PNAS
www.pnas.org

Supplementary Information for

Mechanism for analogous illusory motion perception in flies and humans

Margarida Agrochao[1,*], Ryosuke Tanaka[2,*], Emilio Salazar-Gatzimas[2], Damon A. Clark[1,2,3,4]

1 - Department of Molecular Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA
2 - Interdepartmental Neuroscience Program, Yale University, New Haven, CT 06511, USA
3 - Department of Physics, Yale University, New Haven, CT 06511, USA
4 - Department of Neuroscience, Yale University, New Haven, CT 06511, USA
* Equal contributors

Email: damon.clark@yale.edu

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Movies S1 to S3

www.pnas.org/cgi/doi/10.1073/pnas.2002937117

Methods

Fly strains and husbandry

All flies (*Drosophila melanogaster*) were grown at near 50% relative humidity on a dextrose-based food, and only female flies were used for experiments. Flies for behavioral experiments were grown at 20°C in a 12-hour light/dark cycle. They were staged on CO_2 12-24 hours after eclosion and experiments were conducted at 50% humidity between 12 and 24 hours of staging, either within 3 hours after lights-on or within 3 hours prior to lights-off. Female flies used for the behavioral experiments were staged together with male flies. Flies for imaging experiments were either grown at 25°C (*+; UAS-GcaMP6f; R42H07-GAL4/+*) or 29ºC (all the other genotypes). Flies for imaging were staged 12-24 hours after eclosion on CO_2 or on ice. Flies staged on CO_2 were imaged at least 12 hours post-staging. Flies were typically imaged at between 1 to 2 days after eclosion. In 3 flies used in **Fig. S2B**, we used a sparsening strategy with a flip-out GAL80 line. In those cases, flies were heat-shocked for 10 minutes at 37ºC either at the third instar larval stage or after eclosion to allow sparse labeling of neurons (1).

We used a wild type strain previously reported (2) as the parental strain to generate Gal4-only control flies. Gal4 lines (*+;+;R42F06-GAL4; w^-;+;R42H07-GAL4* (3)) and split Gal4 lines (*w^-;R59E08-AD;R42F06-DBD* (4); *w^-;VT015785-AD;R42F06-DBD* (5); *w^-; VT055812-AD; R47H07-DBD* (6)) used have been reported previously. Enhancer-less GAL4 (*w^-; +; pBDPGAL4Uw* (7)) and split GAL4 lines (*w^-; pBPp65ADZpUw*; *pBPZpGAL4DBDUw* (8)) were used for controls in behavioral experiments. The effector line (*+;UAS-shi^{ts};UAS-shi^{ts}* (9)), the reporter line (*+;UAS-GCaMP6f;+* (10)), and lines for flip-out Gal80 (*w^-,y,hsFLP;+;+* and *w^-;+;tub>FRT.GAL80*) have been also reported previously (1).

Genotypes of flies used in behavioral experiments are as follows:
w^-/+; +; +/ pBDPGAL4 (wild type, **Fig. 1D, S1**),
w^-/+; UAS-shi^{ts}/R59E08-AD; UAS-shi^{ts}/R42F06-DBD (T4T5>shi^{ts}, **Fig. 1E**),
w^-/+;UAS-shi^{ts}/VT015785-AD;UAS-shi^{ts}/R42F06-DBD (T4>shi^{ts}, **Fig. 5E**),
w^-/+; UAS-shi^{ts}/+;UAS-shi^{ts}/ R42H07-GAL4 (T4>shi^{ts}, **Fig. 5D**),
w^-/+; R59E08-AD/+; R42F06-DBD/+ (T4T5/+, **Fig. 1E**),
w^-/+; VT015785-AD/+; R42F06-DBD/+ (T4/+, **Fig. 5E**),
w^-/+; +; R42H07-GAL4/+ (T5/+, **Fig. 5D**),
w^-/+; UAS-shi^{ts}/+; UAS-shi^{ts}/pBDPGAL4 ($shibire^{ts}$/+ (single) , **Fig. 5D**),
w^-/+; UAS-shi^{ts}/BPp65ADZp; UAS-shi^{ts}/BPZpGDBD ($shibire^{ts}$/+ (split), **Figs. 1E, 5E**).

We used F1 hybrids of the previously documented wild type strain (2) with the enhancer-less Gal4 line (7) as our 'wild type' flies in the behavioral experiments (**Fig. 1D, S1**), so that their genetic background matched as closely as possible with the flies used in the other experiments. Note that the effector line (*+;UAS-shi^{ts};UAS-shi^{ts}*) was back-crossed into the same parental wild type strain.

Genotypes of the flies used in the imaging experiments (**Figs. 2, 3, S2, S3**) are as follows:
+; UAS-GcaMP6f/+; R42F06-GAL4/+ (T4T5>GCaMP6f),
+; UAS-GcaMP6f; R42F06-GAL4/+ (T4T5>GCaMP6f),
w^-/+; UAS-GcaMP6f/R59E08-AD; R42F06-DBD/+ (T4T5>GCaMP6f),
+; UAS-GcaMP6f; R42H07-GAL4/+ (T5>GCaMP6f),
w^-/+;UAS-GcaMP6f/VT055812-AD; R47H05-DBD/+ (T5>GCaMP6f).
In some of the experiments, few flies with *hsFLP* and *tub>FRT.GAL80* were used to sparsen the expression of GCaMP6f and improve separation between T4 and T5 neurons (1).

Fly visual stimulus presentation

The stimuli were projected on to the panoramic screens surrounding the flies using Lightcrafter DLPs (Texas Instruments, USA) using monochrome green light (peak 520 nm and mean intensities of 100 cd/m^2 for behavioral experiments, and ~70 cd/m^2 for imaging experiments). Stimuli were projected such that the fly experienced a virtual cylinder (11, 12). This cylinder was vertically oriented for the behavioral experiments and pitched forward by 45º to roughly match the head pitch in the imaging experiments (11, 12).

Fly behavioral experiments

We used a fly-on-a-ball rig to measure flies' turning behavior in response to visual stimulation as described in previous studies (11–14). Each fly was anesthetized on ice, fixed to a needle with UV-cured epoxy and placed above an air-suspended ball that rotated under them as they walked and turned. Rotation of the ball was recorded at 60 samples/sec at a resolution of ~0.5° using an optical mouse sensor. The effective coverage of panoramic screens surrounded the fly was 270° azimuth and 106° elevation.

Two types of visual stimuli were presented: horizontal stationary sawtooth gradients given as

$$C(\theta) = \pm\left(2\frac{\mathrm{mod}(\theta,\lambda)}{\lambda} - 1\right)$$

where $\theta \in [0°, 360°], \lambda = 45°$ (**Figs. 1, 5, S1**), and moving sawtooth gradients given as

$$C(\theta, t) = \pm\left(2\frac{\mathrm{mod}(\theta + \omega t,\lambda)}{\lambda} - 1\right)$$

where $\theta \in [0°, 360°], \lambda = 45°, \omega = 0, \pm 2, \pm 4, \pm 6°/\mathrm{s}$ (**Fig. S1**). The spatial phase was randomized every presentation. Each presentation of stimuli lasted for 5 seconds and was interleaved with 2 seconds of mean gray.

To estimate the mean turning response to visual stimuli, the turning time traces were first averaged across every presentation of each stimulus type within each individual fly ("individual mean response"). Responses to mirror-symmetric pairs of stimuli were subtracted at this step to obtain

the net response to each stimulus type. The group mean responses and their standard errors were then calculated from the individual mean responses. To compare responses to different visual stimuli, the individual mean responses were temporally averaged over the stimulus presentation window ("averaged individual response"). The group means of the averaged individual responses and their standard error were then calculated. Positive responses indicate rightward turning.

To judge whether there was a difference between the magnitude of the effect on turning of silencing T4 neurons and silencing T5 neurons (**Fig. 5**), we calculated the following z-score:

$$z = \frac{\Delta T4 + \Delta T5}{\sqrt{{\sigma_{T4_{silenced}}}^2 + {\sigma_{T4_{controls}}}^2 + {\sigma_{T5_{silenced}}}^2 + {\sigma_{T5_{controls}}}^2}}$$

where

$$\Delta T4 = \text{mean}(T4_{silenced}) - \text{mean}(T4_{controls})$$

Here, $T4_k$ is the turning of $T4_{silenced}$ or $T4_{control}$ flies, σ_{T4_k} is the SEM of $T4_k$ flies. Similar definitions apply for T5 neurons. The summation in the numerator is because the effects are in opposite directions, and we wish to ask if they have different magnitudes. The value for z was 2.14, which corresponds to a two-tailed p-value of 0.03. Thus, the magnitude of the effect of silencing T4 neurons is significantly larger than that of silencing T5 neurons.

