Supporting Information

Fitzgerald et al. 10.1073/pnas.1015680108

SI Text

Analyticity and the Operator Expansion. We begin by deriving the form of the optimal estimator. Applying the product rule of probability to compute the conditional mean, we obtain

$$\begin{aligned}&\mathbb{E}_{\psi(t)|\{V\}}[\dot\psi|\{V\}]\\&=\mathbb{E}_{\psi(t)|\{V\}}[\mathbb{E}_{C(\theta,t)|\{V\},\psi(t)}[\dot\psi(t)|\{V\},\psi(t)]|\{V\}]\\&=\mathbb{E}_{C(\theta,t),\psi(t)|\{V\}}[\dot\psi(t)|\{V\}].\end{aligned}\quad\textbf{[S1]}$$

Bayes theorem and our noise model imply

$$\begin{aligned}P[C,\psi|\{V\}]&=\frac{P_{\rm Gauss}[\{V\}|C,\psi]P_{\rm prior}[C,\psi]}{P[\{V\}]}\\&=\frac{P_{\rm prior}[C,\psi]\exp\left[-\sum_{n=1}^{N}\int\frac{d\omega}{2\pi}\frac{|V_n(\omega)-\bar V_n(\omega|C,\psi)|^2}{2N(\omega)}\right]}{\tilde Z[\{V\}]},\end{aligned}\quad\textbf{[S2]}$$

where $\tilde Z[\{V\}]$ normalizes the probability. Expanding the complex square and absorbing terms independent of C, ψ into $\tilde Z$, we obtain

$$\begin{aligned}\dot\psi_e(t)&=\frac{1}{\tilde{\tilde Z}[\{V\}]}\mathbb{E}_{C,\psi}\dot\psi(t)\\&\times\exp\left[-\sum_{n=1}^{N}\int\frac{d\omega}{2\pi}\left(\frac{|\bar V_n(\omega|C,\psi)|^2}{2N(\omega)}-\frac{\mathrm{Re}(V_n(\omega)\bar V_n^*(\omega|C,\psi))}{N(\omega)}\right)\right],\end{aligned}\quad\textbf{[S3]}$$

where $\mathrm{Re}(z)$ denotes the real part of the complex variable z and $\tilde{\tilde Z}$ is a new normalization function.

Eq. S3 differs from the main text because of the second term within the exponential. However, because the relevant $\{V_n(\omega)\}$ correspond to real-valued $\{V_n(t)\}$, we have

$$V_n^*(\omega)=V_n(-\omega).\quad\textbf{[S4]}$$

An identical argument applies to $\bar V_n$ and δV_n. This argument implies

$$\mathrm{Im}(V_n\bar V_n^*)(\omega)=-\mathrm{Im}(V_n\bar V_n^*)(-\omega).\quad\textbf{[S5]}$$

Thus, $\mathrm{Im}(V_n\bar V_n^*)/N$ is an odd function of ω, and its integral is zero. This result implies that the integral of $\mathrm{Re}(V_n\bar V_n^*)$ can be replaced with one of $V_n\bar V_n^*$ once we restrict to real-valued voltages in the time domain.

When we perform the functional expansion, we seem to be claiming that $\dot\psi_e$ is analytic at all points $\mathbf{V}=(V_1(\omega),\ldots,V_N(\omega))$. Indeed, because $Z[\{V\}]$ is nonzero, holomorphy (and thus analyticity) for the function defined in the text follows easily from the Cauchy–Riemann equations. However, the optimal estimator properly defined by Eq. S3 is nonholomorphic due to its dependence on $\{V_n^*(\omega)\}$. This result implies that although the expansion is not valid for arbitrary $\{V_n(\omega)\}$, it is valid for $\{V_n(\omega)\}$ with real Fourier transforms. In other words, the optimal estimator given by Eq. S3 reduces to the equation in the text for real-valued $\{V_n(t)\}$, so whereas the optimal estimator is not analytic over the entire input space, it can be represented with a functional expansion in the relevant subspace.

Time-Invariant Priors and the Volterra Series. If the prior is invariant to $\{C(\theta,t),\psi(t)\}\rightarrow\{C(\theta,t+\Delta),\psi(t+\Delta)\}$ for arbitrary Δ, then the system possesses time-translation symmetry, and the series of multipoint correlators is a Volterra series. In this context, time-translation symmetry implies that the statistical properties of expected stimuli do not vary over time. Violations of time-translation symmetry may occur due to adaptation if the organism updates its expectations to improve motion estimation. Nevertheless, many situations of interest have time-translation symmetry. Because our theory is carried out in the frequency domain, we obtain a definition of the Volterra series in this domain by Fourier transforming the time-domain Volterra series. Comparing it to the functional expansion, we conclude that our theory is a Volterra series when it can be written in the form

$$\dot\psi_{e,i_1,\ldots,i_l|\mathrm{V}=0}=e^{-i\sum_j\omega_jt}\kappa^{(l)}_{i_1,\ldots,i_l}(\omega_1,\ldots,\omega_l),\quad\textbf{[S6]}$$

where $\{\kappa\}$ are arbitrary time-independent functions.

