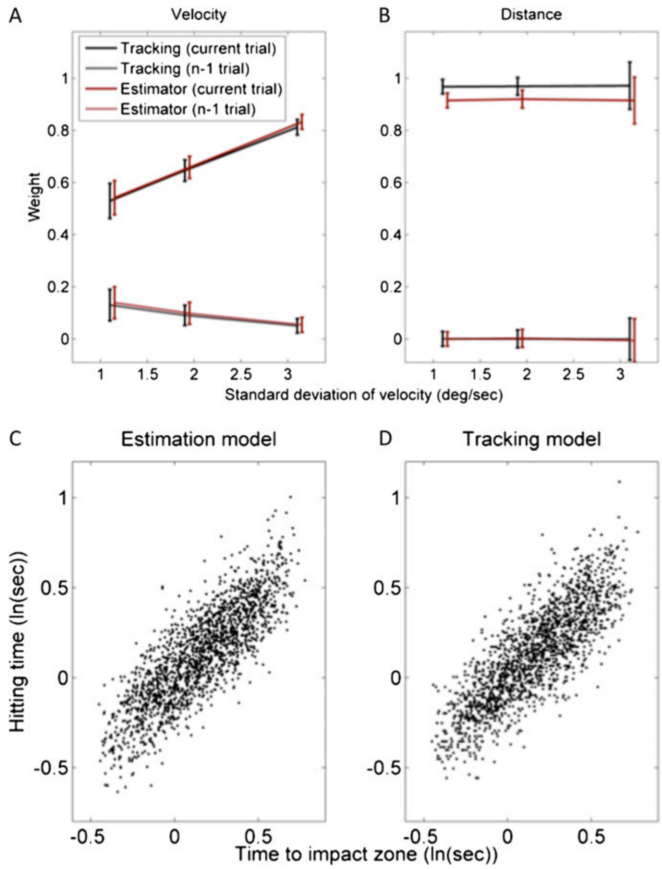

Fig. S2. Comparison between tracking model and estimator model. (*A* and *B*) Results of regression analysis applied to the simulation data generated from the tracking model and the estimator model. The parameters of the estimator model were the population mean parameters estimated from experiment 1. 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)