To estimate the nulling velocity that yields no turning in flies, we presented moving sawtooth gradients with $\omega = 0, \pm 2, \pm 4, \pm 6$ °/s, found the best linear fit to the data of turning velocity versus stimulus velocity, and computed the x-intercept of that best fit. To find the confidence interval for the nulling velocity, we created 1000 bootstrap replicates of the data. For each bootstrap replicate, we found the best linear fit, and computed the nulling velocity. We then computed the 68% confidence intervals on the nulling velocity from that distribution of bootstrap nulling velocities.

Calcium imaging

Neuronal activity at axon terminals of T4 and T5 neurons was recorded using two-photon scanning fluorescence microscopy as described previously (12–14). Flies expressing GCaMP6f in T4 and/or T5 neurons were anesthetized on ice and head-fixed in a stainless-steel shim. To expose the optic lobe, a surgery was performed on the back of the flies' head to remove the cuticle, fat, and trachea above the optic lobe. The exposed brain was submerged in oxygenated sugar-saline solution (15). Imaging was performed with two-photon microscopes (Scientifica, UK) equipped with a 1.05 NA 20x water immersion objective lens (XLUMPlanFL; Olympus, Japan), while the flies were presented with visual stimuli. Stimuli were projected such that the fly experienced a virtual cylinder, which was pitched forward 45° to account for the angle of the head of the fly on the shim. The resulting effective coverage of the screens was 270° azimuth and 69° elevation. The output of the DLP was filtered with two 565/24 filters (Semrock, NY, USA). The input to PMT was also

filtered with a 514/30 in series with a 512/25 filter to block photons from visual stimulus (Semrock, NY, USA). A precompensated eHP MaiTai femtosecond laser (Spectraphysics, CA, USA) provided 930 nm light with power < 30 mW at the sample. Images were acquired at ~13 Hz using ScanImage software (16). The sequence of images acquired was processed off-line to compensate for motion of the brain tissue during imaging.

In each recording session, a set of experimental stimuli was presented along with probe stimuli, which were later used to select responsive structures and to identify cell types (i.e., T4/T5 neurons and their subtypes). The experimental stimuli used were stationary single bars (5º wide stationary vertical bars, either white or black, repeated over horizontal visual space, presented for 1 second) (**Fig. 2D**), stationary sawtooth gradients (the same gradients as ones used in behavioral experiments but with the period of 80º, presented for 1 second) (**Fig. 2E**), stationary square wave gratings (full-contrast horizontal square wave grating with the periods of 80º or 50º, presented for 1 or 5 seconds, respectively) (**Fig. 2F, S2A**), half-contrast stationary square wave gratings (horizontal square wave gratings with 50º period, contrast ranging either from -1 to 0 or 0 to +1, presented for 1 second) (**Fig. 3A**), bar pairs (pairs of 5º wide stationary black or white vertical bars, separated by 10º, repeated every 40º, presented for 1 second) (**Fig. 3B**), and naturalistic contrast patterns (see below) (**Fig. S2B**). The stationary stimuli were presented at all phase configurations with incremental displacements of 5º. Stationary single bar stimuli with appropriate spacing were presented in every experiment to locate receptive fields. The probe stimuli consisted of moving edges (vertical white or black edge moving horizontally on black or white backgrounds, respectively, at 30º/s in each direction) and moving square wave gratings (full contrast square wave gratings with 30º period moving either horizontally or vertically at 30º/s).

Preprocessing of imaging data

After aligning the acquired images to compensate the motion of the sample, frames that moved more than 5 microns from the target were excluded from the analysis. Entire recordings were also discarded if more than 5% of their frames were excluded due to excessive displacement.

Regions of interest (ROIs) were extracted based on a local coherence-based method (17). First, for each pixel in a movie, we calculated the Pearson correlation of its background-subtracted intensity values with those of its 8 neighboring pixels during the probe presentation. The correlations to the 8 neighbors were then averaged for each pixel to generate a single "correlation image". We applied a watershed segmentation algorithm (18) to this correlation image to obtain spatially contiguous clusters of coherently fluctuating pixels as ROIs. Only ROIs found within appropriate anatomical structures (e.g. specific neuropils) were retained. In the "bar pairs" experiments (**Fig. 3B**), an independent component analysis algorithm was performed on the responses to the probe stimuli to find ROIs, as detailed previously (12, 14).

For each ROI, $\Delta F/F$ was calculated as the proxy of intracellular calcium concentration as previously described (14). First, we averaged the time traces of the pixels within each ROI. Then,

to estimate the baseline fluorescence of the ROIs while accounting for the photobleaching of the sensors, a decaying exponential of the form $Ae^{-t/\tau}$ was fitted to the portion of the fluorescence time trace that corresponded to interleaves between stimulus presentations. A single τ was computed to best fit the average fluorescence decay across all ROIs in each movie, while A was individually fit to the mean amplitude of each ROI. This exponential timeseries served as the $F_0(t)$, and the original fluorescence time trace as $F'(t)$, so that $\Delta F/F$ was computed as

$$\frac{\Delta F}{F} = \frac{F' - F_0}{F_0}$$

In the following, the responses were all analyzed in terms of $\Delta F/F$.

ROI selection and cell type identification

ROIs were selected based on the consistency of their response during the probe presentation. The probe stimuli were presented twice before the experimental stimuli and once after. For each ROI, the correlation between responses to the first and second probe presentations ($r_{1/2}$) was computed and thresholded to ensure ROIs had consistent responses ($r_{1/2}$>0.4). A correlation was computed between the average response to the first two probes and that of the third probe ($r_{12/3}$), to ensure responsivity remained unchanged from the beginning to the end of the stimulus ($r_{12/3}$>0.4). ROIs where either of these correlations was smaller than the threshold were excluded from further analysis. A small subset of stimulus presentations did not have a third probe stimulus presentation, and for these flies this threshold was ignored.

Axon terminals of T4 and T5 neurons are interdigitated and separable based on their response properties. Therefore, we further selected ROIs based on the unambiguity of their cellular identity (*i.e.*, T4 vs. T5 neurons, and their subtypes). As previously detailed (14), flies were presented with a probe stimulus consisting of vertical light or dark edges moving horizontally (*i.e.*, front-to-back, FTB, or back-to-front, BTF) as well as square wave gratings moving in the four cardinal directions. The moving square wave gratings were used to ensure that selected ROIs maximally responded to horizontal (FTB or BTF), rather than vertical (up or down) motion. In addition, based on the responses to the four types of edges, edge selectivity index (ESI) and direction selectivity index (DSI) were calculated as

$$ESI = \frac{\Delta R_{light} - \Delta R_{dark}}{\Delta R_{light} + \Delta R_{dark}}$$

and

$$DSI = \frac{\Delta R_{FTB} - \Delta R_{BTF}}{\Delta R_{FTB} + \Delta R_{BTF}},$$

where $\Delta R_{light} = (\Delta R_{light,FTB} + \Delta R_{light,BTF})/2$ and $\Delta R_{prog} = (\Delta R_{light,FTB} + \Delta R_{dark,BTF})/2$ (and similarly for ΔR_{dark} and ΔR_{BTF}). The ΔR value for each edge type was calculated as the difference between the 99th and 50th percentile values of the response time trace averaged over two presentations. ROIs were selected based on the following thresholds on the selectivity indices (12, 14): ESI>0.3 for T4 neurons, ESI<–0.4 for T5 neurons, DSI>0.4 for FTB, DSI<–0.4 for BTF. ESI- and DSI-based ROI selection was done using only the pre-experiment probe presentations. In **Fig. S2B**, in which stimulus presentation lasted for more than 18 minutes, the ESI- and DSI- based selection was repeated again using the post-stimulus probe presentations.