We will see that the time variance of the estimator is determined by the time variance of the priors. Because $\dot\psi_{e,i_1,\ldots,i_l}$ is a product of Ψ, Z, and the derivatives, we examine their form at zero voltage. First, note that

$$f_\Delta(t)=f(t+\Delta)\Rightarrow f_\Delta(\omega)=e^{-i\omega\Delta}f(\omega).\quad\textbf{[S7]}$$

Thus,

$$\begin{aligned}\Psi_{,i_1,\ldots,i_l|V=0}&=\mathbb{E}_{C(\theta,t),\psi(t)}\exp\left[-\sum_{n=1}^{N}\int\frac{d\omega}{2\pi}\frac{|\bar V_{n\Delta}(\omega|C,\psi)|^2}{2N(\omega)}\right]\\&\times e^{-i\sum_j\omega_j\Delta}\frac{\dot\psi(t)}{(2\pi)^l}\frac{\bar V^*_{i_1\Delta}(\omega_1|C,\psi)\ldots\bar V^*_{i_l\Delta}(\omega_l|C,\psi)}{N(\omega_1)\ldots N(\omega_l)}.\end{aligned}\quad\textbf{[S8]}$$

Because $\bar V_n$ is time invariant, its time translation is equivalent to time translating its inputs,

$$\begin{aligned}\Psi_{,i_1,\ldots,i_l|V=0}&=\mathbb{E}_{C(\theta,t),\psi(t)}\exp\left[-\sum_{n=1}^{N}\int\frac{d\omega}{2\pi}\frac{|\bar V_n(\omega|C_\Delta,\psi_\Delta)|^2}{2N(\omega)}\right]\\&\times e^{-i\sum_j\omega_j\Delta}\frac{\dot\psi_\Delta(t-\Delta)}{(2\pi)^l}\frac{\bar V^*_{i_1}(\omega_1|C_\Delta,\psi_\Delta)\ldots\bar V^*_{i_l}(\omega_l|C_\Delta,\psi_\Delta)}{N(\omega_1)\ldots N(\omega_l)}.\end{aligned}\quad\textbf{[S9]}$$

If the prior is time-translation invariant, we can replace expectations over C, ψ with expectations over C_Δ, ψ_Δ:

$$\Psi_{,i_1,\ldots,i_l|V=0}(t)=e^{-i\sum_j\omega_j\Delta}\Psi_{,i_1,\ldots,i_l|V=0}(t-\Delta).\quad\textbf{[S10]}$$

This equation must hold for all t and Δ. If we define some constant reference time t_0, then

$$\Psi_{,i_1,\ldots,i_l|V=0}(t) = e^{-i\sum_j \omega_j (t-t_0)} \kappa^{(l)}_{i_1,\ldots,i_l}(\omega_1, \ldots, \omega_l), \quad \textbf{[S11]}$$

where

$$\kappa^{(l)}_{i_1,\ldots,i_l}(\omega_1, \ldots, \omega_l) = \mathbb{E}_{C(\theta,t),\psi(t)} \frac{\dot\psi(t_0)}{(2\pi)^l} \exp\left[-\sum_{n=1}^{N} \int \frac{d\omega |\bar V_n(\omega|C,\psi)|^2}{2\pi \quad 2N(\omega)}\right] \times \frac{\bar V^*_{i_1}(\omega_1|C,\psi)\ldots \bar V^*_{i_l}(\omega_l|C,\psi)}{N(\omega_1)\ldots N(\omega_l)}. \quad \textbf{[S12]}$$

The identical argument leading to Eq. **S9** can be applied to write

$$Z_{,i_1,\ldots,i_l|V=0} = \mathbb{E}_{C(\theta,t),\psi(t)} \exp\left[-\sum_{n=1}^{N} \int \frac{d\omega}{2\pi} \frac{|\bar V_n(\omega|C_\Delta,\psi_\Delta)|^2}{2N(\omega)}\right] \times \frac{1}{(2\pi)^l} e^{-i\sum_j \omega_j \Delta} \frac{\bar V^*_{i_1}(\omega_1|C_\Delta,\psi_\Delta)\ldots \bar V^*_{i_l}(\omega_l|C_\Delta,\psi_\Delta)}{N(\omega_1)\ldots N(\omega_l)}. \quad \textbf{[S13]}$$

Assuming that $P[C,\psi]$ is time-translation invariant, we may replace the expectation over C, ψ with an expectation over C_Δ, ψ_Δ and obtain that either $Z_{,i_1,\ldots,i_l|V=0} = 0$ or $\sum_j \omega_j = 0$. In either case,

$$Z_{,i_1,\ldots,i_l|V=0} = e^{-i\sum_j \omega_j t} \mathbb{E}_{C(\theta,t),\psi(t)} \frac{1}{(2\pi)^l} \exp\left[-\sum_{n=1}^{N} \int \frac{d\omega |\bar V_n(\omega|C,\psi)|^2}{2\pi \quad 2N(\omega)}\right] \times \frac{\bar V^*_{i_1}(\omega_1|C,\psi)\ldots \bar V^*_{i_l}(\omega_l|C,\psi)}{N(\omega_1)\ldots N(\omega_l)}. \quad \textbf{[S14]}$$

Because both Eqs. **S11 and S14** have the prefactor $e^{-i\sum_j \omega_j t}$, Eq. **S6** holds when the prior is time-translation invariant.

