Maximally efficient prediction in the early fly visual system may support evasive flight maneuvers

Siwei Wang[1,2], Idan Segev[2], Alexander Borst[3], Stephanie Palmer[1],

1 Department of Organismal Biology and Anatomy, Department of Physics, University of Chicago, Chicago, IL USA
2 Edmond J Safra Center for Brain Sciences, Hebrew University of Jerusalem, Jerusalem, Israel
3 Max Planck Institute of Neurobiology, Martinsried Germany

*sepalmer@uchicago.edu

Abstract

The visual system must make predictions to compensate for inherent delays in its processing. Yet little is known, mechanistically, about how prediction aids natural behaviors. Here, we show that despite a 20-30ms intrinsic processing delay, the vertical motion sensitive (VS) network of the blowfly achieves maximally efficient prediction. This prediction enables the fly to fine-tune its complex, yet brief, evasive flight maneuvers according to its initial ego-rotation at the time of detection of the visual threat. Combining a rich database of behavioral recordings with detailed compartmental modeling of the VS network, we further show that the VS network has axonal gap junctions that are critical for optimal prediction. During evasive maneuvers, a VS subpopulation that directly innervates the neck motor center can convey predictive information about the fly's future ego-rotation, potentially crucial for ongoing flight control. These results suggest a novel sensory-motor pathway that links sensory prediction to behavior.

Author summary

Survival-critical behaviors shape neural circuits to translate sensory information into strikingly fast predictions, e.g. in escaping from a predator faster than the system's processing delay. We show that the fly visual system implements fast and accurate prediction of its visual experience. This provides crucial information for directing fast evasive maneuvers that unfold over just 40ms. Our work shows how this fast prediction is implemented, mechanistically, and suggests the existence of a novel sensory-motor pathway from the fly visual system to a wing steering motor neuron. Echoing and amplifying previous work in the retina, our work hypothesizes that the efficient encoding of predictive information is a universal design principle supporting fast, natural behaviors.

Introduction

Escape is essential for survival. Escape behaviors can take on a variety of forms, from the triggering of a reflexive startle response (e.g., the c-bend escape in zebrafish [1], escape take-offs in flies [2]), to in-flight escape turns, also in flies. These in-flight

maneuvers interrupt ongoing behaviors and thus require continuous sensory processing (e.g. escape from a looming threat during flight [3, 4]). For flying animals, such in-flight escapes emerged as early as flight itself [5]. In this kind of escape, here called an 'evasive maneuver', the organism must adjust its motor program for 'steering away from the threat' dynamically according to both the sensory input of threat and its variable headings from where the animal unfolds the escape. This requires active control. This is because: 1) These escape trajectories must be tuned to both the looming threat location/dynamics and the animal's heading at the moment of detection; 2) The success of such in-flight evasive maneuvers is determined by how well the animal is turned away from the threat at the end of the maneuver, thus how the escape unfolds matters, in addition to its initiation; 3) Flight is aerodynamically unstable. Evasive maneuvers take place while the animal is in-flight, and thus require active control; and finally 4) Active control enables evasive trajectories to be not precisely identical each time the fly reacts, and prevents predators from adapting to a repetitive pattern (e.g. an electric snake can mitigate its prey's predictable, reflexive C-start [6] and thus foil the escape). Here we investigate how sensory information helps to control the finely visual input-tuned in-flight evasive maneuvers at the timescale necessary for survival.

Insects, especially diptera, are excellent models for exploring this problem. The neural architecture of insect brains is highly conserved. It is hypothesized that the neural circuits and the escape behaviors they facilitate co-evolved, as early as flight itself [5]. Many modern arthropod species inherited these core sensory-behavioral modules. In this work, we use *Drosophila* and the blowfly as our animal models. This is because: a) we have precise experimental measurements of motion executed during evasive maneuvers in *Drosophila* [7]; b) there exists experimentally validated, mechanistic-level understanding of neural circuits in the blowfly visual system that may be involved in encoding these sensory inputs; and c) previous experiments suggest that these two species have tight homologies in both their neural architectures and their respective escape behaviors. Both the blowfly and *Drosophila* use banked turns to change their heading direction [7–9]. Although blowflies in general execute higher velocity saccadic turns and they have higher acuity in their compound eyes than *Drosophila*, both animals share similar wingbeat frequencies [10, 11]. This further suggests similar time courses in their escape trajectories. Despite their size difference (a blowfly is roughly four times bigger than a *Drosophila*), their motion sensitive neurons have analogous electrotonic structures [12]. Similarly, their mechanosensory and motor systems are scaled according to this size difference [10, 13–19]. Because the only precise measurements of the fly's evasive maneuver are available in *Drosophila* [20] and the only experimentally validated neural circuit that processes the visual input during these maneuvers is available from the blowfly [21–23], but the two species are so similar overall [24], we investigate how the blowfly's motion sensing circuit extracts behaviorally relevant information based on behavioral measurements from *Drosophila*. This investigation will at least show the minimum necessary computation that the blowfly's neural circuit must execute during fast evasive maneuvers.