To further ensure the cell-type identities of ROIs, a polarity selectivity index (PSI) was calculated based on the responses to stationary bars, except for the analysis of the "bar pairs" experiments (**Fig. 3B**). Stationary bars with either positive (white) or negative (black) contrast and with various spatial phase configurations were presented among experimental stimuli, and for each ROI, PSI was calculated as

$$PSI = \frac{\Delta S_+ - \Delta S_-}{\Delta S_+ + \Delta S_-},$$

where ΔS_+ and ΔS_- were difference between time-averaged responses at maximally and minimally stimulating spatial phase configurations of either positive (+) or negative (-) contrast bars, respectively. ROIs were selected based on the thresholds of PSI>0.4 for T4 neurons and PSI<-0.4 for T5 neurons.

Receptive field mapping and response alignment

Receptive field locations were identified using the responses to either bright (T4 neurons) or dark (T5 neurons) bars. The stationary bars stimulus contained 5º wide bars repeated either every 40º (**Fig. 3B**), 50º (**Fig. 3A, S2A**), 80º (**Figs. 2E** and **F**), or 90º (**Fig. S2B**) of visual space, allowing corresponding either 8, 10, 16, or 18 non-overlapping phase configurations with 5º phase increments. Responses of each ROI were averaged across repetitions and integrated over the presentation period (1 second) for each phase, resulting in a spatial tuning curve across the eight phases. The phase that gave rise to the peak of the spatial tuning curve was considered to be the receptive field center, and all other stationary, periodic stimuli were aligned to this receptive field center.

Naturalistic contrast patterns

For presentation of naturalistic contrast patterns (**Fig. S2B**), one natural 360º panorama was selected from a database (19) for its many edges and high contrast. A single horizontal line was chosen from the panorama, and then repeated vertically, creating a 1-dimensional 'naturalistic contrast pattern', as in previous work (14). The naturalistic contrast pattern was presented at 72 all possible phase configuration with 5º phase increments. In each imaging session, the contrast pattern was presented twice at each of a half of the all possible phase configurations (*i.e.*, at 36

phases). The receptive field center of each ROI was located using stationary bar stimuli as described above, but with bars with 90º spacing (= 5º phase increments x 18 phase configuration). The period of 90º was enough to uniquely determine the location of the receptive field of an ROI, because we only imaged the portion of the lobula plate that corresponds to the frontal visual field, based on anatomy and responses to moving edge stimuli. The responses of ROIs to different positions in the natural image stimuli were aligned using the estimated receptive field center locations, and averaged over ROIs within each cell type (T4 and T5 subtypes). Note that some flies were only presented with one of the two halves of the all possible phase configurations.

Models of motion detectors

We performed numerical simulation of how different models of elementary motion detectors respond to our stationary stimuli (stationary sawtooth gradients, full- and half-contrast stationary square wave gratings, pairs of bars) as well as to half-contrast drifting sinusoidal gratings with various spatiotemporal frequencies. The visual stimuli were simulated at the spatiotemporal resolution of 0.1º and 240 Hz.

The Barlow-Levick model (20) (**Fig. S5A**) had two input branches spaced 5º apart. Each branch had a Gaussian spatial filter h with full-width half-maximum (FWHM) of 5.7º. The first arm was then convolved with a temporal low-pass filter:

$$f_1(t) = \begin{cases} kte^{-t/\tau} & (t \geq 0) \\ 0 & (t < 0) \end{cases}$$

and the second branch with a band-pass filter:

$$f_2(t) = \begin{cases} 0.2f_1(t) + k(\tau - t)e^{-\frac{t}{\tau}} & (t \geq 0) \\ 0 & (t < 0) \end{cases}$$

where τ = 100 ms. k was fixed so that $f_1(t)$ has unit L_2 norm. A small, low-pass term ($0.2f_1(t)$) was added to the temporal filter on the second arm to sustained responses to stationary stimuli. The response of the model was given as

$$r_{\mathrm{BL}}(t,x) = R\left(R\big((f_1 h * c)(t,x)\big) - 2R\big((f_2 h * c)(t,x+5°)\big)\right)$$

where $c(t,x)$ is the input contrast, positive half-wave rectification is denoted as $R(\cdot)$, and spatiotemporal convolution is denoted by $*$.

The response of motion energy model (21) (**Fig. S5B**) was given as

$$r_{\mathrm{ME}}(t,x) \coloneqq R^2\big((k * c)(t,x)\big)$$

where spatiotemporal filter $k(x,t)$ is defined as:

$$k(x,t) = h_1(x)f_1(t) + h_2(x)f_2(t)$$

where $f_1(t)$ and $f_2(t)$ are the same temporal filters used in the Barlow-Levick model. $h_1(x)$ and $h_2(x)$ are odd and even one-dimensional spatial Gabor filters given as

$$h_1(x) = C_1 \exp\left(-\frac{x^2}{2\sigma^2}\right) \sin\left(\frac{2\pi x}{\lambda}\right)$$

$$h_2(x) = C_2 \exp\left(-\frac{x^2}{2\sigma^2}\right) \cos\left(\frac{2\pi x}{\lambda}\right)$$

where $\sigma = 2.42°$, giving a Gaussian envelope with a FWHM of 5.7º, and we fixed the carrier wavelength $\lambda = 22.8°$. C_1 and C_2 were fixed such that the kernels had unit L_1 norm.

The synaptic model of T4 neurons (**Fig. 4A**, *top*) (22, 23) had three input branches that were spatially separated by 5º. Each branch had a Gaussian spatial filter (FWHM = 5.7º) followed by temporal low-pass (flanking branches) or high-pass (center branch) filtering, which were respectively identical to $f_1(t)$ and $f_2(t)$ used in the Barlow-Levick model. The sign of the output of the first branch was inverted, and then the outputs of all the branches were half-wave rectified. The polarity preferences of the three branches mimic the primary input neurons of T4 neurons (24, 25). The pseudo-steady state membrane voltage of a T4 neuron, $V_m(t)$, was modeled by neglecting capacitive currents (5, 26), such that

$$V_m(t) = \frac{g_1(t)V_i + g_2(t)V_e + g_3(t)V_i}{g_{leak} + g_1(t) + g_2(t) + g_3(t)}.$$

Here, $g_1(t), g_2(t), g_3(t)$ correspond to the outputs of the three input branches, weighted by 0.2, 0.1, 0.2, respectively, $g_{leak} = +1$, $V_i = -30$ mV, and $V_e = +60\ mV$. Note that V_m is defined such that reversal potential for leak current is 0 mV. Finally, the transformation from the membrane voltage to calcium responses were modeled as a half-wave rectifying quadratic nonlinearity.

We modified this three-input model of T4 neurons (**Fig. 4A**, *bottom*) to recapitulate the responses of T4 neurons to stimuli that do not involve contrast increments (**Fig. 3A**). A constant bias term of +4 was added to the third branch after spatiotemporal filtering and before rectification, and a constant bias term was added to $V_m(t)$ so that $V_m(t) = 0\ mV$ when there is no input. The increased baseline activity of the third arm was introduced to enable the model T4 neuron to be disinhibited by the presence of OFF contrast on one side of the receptive field (**Fig. 4C**). Previous physiological measurement has shown that Mi4, an inhibitory neuron that corresponds to the third arm in this model, is inhibited by OFF contrasts in addition to being excited by ON contrasts (14).

The T5 model (**Fig. S4A-D**) had the same architecture as the unmodified T4 model. The differences from the T4 model are as follows: (a) the output of all input branches were sign-inverted before the half-wave rectification (rather than just the null side branch in the T4 model), (b) when converting the output of the branches to conductance $g_1(t), g_2(t), g_3(t)$, weightings of

0.05, 0.1, 0.2 were used (instead of 0.2, 0.1, 0.2), and (c) when calculating $V_m(t)$, $g_1(t)$ was multiplied with V_e, rather than V_i.

Human psychophysics

Twelve volunteers including two of the authors participated in the human psychophysics experiment (5 male, 7 female, ages 23-35 years). One participant was excluded from the analysis for not perceiving any illusory motion without adaptation. All participants provided written informed consent and were compensated for participation in the experiment. The experimental protocol was approved by the Institutional Review Board of Yale University and was in accordance with the Declaration of Helsinki. All participants had normal or corrected normal vision.