Note that assuming white noise ($N(\omega) \equiv \mathcal{N}$) and time-translation symmetry, the derivatives are simple in the time domain,

$$\Psi_{,i_1,\ldots,i_l}(t_1, \ldots, t_l) = \mathbb{E}_{C(\theta,t),\psi(t)} \exp\left[-\sum_n \int dt \frac{(\bar V_n(t|C,\psi))^2}{2\mathcal{N}}\right] \times \frac{1}{\mathcal{N}^l} \dot\psi(t_0) \bar V_{i_1}(t_0 - t_1)\ldots \bar V_{i_l}(t_0 - t_l) \quad \textbf{[S15]}$$

and

$$Z_{,i_1,\ldots,i_l}(t_1, \ldots, t_l) = \mathbb{E}_{C(\theta,t),\psi(t)} \exp\left[-\sum_n \int dt \frac{(\bar V_n(t|C,\psi))^2}{2\mathcal{N}}\right] \times \frac{1}{\mathcal{N}^l} \bar V_{i_1}(t_0 - t_1) \bar V_{i_2}(t_0 - t_2)\ldots \bar V_{i_l}(t_0 - t_l). \quad \textbf{[S16]}$$

Because t_0 is arbitrary, it is convenient to average over it when computing the temporal domain Volterra kernels. Intuitively, Eq. **S15** computes correlations between the velocity and the stimulus whereas Eq. **S16** computes pure stimulus correlations. Clearly, if the system is not time-translation invariant, then this procedure to obtain the time-domain kernels would not apply.

Photovoltage Expectations.

In the main text we applied some basic facts regarding the structure of

$$\mathbb{E}_{\{V(t)\}|C(\theta,t),\psi(t)} V_{i_1}(\omega_1)\ldots V_{i_l}(\omega_l). \quad \textbf{[S17]}$$

This section discusses how functional integral methods borrowed from physics allow the evaluation of these expectations.

First, it is worth pointing out a source of potential confusion. Given a single Gaussian-distributed complex variable,

$$P(z) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{|z|^2}{2\sigma^2}\right], \quad \textbf{[S18]}$$

it is straightforward to show that

$$\mathbb{E}_z |z|^2 = 2\sigma^2 \quad \textbf{[S19]}$$

$$\mathbb{E}_z z^2 = 0. \quad \textbf{[S20]}$$

From Eq. **S20**, one might expect that Eq. **S17** be exactly $\bar V_{i_1}(\omega_1)\ldots \bar V_{i_l}(\omega_l)$ with the noise playing no role. This conclusion would be incorrect because we must take all expectations over $\{V(t)\}$ rather than $\{V(\omega)\}$.

To work with zero-mean Gaussian distributions we begin by rewriting Eq. **S17** as

$$\mathbb{E}_{\{\delta V(t)\}} (\bar V_{i_1}(\omega_1|C,\psi) + \delta V_{i_1}(\omega_1))\ldots(\bar V_{i_l}(\omega_l|C,\psi) + \delta V_{i_l}(\omega_l)). \quad \textbf{[S21]}$$

After expanding this product, each term will contain a variable number of $\bar V$s (that can be factored out of the expectation) together with noise terms. For now, we ignore copies of $\bar V$ and evaluate expectations of the form

$$\mathbb{E}_{\{\delta V(t)\}} \delta V_{i_1}(\omega_1)\ldots \delta V_{i_m}(\omega_m), \quad \textbf{[S22]}$$

where $0 \le m \le l$.

To evaluate these expectations, we use the method of functional integration and rewrite **S22** as

$$\frac{\int D\{\delta V_n(t)\}\, \delta V_{i_1}(\omega_1)\ldots \delta V_{i_m}(\omega_m) \exp\left[-\sum_n \int \frac{d\omega}{2\pi} \frac{|\delta V_n(\omega)|^2}{2N(\omega)}\right]}{\int D\{\delta V_n(t)\} \exp\left[-\sum_n \int \frac{d\omega |\delta V_n(\omega)|^2}{2\pi \quad 2N(\omega)}\right]}. \quad \textbf{[S23]}$$