Visual stimuli were generated using Psychtoolbox 3 (27) and presented on a 27-in. LED monitor (1920 x 1080 pixels, 60 Hz refresh rate). The mean luminance was ~300 cd/m^2 and the viewing distance was maintained at 43 cm using a chin rest. Participants observed the stimuli binocularly in a dark room. All visual stimuli were gamma-corrected in software.

Adaptor stimuli (**Movie S3**) were presented within a ring-shaped aperture with retinal eccentricity of 11º (at the center of the annulus) and thickness of 4º. The aperture contained either uniform mean gray, moving dark edges, or moving light edges. The moving edges had a central angle of 10º and temporal frequency of 2 Hz, resulting in the retinal angular velocity of ~3.8º/s. The directions of the edges were inverted and the phase of the edges were randomized every temporal period. The goal was to expose every retinotopic location to equal amounts of clockwise and counterclockwise motion and thus to prevent directional motion after effects. The sawtooth gradient had eccentricity of 11º, thickness of 2º, and central angle of 10º, chosen from previous psychophysical experiments (28). The phase and direction of the gradient was randomized in every presentation. The gradient rotated about the fixation point at a constant velocity determined on trial-by-trial basis as described below. A central fixation point was always presented during the presentation of adaptor or sawtooth gradient stimuli. All visual stimuli were spatially low-pass filtered with a Gaussian kernel with a standard deviation of 0.021º, and presented against a field of static random dots of single-pixel resolution with equal probabilities of black and white.

The experiment consisted of three blocks, and a single type of adaptor stimuli out of three was presented in each block. Each block started with a minute-long presentation of an adaptor stimulus, followed by 100 presentations of sawtooth gradients. The sawtooth gradient was presented for 500 ms, and each presentation of sawtooth gradients was preceded by 3 s presentation of a top-up adaptor stimulus, except for the first trial of each block. After each presentation of the sawtooth gradient, the subject made a two-alternative forced choice by pressing a key to indicate the perceived direction of the rotation of the gradient (either clockwise or counterclockwise). For each adaptor type, a logistic psychometric curve given as

$$p(\omega) = \frac{1}{1 + \mathrm{e}^{-k(\omega - T)}}$$

was fitted to the observer's responses, where $p(\omega)$ was the probability that a motion toward the lighter shade along the shallow gradient is perceived, and ω was the rotational velocity of the sawtooth gradient (positive ω indicates rotation toward the lighter shade along the shallow gradient). The psychometric curve was updated every trial using a Bayesian adaptive method. As a joint prior distribution of threshold T and log base 2 of slope k, we assumed separable normal distributions

$$T \sim \mathcal{N}(\mu_T, \sigma_T)$$
$$\log_2 \frac{k}{k_0} \sim \mathcal{N}(\mu_k, \sigma_k)$$

where $k_0 = 1$ s/° and $\mathcal{N}(\mu, \sigma)$ is a normal distribution with mean μ and standard deviation σ. The prior means were μ_T = –0.3°/s and μ_k = 3, and standard deviations were σ_T = 1 °/s and σ_k = 3.5, based on the results of preliminary experiments. In each trial, the rotational velocity of the gradient was determined to minimize expected entropy of the parameter distribution (29). The means of the posterior distributions were reported as point estimates of the threshold and slope, and errors around the individual psychometric curve correspond to 68% credible intervals. Additionally, we calculated the probability that the posterior distributions of the threshold from each pair of adaptor conditions differ within each individual observer as

$$p_{\mathrm{diff}} = 2\,|p(T_1 > T_2) - 0.5|$$

where

$$p(T_1 > T_2) = \int_{-\infty}^{\infty} p(T_1) \int_{-\infty}^{T_1} p(T_2) dT_2 dT_1$$

and $p(T_1)$ and $p(T_2)$ are marginal posterior distributions for two adaptor conditions. p-values presented indicate 1 – p_{diff}.

Statistics

Each fly or human subject was counted as an independent measurement for statistical purposes. For imaging data, ROIs of a particular cell type within a fly were averaged together to produce each measurement for the fly-cell type combination. For fly data, all p-values presented are a result of a Wilcoxon rank-sum test across flies. For the across subject tests in human experiments, p-values presented are a result of a Wilcoxon signed-rank test. For within subject test of threshold distributions (**Fig. 6C, S7**), the calculated probability that distributions were different was Bonferroni-corrected in each case by multiplying by n = 11.

Data and code availability

The data analyzed in this work are available from the Dryad Digital Repository: https://datadryad.org/stash/share/RQqRF0DX9ZgNSHNOMzikGQcMG7fJXSYtGtaQp6HpaE0 [note: this link will be updated to a public link once the paper is accepted]. MATLAB code to reproduce all analyses and modelling results is available as a GitHub repository at github.com/ClarkLabCode/IllusionPaperCode.

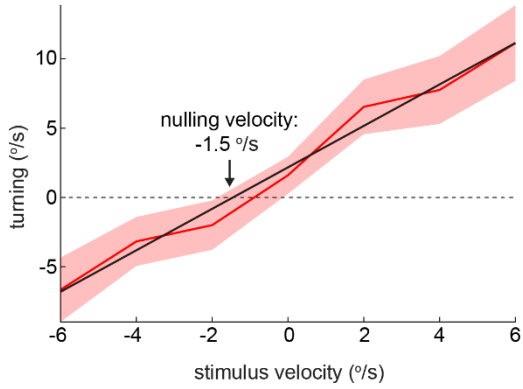

Fig. S1. Estimating the nulling velocity of the sawtooth gradient illusion.

Time-averaged turning response of wild type flies to a 5 s presentation of 45º-period sawtooth gradients moving horizontally at different velocities (mean ± SEM) (*red*), linear regression to the data (*black*) (n = 15 flies). The arrow indicates that the nulling velocity was –1.5 °/s, with a 68% confidence interval of [-2.1, -0.9] °/s, found by bootstrapping over flies (see **Methods**).

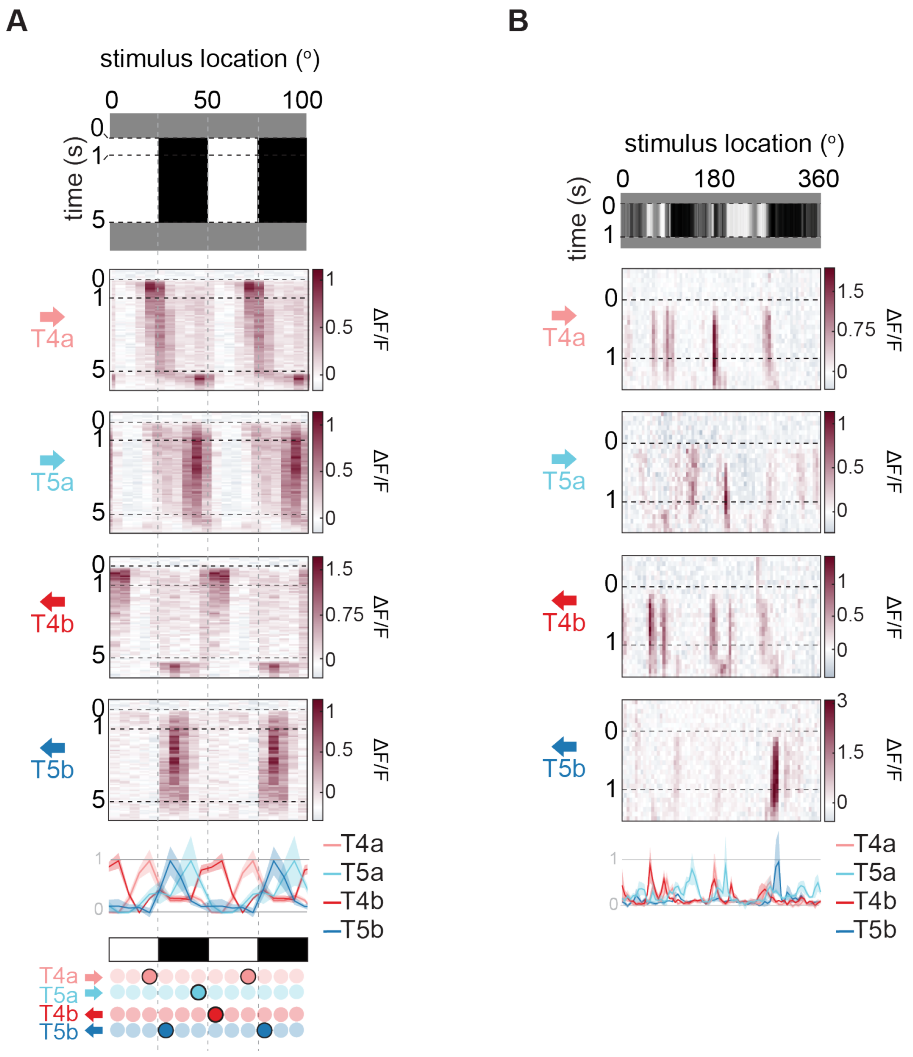

Fig. S2. T4 and T5 neurons show sustained responses to contrast edges and to naturalistic stimuli.