These are Gaussian functional integrals and can be exactly evaluated. The denominator is evaluated to give

$$\int D\{\delta V_n(t)\} \exp\left[-\sum_n \int \frac{d\omega}{2\pi} \frac{|\delta V_n(\omega)|^2}{2N(\omega)}\right] = \left(\prod_\omega \sqrt{2\pi V N(\omega)}\right)^N, \quad \textbf{[S24]}$$

where N is the number of photoreceptors and V is a constant arising from a discrete approximation to the functional integral. To evaluate the general form of the numerator, it is important to note that because the noise has zero mean, the $m = 1$ case vanishes. The $m = 2$ case can be evaluated as

$$\int D\{\delta V_n(t)\}\delta V_{i_1}(\omega_1)\delta V_{i_2}(\omega_2)\exp\left[-\sum_n\int\frac{d\omega}{2\pi}\frac{|\delta V_n(\omega)|^2}{2N(\omega)}\right] = \delta_{i_1,i_2}\delta(\omega_1+\omega_2)N(\omega_1)\left(\prod_\omega\sqrt{2\pi VN(\omega)}\right)^N. \quad \text{[S25]}$$

Note that the last factor in Eq. S25 cancels the denominator and removes the dependence on the discrete approximation. Therefore, we take the continuum limit and find

$$\mathbb{E}_{\{\delta V(t)\}}\delta V_{i_1}(\omega_1)\delta V_{i_2}(\omega_2) = \delta_{i_1,i_2}\delta(\omega_1+\omega_2)N(\omega_1) \quad \text{[S26]}$$

as intuitively expected. This result consequently provides

$$\mathbb{E}_{\{V(t)\}|C(\theta,t),\psi(t)}V_{i_1}(\omega_1)V_{i_2}(\omega_2) = \bar V_{i_1}(\omega_1|C,\psi)\bar V_{i_2}(\omega_2|C,\psi) + \delta_{i_1,i_2}\delta(\omega_1+\omega_2)N(\omega_1). \quad \text{[S27]}$$

When $m > 2$, the functional integral is zero unless each index can be paired with another index (as in Eq. S25). Such a pairing is referred to as a contraction. Generally,

$$\mathbb{E}_{\{\delta V(t)\}}\delta V_{i_1}(\omega_1)\ldots\delta V_{i_m}(\omega_m) = \begin{cases}0, & m \text{ odd}\\ \text{sum over all contractions}, & m \text{ even.}\end{cases} \quad \text{[S28]}$$

Eq. S28 is best demonstrated by example on the quartic expectation,

$$\begin{aligned}\mathbb{E}_{\{\delta V(t)\}}&\delta V_{i_1}(\omega_1)\delta V_{i_2}(\omega_2)\delta V_{i_3}(\omega_3)\delta V_{i_4}(\omega_4)\\ &= \delta_{i_1,i_2}\delta(\omega_1+\omega_2)N(\omega_1)\delta_{i_3,i_4}\delta(\omega_3+\omega_4)N(\omega_3)\\ &+ \delta_{i_1,i_3}\delta(\omega_1+\omega_3)N(\omega_1)\delta_{i_2,i_4}\delta(\omega_2+\omega_4)N(\omega_2)\\ &+ \delta_{i_1,i_4}\delta(\omega_1+\omega_4)N(\omega_1)\delta_{i_2,i_3}\delta(\omega_2+\omega_3)N(\omega_2).\end{aligned} \quad \text{[S29]}$$

and,

$$\begin{aligned}\mathbb{E}_{\{V(t)\}|C(\theta,t),\psi(t)}&V_{i_1}(\omega_1)V_{i_2}(\omega_2)V_{i_3}(\omega_3)V_{i_4}(\omega_4)\\ &= \bar V_{i_1}(\omega_1|C,\psi)\bar V_{i_2}(\omega_2|C,\psi)\bar V_{i_3}(\omega_3|C,\psi)\bar V_{i_4}(\omega_4|C,\psi)\\ &+ \bar V_{i_1}(\omega_1|C,\psi)\bar V_{i_2}(\omega_2|C,\psi)\delta_{i_3,i_4}\delta(\omega_3+\omega_4)N(\omega_3)\\ &+ \bar V_{i_1}(\omega_1|C,\psi)\bar V_{i_3}(\omega_3|C,\psi)\delta_{i_2,i_4}\delta(\omega_2+\omega_4)N(\omega_2)\\ &+ \bar V_{i_1}(\omega_1|C,\psi)\bar V_{i_4}(\omega_4|C,\psi)\delta_{i_2,i_3}\delta(\omega_2+\omega_3)N(\omega_2)\\ &+ \bar V_{i_2}(\omega_2|C,\psi)\bar V_{i_3}(\omega_3|C,\psi)\delta_{i_1,i_4}\delta(\omega_1+\omega_4)N(\omega_1)\\ &+ \bar V_{i_2}(\omega_2|C,\psi)\bar V_{i_4}(\omega_4|C,\psi)\delta_{i_1,i_3}\delta(\omega_1+\omega_3)N(\omega_1)\\ &+ \bar V_{i_3}(\omega_3|C,\psi)\bar V_{i_4}(\omega_4|C,\psi)\delta_{i_1,i_2}\delta(\omega_1+\omega_2)N(\omega_1)\\ &+ \delta_{i_1,i_2}\delta(\omega_1+\omega_2)N(\omega_1)\delta_{i_3,i_4}\delta(\omega_3+\omega_4)N(\omega_3)\\ &+ \delta_{i_1,i_3}\delta(\omega_1+\omega_3)N(\omega_1)\delta_{i_2,i_4}\delta(\omega_2+\omega_4)N(\omega_2)\\ &+ \delta_{i_1,i_4}\delta(\omega_1+\omega_4)N(\omega_1)\delta_{i_2,i_3}\delta(\omega_2+\omega_3)N(\omega_2).\end{aligned} \quad \text{[S30]}$$

All other expectations can be written in the same manner.

Features of the *lth* Derivatives. It is important to verify the three conditions that were used in the symmetry arguments in the main text: (*i*) the denominator of $\psi_{e,i_1,\ldots,i_l}$ is Z^{2^l}, (*ii*) the numerators sum terms that distribute l derivatives among Ψ and Z, and (*iii*) each term in the numerator multiplies 2^l copies of Ψ, Z, or their derivatives.

We proceed by induction. The base case is apparent. Suppose that all three claims are true for the $(l-1)$st derivatives and write

$$\dot\psi_{e,i_1,\ldots,i_{l-1}} = \frac{\mathcal{M}^{(l-1)}}{Z^{2^{l-1}}}, \quad \text{[S31]}$$

where $\mathcal{M}^{(l-1)}$ satisfies the second and third conditions. Taking one more derivative we obtain

$$\dot\psi_{e,i_1,\ldots,i_l} = \frac{\mathcal{M}^{(l-1)}_{,i_l}Z^{2^{l-1}} - \mathcal{M}^{(l-1)}2^{l-1}Z^{2^{l-1}-1}Z_{,i_l}}{(Z^{2^{l-1}})^2}. \quad \text{[S32]}$$

From this expression we can verify all three conditions: (*i*) The new denominator is explicitly $Z^{2^{l-1}2} = Z^{2^l}$. (*ii*) In the first term in the numerator, because each term in $\mathcal{M}^{(l-1)}$ distributes $l-1$ derivatives across Ψ and Z, $\mathcal{M}^{(l-1)}_{,i_l}$ distributes the l derivatives appropriately. The second term is already in the desired form. (*iii*) The first term in the numerator mutiplies $2^{l-1} + 2^{l-1} = 2^{(l-1)}2 = 2^l$ terms whereas the second term multiplies $2^{l-1} + 2^{l-1} - 1 + 1 = 2^l$ terms.

These conditions were used in the development of the main text to conclude the following: (*i*) Because $Z[V]_{|V=0} \neq 0$, the denominator is nonzero. Thus, a vanishing numerator implies the whole expression vanishes. (*ii*) Because the number of derivatives in a term must sum to l, if l is odd then the lth derivative can't be written as the product of only even-ordered derivatives. (*iii*) If the number of factors in the numerator were different from the number in the denominator, then all functional integrals would diverge or vanish. Numerically, we evaluated the first three kernels according to equations in the text and

$$\begin{aligned}\dot\psi_{e,i_1,i_2,i_3} = \frac{1}{Z^8}(&\Psi_{,i_1,i_2,i_3}Z^7 - \Psi_{,i_1,i_2}Z_{,i_3}Z^6\\ &-\Psi_{,i_1,i_3}Z_{,i_2}Z^6 - \Psi_{,i_2,i_3}Z_{,i_1}Z^6 - \Psi_{,i_1}Z_{,i_2,i_3}Z^6\\ &-\Psi_{,i_2}Z_{,i_1,i_3}Z^6 - \Psi_{,i_3}Z_{,i_1,i_2}Z^6 + 2\Psi_{,i_1}Z_{,i_2}Z_{,i_3}Z^5\\ &+2\Psi_{,i_2}Z_{,i_1}Z_{,i_3}Z^5 + 2\Psi_{,i_3}Z_{,i_1}Z_{,i_2}Z^5 - \Psi Z_{,i_1,i_2,i_3}Z^6\\ &+2\Psi Z_{,i_1}Z_{,i_2,i_3}Z^5 + 2\Psi Z_{,i_2}Z_{,i_1,i_3}Z^5 + 2\Psi Z_{,i_3}Z_{,i_1,i_2}Z^5\\ &-6\Psi Z_{,i_1}Z_{,i_2}Z_{,i_3}Z^4).\end{aligned} \quad \text{[S33]}$$

Linear Motion Estimation. The discussion in the main text shows that to support a linear term in the expansion, there must be a nonzero average image $< C(k,t) >$ in $\mathcal{P}$. It is important to realize that as a *function* this is not constrained to be zero. Because we separated self-motion from the dynamics of the world, nonzero $< C(k,t) >$ does not imply that photoreceptors experience a nonzero average contrast. Because nonzero $< C(k,t) >$ does not seem likely for a static prior, linear estimators may be more appealing for adaptive motion estimation.