A, Calcium responses of T4a, T5a, T4b, and T5b neurons to 5 s presentation of stationary square wave gratings. On the bottom are time-averaged responses (mean ± SEM) and a summary of the responses, similarly to **Figure 2E, F**. n = 7, 4, 2, 6 flies for T4a, T4b, T5a and T5b neurons. **B,** Calcium responses of T4a, T5a, T4b, and T5b neurons to 1 s presentation of a stationary naturalistic contrast pattern (see **Methods**). Below are time-averaged responses (mean ± SEM). n = 14, 9, 7, 5 flies for T4a, T4b, T5a and T5b neurons.

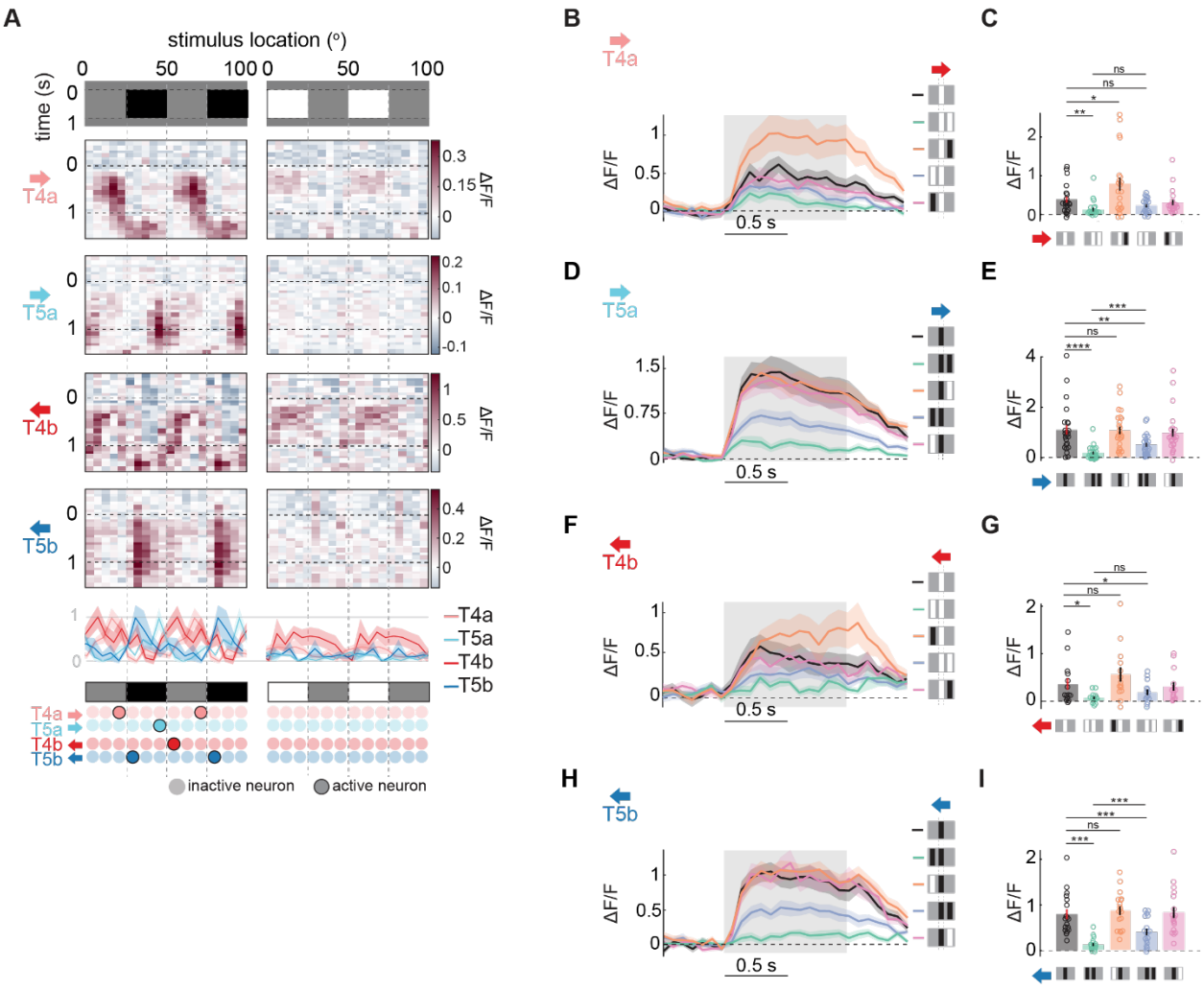

Fig. S3. Responses to stationary edges emerge through different phenomenology in T4 and T5 neurons.

This figure replicates plots in **Figure 3**, but shows T4a, T4b, T5a, and T5b split out. **A,** Calcium responses of T4 and T5 neurons to 1 s presentation of half-contrast square wave gratings, gray-to-black (*left*) or white-to-gray (*right*). Below are time-averaged responses of T4 and T5 neurons over space, normalized within each cell type (mean ± SEM), and a summary of the responses. n = 10, 8, 2, 4 flies for T4a, T4b, T5a, T5b neurons. **B-I,** Calcium response of (**B, C**) T4a, (**D, E**) T5a, (**F, G**) T4b and (**H, I**) T5b neurons to 1 s presentation of stationary single or paired bars. See **Fig. 3** for description of stimuli. **B, D, F, H** Time traces of the calcium responses, and **C, E, G, I** time-averaged responses. All are mean ± SEM. n = 21, 14, 24, 18 flies for T4a, T4b, T5a, T5b neurons. * p≤0.05; ** p≤0.01; *** p≤0.001; **** p≤0.0001 by a Wilcoxon rank-sum test across flies.