This section shows how linear estimators may arise with an example. We consider an arbitrary static stimulus, $\mathcal{C}(\theta)$. For simplicity, we suppose that the animal rotates with constant speed in either direction. Then time-reversal symmetry is satisfied and the prior takes the form

$$P[C(\theta,t),\psi(t)] = \frac{1}{2}\delta[C(\theta,t)-\mathcal{C}(\theta)] \times(\delta[\psi(t)-\omega_0 t]+\delta[\psi(t)+\omega_0 t]). \quad \text{[S34]}$$

Because of time-reversal symmetry, $\Psi_{\{V\}=0} = 0$ and $|\bar V_n(\omega|\mathcal{C}(\theta),\omega_0 t)|^2 = |\bar V_n(\omega|\mathcal{C}(\theta),-\omega_0 t)|^2$, and it is simple to evaluate the required derivatives,

$$Z_{\{V\}=0} = \exp\left[-\sum_n \int \frac{d\omega}{2\pi} \frac{|\bar{V}_n(\omega|\mathcal{C}(\theta),\ \omega_0 t)|^2}{2N(\omega)}\right] \quad \textbf{[S35]}$$

$$\Psi_i(\omega_1)_{\{V\}=0} = \exp\left[-\sum_n \int \frac{d\omega}{2\pi} \frac{|\bar{V}_n(\omega|\mathcal{C}(\theta),\ \omega_0 t)|^2}{2N(\omega)}\right] \times \frac{\omega_0}{4\pi N(\omega_1)}(\bar{V}_i^*(\omega_1|\mathcal{C}(\theta),\ \omega_0 t) - \bar{V}_i^*(\omega_1|\mathcal{C}(\theta),\ -\omega_0 t)). \quad \textbf{[S36]}$$

This calculation immediately gives the linear estimation kernel as

$$k_i^{(1)}(\omega_1) = \frac{\omega_0}{4\pi N(\omega_1)}(\bar{V}_i^*(\omega_1|\mathcal{C}(\theta),\ \omega_0 t) - \bar{V}_i^*(\omega_1|\mathcal{C}(\theta),\ -\omega_0 t)). \quad \textbf{[S37]}$$

Note that the linear kernel is proportional to the speed.

Because the image is static, it is straightforward to evaluate the average voltages. Performing first the time integral,

$$\begin{aligned}\bar{V}_i(\omega_1|\mathcal{C}, \pm\omega_0 t) &= T(\omega_1)\int dt e^{i\omega_1 t}\int \frac{dk}{2\pi} e^{-ik(\theta_i \mp \omega_0 t)} M(-k)\mathcal{C}(k) \\ &= T(\omega_1)\int \frac{dk}{2\pi} e^{-ik\theta_i} M(-k)\mathcal{C}(k)\delta(\omega_1 \pm k\omega_0).\end{aligned} \quad \textbf{[S38]}$$

We now recall that $\delta(\alpha x) = \delta(x)/|\alpha|$ to convert the δ-functions over ω_1 to δ-functions over k. Then,

$$\bar{V}_i(\omega_1|\mathcal{C}, \pm\omega_0 t) = \frac{T(\omega_1)}{2\pi\omega_0} e^{\pm i\omega_1\theta_i/\omega_0} M(\pm\omega_1/\omega_0)\mathcal{C}(\mp\omega_1/\omega_0). \quad \textbf{[S39]}$$

The first thing to note is that the linear estimator depends upon the photoreceptor (i.e., it is not independent of θ_i). This dependence is a reflection of the fact that the estimator needs the appropriate phase to perform the motion estimation.

If, as above, the phase is known, we may compute the average estimate of the velocity. For simplicity, suppose that the animal is stimulated with positive angular frequency. Then,

$$\begin{aligned}\langle\dot{\psi}_e^{(1)}\rangle_{\omega_0 t} &= \sum_i \int d\omega k_i^{(1)}(\omega)\bar{V}_i(\omega|\mathcal{C}, \omega_0 t) \\ &= \int \frac{d\omega}{2\pi}\frac{\omega_0}{2N(\omega)}\left(\sum_i |\bar{V}_i(\omega|\mathcal{C}(\theta), \omega_0 t)|^2 \right. \\ &\quad \left. - \sum_i \bar{V}_i^*(\omega|\mathcal{C}(\theta),\ -\omega_0 t)\bar{V}_i(\omega|\mathcal{C}(\theta),\ \omega_0 t)\right).\end{aligned} \quad \textbf{[S40]}$$