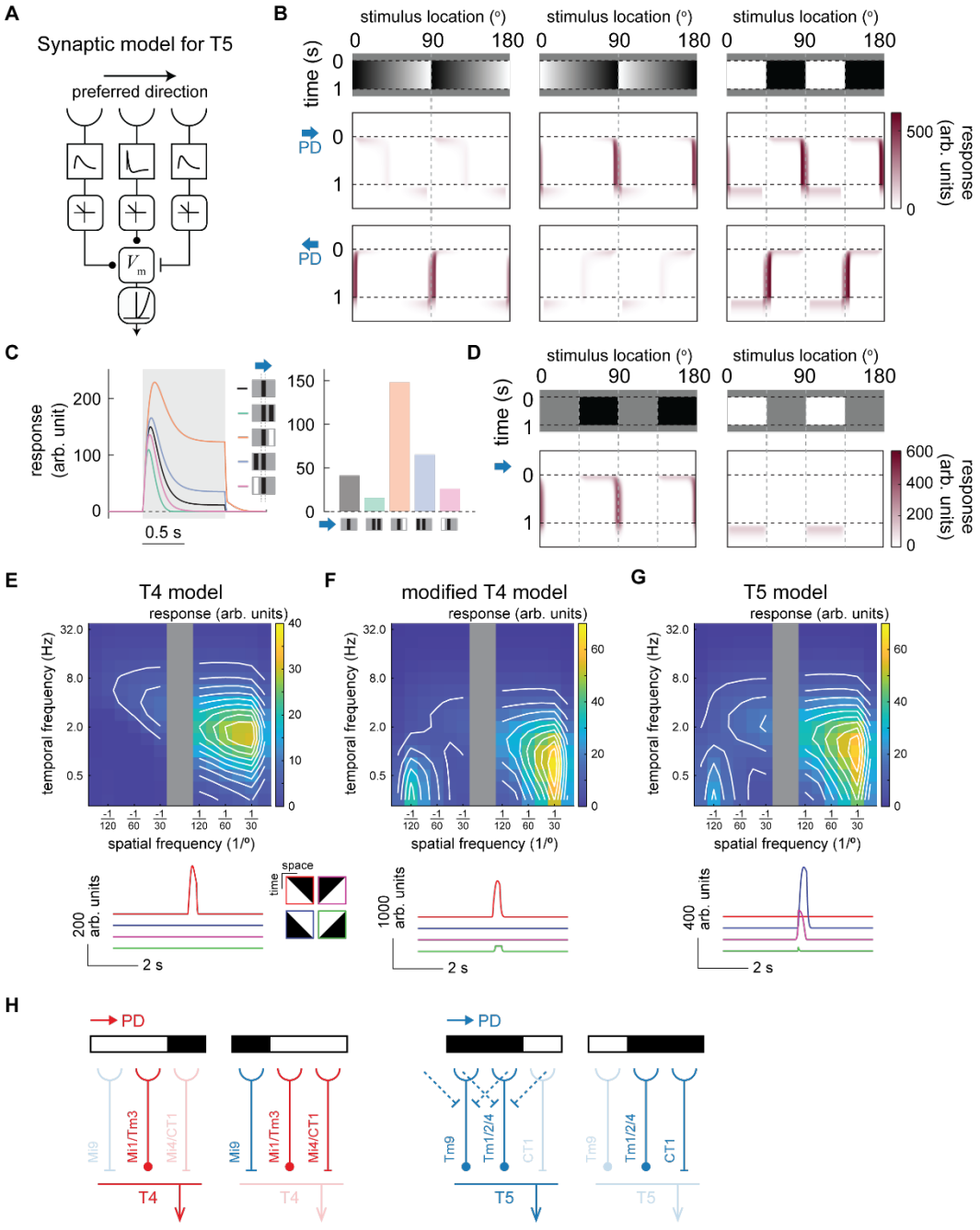

Fig. S4. Characterization of the synaptic models of T4 and T5.

A, A minimal synaptic model of T5 neurons, designed similarly to the model of T4 in **Fig. 4A** (see **Methods**). Unlike the T4 models, the arm on the null side is excitatory, and all the input arms undergo sign inversion (making them OFF cells), consistent with known anatomy and physiology of inputs into T5 neurons (24, 30, 31). **B-D,** The model responses to (**B**) stationary sawtooth gradients and square wave grating, (**C**) pairs of bars, and (**D**) half-contrast square wave gratings (similar to **Fig. 4B-D**). Preferred directions of the models are indicated with the blue arrows. Replicating physiological observations, the model shows stationary edge responses when dark

contrast is on its receptive field center but not on the preferred side (**B, D**), replicating imaging results (**Fig. 2E, 2F, 3A**). On the other hand, the model showed sensitivity to light bars, which we did not observe physiologically (**Fig. 3D, 3E**). This is because the dark bar at the center, after spatial low-pass filtering, excited the two side input arms, effectively serving as baseline excitation which allowed them to be inhibited by light bars. **E-G,** Characterization of (*top*) spatiotemporal frequency tuning and (*bottom*) moving edge selectivity in the (**E**) original and (**F**) modified models of T4 and (**G**) the model of T5. The spatiotemporal frequency tuning was probed with half-contrast drifting sinusoidal gratings with various spatiotemporal frequency tuning (see **Methods**). Positive spatial frequencies correspond to motion in the preferred direction. The white contours indicate iso-response lines. Overall, the models showed realistic-looking frequency tuning peaking at 1 Hz of temporal frequency (13). The T4 models showed selective response to a light edge moving rightward (i.e. preferred direction) (**E, F**), and the T5 model to a dark edge moving rightward (**G**), while it also showed smaller response to a light edge moving leftward. **H**, Schematic circuit diagrams of input neurons into T4 and T5 potentially contributing to the stationary edge responses. ON and OFF cells are indicated by red and blue, respectively. Excitatory and inhibitory synapses are represented as filled circles or T-junctions, respectively. Inactive cells are grayed out. Blue dotted line indicates hypothetical lateral inhibition happening at the level of T5 inputs, which could be responsible for lowered activity in T5 when a dark bar is present at the null side of the receptive field (**Fig. 3D, 3E**).

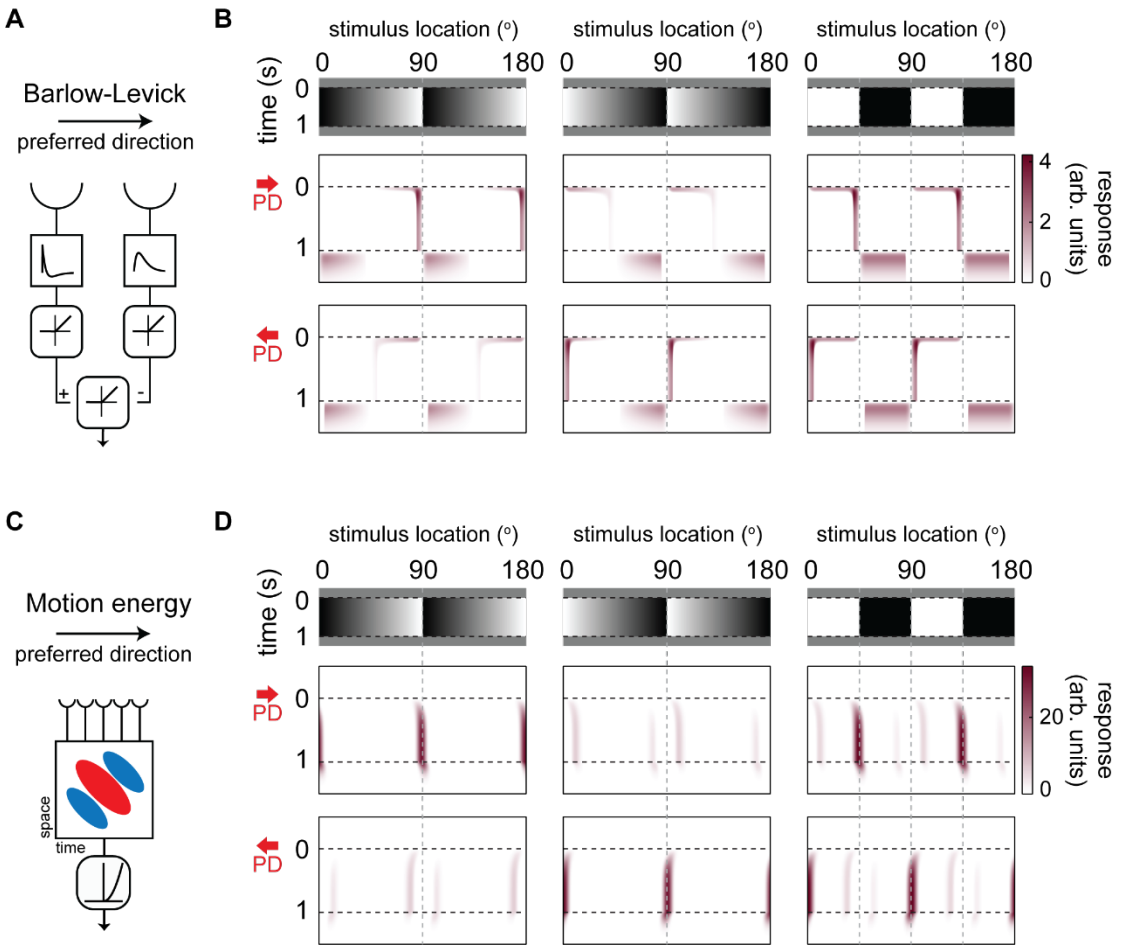

Fig. S5. Biologically plausible models of motion detection in different visual systems exhibit responses to stationary contrast edges.

Schematics of the numerical models of motion detectors (**A, C**) and their responses to stationary sawtooth gradients and stationary full-contrast square wave grating (**B, D**). The responses of the models to stationary stimuli (**B, D**) are shown as space-time plots, as in **Fig. 2**. The preferred directions of the models are indicated with red arrows. The strong responses of the both models to the stationary stimuli were confined about the sharp contrast edges. Specifically, the models responded when light contrast was presented on their receptive field center and dark presented on the preferred side, paralleling the response pattern of T4 neurons (**Figure 2D, E**). Note that the sign of the filters were arbitrarily chosen such that their responses are similar to those of T4 rather than T5, but one can simply invert them to make the models generate more T5-like responses. **A,** The Barlow-Levick model (20) consists of two spatially separated input branches, each of which undergoes either high- or low-pass temporal filtering followed by half-wave rectification. The output of the second branch is subtracted from the output of first branch, followed by another half-wave rectification. **B,** The motion energy model (21) consists of an oriented spatiotemporal filter followed by a positively rectifying quadratic operation. Note that in both models, 'high-pass' filters had small low-passed component to allow sustained responses in the models (see **Method**).

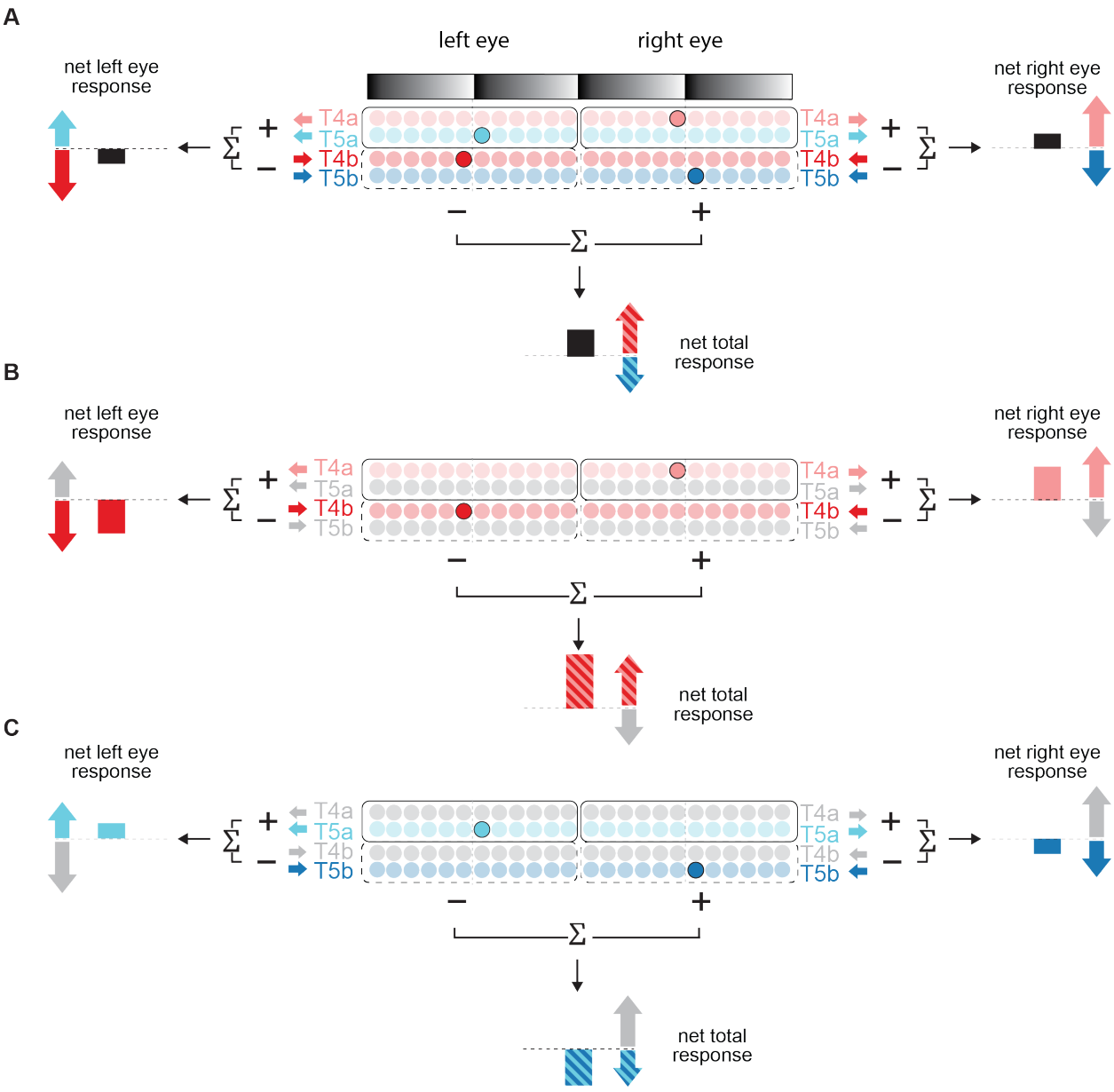

Fig. S6. Mapping the activity of T4 and T5 neurons to turning behavior when taking into account both eyes.

Optomotor turning of flies reflects the difference between net activity of T4 and T5 neuron populations tuned to rightward motion (*i.e.*, T4a and T5a neurons in the right eye, T4b and T5b neurons in the left eye) and leftward motion (*i.e.*, T4b and T5b neurons in the right eye, T4a and T5a neurons in the left eye). This model assumes that, within each eye, behavior sums over space the outputs of T4 and T5 neurons tuned to front-to-back motion (T4a and T5a), and then subtracts the spatially summed outputs of T4 and T5 neurons tuned to back-to-front motion (T4b and T5b). The net motion signal in the left eye is then subtracted from net motion signal in the right eye, resulting in the net turning signal (positive indicates rightward). **A,** When flies are presented with

a sawtooth gradient with contrast increasing to the right, T4a and T5b neurons are active in the right eye, and T5a and T4b neurons are active in the left eye. As a result of subtractions within and across the eyes, activities of right-eye T4a neurons and left-eye T4b neurons contribute positively to the net turning signal, whereas activities of right-eye T5b neurons and left-eye T5a neurons contributes negatively to the net turning signal. In this model, since wild type flies turned rightward to this stimulus, the contribution of T4 neurons to the net turning signal slightly outweighs that of T5 neurons. **B,** By genetically silencing T5 neurons, the net motion signal becomes more positive in the right eye and more negative in the left eye, resulting in enhanced rightward turning. **C,** By genetically silencing T4 neurons, the net motion signal becomes negative in the right eye and positive in the left eye, resulting in reversed, leftward turning.

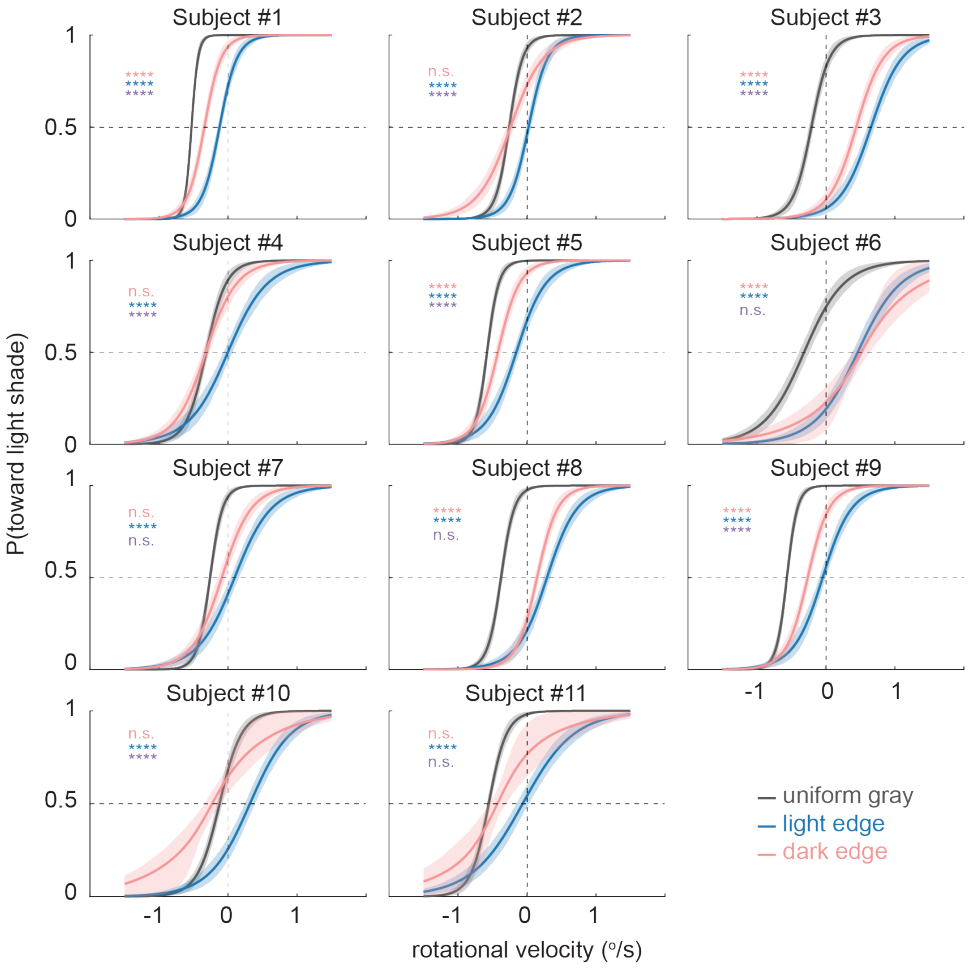

Fig. S7. Psychometric curves for individual subjects.