However, consulting our equation for the average voltage, we see that the second term is proportional to $\sum_i e^{2i\omega\theta_i/\omega_0}$. This sum of phases will largely cancel, making the second term small compared with the first. For clarity we ignore it:

$$\langle\dot{\psi}_e^{(1)}\rangle_{\omega_0 t} = \frac{N}{8\pi^2\omega_0}\int \frac{d\omega}{2\pi}\frac{|T(\omega)|^2|M(\omega/\omega_0)|^2|\mathcal{C}(\omega/\omega_0)|^2}{N(\omega)}. \quad \textbf{[S41]}$$

This result is strictly positive. It is easy to see that when the organism is stimulated with negative angular frequency, the magnitude of the estimate is the same but the sign is reversed:

$$\langle\dot{\psi}_e^{(1)}\rangle_{-\omega_0 t} = -\frac{N}{8\pi^2\omega_0}\int \frac{d\omega}{2\pi}\frac{|T(\omega)|^2|M(\omega/\omega_0)|^2|\mathcal{C}(\omega/\omega_0)|^2}{N(\omega)}. \quad \textbf{[S42]}$$

As desired, the linear estimator on average estimates the sign of the velocity correctly. This example shows that if an average image can be inferred by the animal, linear motion estimators do in fact exist to extract motion signals from this information. It remains unclear whether a biological system can estimate an average image over the relevant timescale to facilitate such a motion estimation strategy.

When the Prior Matters. In this section we investigate the question of when the prior significantly influences the estimation strategy. At sufficiently low signal-to-noise ratio (SNR) the neural signals are too noisy to provide a precise motion estimate and prior expectations dominate the estimation strategy. Contrastingly, at high SNR the prior is less important for estimation than the high-fidelity signal. The prior and neural signals compete for relevance in the crossover regime, where $\log[P_{\text{prior}}[C, \psi]] \approx \sum_n^N \int \frac{d\omega|\bar{V}_n(\omega|C, \psi)|^2}{2\pi \quad 2N(\omega)}$. If we assume white noise, then we may rearrange this to read $\mathcal{S}[C, \psi] \approx \frac{2\log[P_{\text{prior}}[C, \psi]]}{NT}$, where $\mathcal{S}[C,\psi]$ is the average SNR per photoreceptor per unit time, N is the number of photoreceptors, and T is the time over which signal power and noise power are integrated. At $\mathcal{S}$, the signal and prior have comparable influence on motion estimation. We refer to $\mathcal{S}$ as the *equivalent SNR*. The equivalent SNR is a function and should be interpreted as determining how important the prior probability distribution is for assigning the appropriate weight for the pair $\{C(\theta, t), \psi(t)\}$. Whereas $\mathcal{S}$ is a random variable, if we consider the simple example where there are $\mathcal{R}$ rotational states per unit time and $\mathcal{C}$ contrast states per photoreceptor per unit time and assume all states are equally probable, then the equivalent SNR simplifies considerably, $\mathcal{S} \sim \log(\mathcal{R}^{1/N}\mathcal{C})$. When the direction of motion is to be inferred from binary random dots, $\mathcal{R} \sim \mathcal{C} \sim 2$. Assuming a one-dimensional fly eye, $N \sim 75$, we obtain S ~ 0.7. Whereas the equivalent SNR depends weakly on $\mathcal{R}$, it increases faster with stimulus complexity, $\mathcal{C} \sim 10 \Rightarrow \mathcal{S} \sim 2.3$, $\mathcal{C} \sim 100 \Rightarrow \mathcal{S} \sim 4.6$, $\mathcal{C} \sim 1{,}000 \Rightarrow \mathcal{S} \sim 6.9$. For naturalistic stimuli we expect a large number of contrast states and the photoreceptor SNR may become comparable to the equivalent SNR (1). In this scenario a prior is needed to successfully analyze natural stimuli. On the other hand, many experiments operate in an artificial signal-dominated regime that does not require the prior. Overly simplistic stimuli may lead to errors and misrepresent the animal's ability to estimate motion.

The expectations that we need to evaluate can be written abstractly in the form, $Y = \mathbb{E}_{C,\psi}X[C, \psi]$. The prior may have an influence in the signal-dominated regime if it breaks the symmetry of X. Suppose $\mathcal{T}$ is a transformation that satisfies $\mathcal{T}^2 = \text{Id}$ and $X[\mathcal{T}C, \mathcal{T}\psi] = \pm X[C, \psi]$. When averaging over P_{prior}, the weight of $\{C, \psi\}$ and $\{\mathcal{T}C, \mathcal{T}\psi\}$ will be $P_{\text{prior}}[C, \psi] \pm P_{\text{prior}}[\mathcal{T}C, \mathcal{T}\psi]$. If $\{C, \psi\}$ is favored by the prior, the sum is essentially $P_{\text{prior}}[C, \psi]$, and if $\{\mathcal{T}C, \mathcal{T}\psi\}$ is favored, it is $\pm P_{\text{prior}}[\mathcal{T}C, \mathcal{T}\psi]$. At high SNR, the signal selects pairs of stimuli that cannot be distinguished on the basis of neural signals whereas the prior selects within pair. In particular, departures in the prior from the symmetry can have a significant impact on the estimator by destroying the cancellation facilitated by X.

We examine the magnitude of the effects produced by the symmetry breaking, using a simple example. Suppose $\mathcal{T}$ is the contrast inversion operator and consider a stimulus composed of binary random dots. We numerically evaluate the second- and third-order estimators ($P_t = 0.01$, $P_v = 0.01$) as a function of the probability of a dark dot, P_{dark}, and quantify their performance with the fraction of trials in which they properly distinguish leftward from rightward motion (Fig. S1*A*) and their correlation with the true stimulus velocity (Fig. S1*B*). The Reichardt correlator performs well independently of P_{dark} whereas the third-order correlator requires asymmetry. For large asymmetries, both the Reichardt and three-point correlators are suitable estimators.

Their inclusion reliably decreases the squared error (Fig. S1*C*), but they both underestimate the true velocity compared with the full estimate (Fig. S1*D*). As SNR increases, the full estimator improves, low-order estimators lose importance, and the benefit from the three-point correlator becomes comparable to the Reichardt correlator. We tested the estimators using 200-ms probe stimuli ($P_v = 0$). In Fig. S1, we used 10,000 simulations for *A* and *B*, used 100 simulations for *C* and *D*, and compared the true velocity to its estimate at 200 ms. We considered the kernels $\{k^{(2)}_{i,i+1}, k^{(2)}_{i,i+2}, k^{(3)}_{i,i+1,i+2}\}$ and the full optimal estimator.

Despite the Reichardt correlator's correlation with the velocity and success in direction discrimination, it yields only a small fraction of the true velocity (2) captured by the complete estimator. The correlation implies that scaling the Reichardt output would often produce a reasonable estimate of velocity. In fact, experimental evidence suggests that this technique, termed "contrast normalization", may be used biologically to repair some of the failures of pairwise models (3, 4). Because optimality determines the size of each term in our theory, we cannot include contrast normalization explicitly in the model. However, it should be noted that although scaling the response has been shown to work for some stimuli, it does not apply generally (3). Furthermore, our theory shows that quantitatively better performance is obtained by including higher-order estimators and long-range interactions. Even in simple situations, higher-order statistics contain useful information that cannot be obtained by two-point correlations alone, and more complex situations may make higher-order statistics more prominent. Multipoint correlators can also be used to estimate acceleration or translational velocity, so if the organism estimates these complex parameters directly (5), higher-order correlators may play a central role.

1. Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D (1991) Reading a neural code. *Science* 252:1854–1857.
2. Egelhaaf M, Borst A, Reichardt W (1989) Computational structure of a biological motion-detection system as revealed by local detector analysis in the fly's nervous system. *J Opt Soc Am A* 6:1070–1087.
3. Dror RO, O'Carroll DC, Laughlin SB (2001) Accuracy of velocity estimation by Reichardt correlators. *J Opt Soc Am A Opt Image Sci Vis* 18:241–252.
4. Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9: 181–197.
5. Thiel A, Greschner M, Eurich CW, Ammermüller J, Kretzberg J (2007) Contribution of individual retinal ganglion cell responses to velocity and acceleration encoding. *J Neurophysiol* 98:2285–2296.

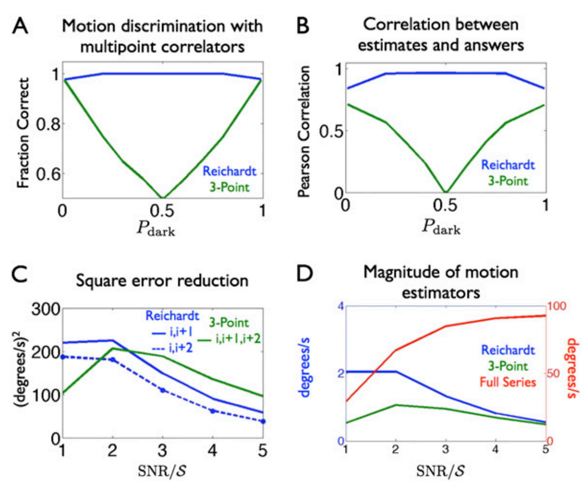

Fig. S1. The onset of a three-point correlator. To test the estimators we numerically evaluated $k^{(2)}_{i,i+1}, k^{(2)}_{i,i+2}, k^{(3)}_{i,i+1,i+2}$ for binary random dots rotating at 100°/s. (*A*) The fraction of trials where the estimators correctly determined the motion direction as a function of the probability of dark dots. (*B*) The correlation between estimated and true velocities. (*C*) The error reduces upon sequentially adding contributions from $k^{(2)}_{i,i+1}$,$k^{(2)}_{i,i+2}$,and $k^{(3)}_{i,i+1,i+2}$. (*D*) The velocity estimates from Reichardt and three-point correlators (left axis) are small compared with the full estimate (right axis).