Psychometric curves for the all subjects (n = 11) by adaptor conditions (*gray*: uniform gray, *red*: dark edge, *blue*: light edge). Each curve shows the estimated probability of the subject seeing rotation towards the light shade along the shallow gradient, as a function of rotational velocity (positive velocity indicates rotation towards the light shade). The curves were drawn using the means of the joint posterior distributions of the threshold and slope of logistic psychometric functions, with 68% credible shaded regions. n.s., non-significant; **** $p \leq 0.0001$, analytically calculated probability that a pair of posterior distributions of illusory velocity are not different (see **Methods**). Blue stars indicate test between uniform gray and light edge conditions, pink between uniform gray and dark edge, and violet between light and dark edge.

Movie S1. Stationary sawtooth gradients induce the perception of motion.

Stationary sawtooth gradients induce the perception of motion (32) when viewed peripherally. With the gradients oriented with the light shade towards the right, the perception of motion is clockwise. The perception of motion and it is reversed with the gradients oriented with the light shade towards the left. The illusion is sustained and it is refreshed when the pattern is flashed in different locations in the visual field.

Movie S2. Each T4 and T5 neuron type responds most to the stationary version of its preferred moving edge.

This movie shows the preferred moving edge of each T4 and T5 cell type, followed by the preferred stationary edge.

Movie S3. Adaptor stimuli used in the human psychophysics experiment.

Human subjects were first exposed to ring-shaped adaptor stimuli, followed by 0.5 s of stationary, sawtooth gradients, on which they made a decision of the direction of rotation (clockwise or counterclockwise) by pressing a key. Adaptor stimuli contained either mean gray, light moving edges, or dark moving edges.

SI References

1. M. D. Gordon, K. Scott, Motor Control in a Drosophila Taste Circuit. *Neuron* **61**, 373–384 (2009).
2. D. M. Gohl, *et al.*, A versatile in vivo system for directed dissection of gene expression patterns. *Nat. Methods* **8**, 231–237 (2011).
3. M. S. Maisak, *et al.*, A directional tuning map of Drosophila elementary motion detectors. *Nature* **500**, 212–6 (2013).
4. T. Schilling, A. Borst, Local motion detectors are required for the computation of expansion Flow-Fields. *Biol. Open* **4**, 1105–1108 (2015).
5. E. Gruntman, S. Romani, M. B. Reiser, Simple integration of fast excitation and offset, delayed inhibition computes directional selectivity in Drosophila. *Nat. Neurosci.* **21**, 250–257 (2018).
6. E. Gruntman, S. Romani, M. B. Reiser, The computation of directional selectivity in the Drosophila OFF motion pathway. *Elife* **8**, 721902 (2019).
7. B. D. Pfeiffer, *et al.*, Refinement of tools for targeted gene expression in Drosophila. *Genetics* **186**, 735–755 (2010).
8. S. Hampel, R. Franconville, J. H. Simpson, A. M. Seeds, A neural command circuit for grooming movement control. *Elife* **4**, 1–26 (2015).
9. T. Kitamoto, Conditional modification of behavior in drosophila by targeted expression of a temperature-sensitive shibire allele in defined neurons. *Dev. Neurobiol.* **47**, 81–92 (2001).
10. T.-W. Chen, *et al.*, Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
11. M. S. Creamer, O. Mano, R. Tanaka, D. A. Clark, A flexible geometry for panoramic visual and optogenetic stimulation during behavior and physiology. *J. Neurosci. Methods* **323**, 48–55 (2019).
12. E. Salazar-Gatzimas, *et al.*, Direct Measurement of Correlation Responses in Drosophila Elementary Motion Detectors Reveals Fast Timescale Tuning. *Neuron* **92**, 227–239 (2016).
13. M. S. Creamer, O. Mano, D. A. Clark, Visual Control of Walking Speed in Drosophila. *Neuron* **100**, 1460-1473.e6 (2018).
14. E. Salazar-Gatzimas, M. Agrochao, J. E. Fitzgerald, D. A. Clark, The Neuronal Basis of an Illusory Motion Percept Is Explained by Decorrelation of Parallel Motion Pathways. *Curr. Biol.* **28**, 3748-3762.e8 (2018).
15. R. I. Wilson, G. C. Turner, G. Laurent, Transformation of Olfactory Representations in the Drosophila Antennal Lobe. *Science (80-.).* **303**, 366–370 (2004).
16. T. A. Pologruto, B. L. Sabatini, K. Svoboda, ScanImage: Flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 1–9 (2003).
17. S. L. Smith, M. Häusser, Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nat. Neurosci.* **13**, 1144–1149 (2010).
18. F. Meyer, Topographic distance and watershed lines. *Signal Processing* **38**, 113–125 (1994).
19. H. G. Meyer, A. Schwegmann, J. P. Lindemann, M. Egelhaaf, Panoramic high dynamic range images in diverse environments (2014) https:/doi.org/10.4119/unibi/2689637.
20. H. B. Barlow, W. R. Levick, The mechanism of directionally selective units in rabbit's retina. *J. Physiol.* **178**, 477–504 (1965).

21. E. H. Adelson, J. R. Bergen, Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* **2**, 284–299 (1985).
22. J. A. Zavatone-Veth, B. A. Badwan, D. A. Clark, A minimal synaptic model for direction selective neurons in Drosophila. *J. Vis.* **20**, 1–22 (2020).
23. B. A. Badwan, M. S. Creamer, J. A. Zavatone-Veth, D. A. Clark, Dynamic nonlinearities enable direction opponency in Drosophila elementary motion detectors. *Nat. Neurosci.* **22**, 1318–1326 (2019).
24. A. Arenz, M. S. Drews, F. G. Richter, G. Ammer, A. Borst, The Temporal Tuning of the Drosophila Motion Detectors Is Determined by the Dynamics of Their Input Elements. *Curr. Biol.* **82**, 887–895 (2017).
25. J. A. Strother, *et al.*, The emergence of directional selectivity in the visual motion pathway of Drosophila. *Neuron* **94**, 168-182.e10 (2017).
26. V. Torre, T. A. Poggio, A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. B Biol. Sci.* **202**, 409–416 (1978).
27. M. Kleiner, D. H. Brainard, D. Pelli, "Whats new in Psychtoolbox-3?" *Percept 36 ECVP Abstr. Suppl.* (2007).
28. R. Hisakata, I. Murakami, The effects of eccentricity and retinal illuminance on the illusory motion seen in a stationary luminance gradient. *Vision Res.* **48**, 1940–1948 (2008).
29. L. L. Kontsevich, C. W. Tyler, Bayesian adaptive estimation of psychometric slope and threshold. *Vision Res.* **39**, 2729–2737 (1999).
30. M. Meier, A. Borst, Extreme Compartmentalization in a Drosophila Amacrine Cell. *Curr. Biol.* **29**, 1545-1550.e2 (2019).
31. K. Shinomiya, *et al.*, Comparisons between the ON- and OFF-edge motion pathways in the Drosophila brain. *Elife* **8**, 1–19 (2019).
32. A. Fraser, K. J. Wilcox, Perception of illusory movement. *Nature* **281**, 565–566 (1979).