Drosophila perform rapid in-flight evasive maneuvers after a sensory-motor delay (e.g., $\approx 60 \pm 20$ ms in [20]). During this delay, after the fly recognizes the looming threat, it processes the location of that threat and initiates an appropriate evasive maneuver. These maneuvers are banked turns, an initial rotation immediately followed by an active counter-rotation. This reorients the animal's heading and accelerates it away from the threat in a mere 40 ms. The motion experience of an animal in such an escape is distinct from motion stimuli investigated in previous work. First, it varies at the millisecond level, whereas recent work [25, 26] on motion encoding focused on constant velocity stimuli. Second, evasive maneuvers can have angular velocities of up to 5300° per second. This is at least five times faster than the saccadic banked turns

investigated in most previous work on flight control during turns [8, 27]. Such high speeds suggest that the entire course of the evasive maneuver requires active control, which is another key difference from escape take-offs that only require initiation [2, 28]. During this brief temporal window, the fly must use sensory information about its own ego-rotation, i.e. $\vec{\theta}$ (we simplify the notation and use θ in the rest of the paper) to control the banked turns. We hypothesize that information about the fly's initial ego-rotation, which is present in the visual system upon the detection of threat and initiation of the escape, is important for how the maneuver unfolds and is controlled.

Because of the short timescale of the escape turns, the long-standing hypothesis has been that only the mechanosensory circuit, i.e. the halteres, are fast enough to perform active control [5]. However, the halteres themselves are insufficient for controlling evasive maneuvers [20, 29, 30] unless they integrate ego-motion inputs encoded by the visual system [13]. Halteres are known to be gyroscopic sensors [31]; they help a fly keep its aerodynamic balance [32]. Like the vestibular system, they respond to short timescale rotational perturbations (e.g., air anisotropies) and experience the Coriolis forces during a fly's body rotations. Previous work showed that the halteres can sense ego-rotation with angular velocities below $2500°/s$ [29, 30]. This is dwarfed by the angular velocities achieved during evasive maneuvers, which can be as high as $5300°/s$. Indeed, most previous work about how the halteres control behavior focused on voluntary saccadic turns, i.e. a fixed stereotyped rotation with angular velocities around $1000°/s$. To cope with high angular velocity during evasive maneuvers, the halteres can only engage in active control by adjusting their dynamic range using descending visual inputs about the fly's ego-motion. This is the so-called control loop hypothesis [33]. This vision-mediated control has been observed in other behaviors [34–36]. Recent work further clarified that the ego-rotation encoded by the visual input can broaden the dynamic range of the haltere steering neurons [13]. In this work, we show how visually-encoded information about the fly's ego-motion can guide evasive maneuvers through the VS system's connection to the halteres via the neck motor center.

Because the visual systems in both the blowfly and *Drosophila* have a 30 ms processing lag [26, 37], it is unlikely that evasive maneuvers can access visual information through feedback within the 40 ms time span. However, if a fly can use past visual input before evasive maneuvers to predict its own future visual experience during evasive maneuvers, it can bypass this processing lag and still use visual inputs for active control. This is the so-called bottom-up prediction which exploits the temporal correlations between past and future visual stimuli during evasive maneuvers. Such bottom-up prediction exists in the vertebrate retina [38]. It ensures fluid interaction with the external environment. It was also shown to be important in the formation of long-term memory [39]. In our case, the fly's evasive maneuver is initiated with a particular heading change that is selected in the moment and based on the purely visual threat cue. The escape trajectory depends on the threat angle relative to the fly's heading [20]. Where and how the escape maneuver begins constrains how it will unfold, giving the visual system ample predictive power with which to guide active flight control. Hence, we hypothesize that such bottom-up prediction helps the fly to use its ego-motion to control brief evasive maneuvers. Namely, we show how this bottom-up prediction provides information about future sensory input, subverting delays in the visual input stream. In addition, because blowflies are known to execute higher velocity saccadic turns than *Drosophila*, this bottom-up prediction will be even more important for blowflies' shorter/faster evasive maneuvers.

The fly visual system is organized in four largely feedforward layers: retina, lamina, lobula and lobula plate. Photoreceptor signals are processed in these layers, which are all arranged in a columnar, retinotopic fashion. In a striking parallel to the vertebrate retina [40, 41], the direction of local visual motion is computed in parallel ON and OFF

motion pathways [42]. These local motion detectors show temporal frequency optimum [43, 44], i.e. they are not speedometers. Their outputs go to columnar T4 and T5 cells, which are essential for the fly to initiate escape [3] . Specific subpopulations of T4 and T5 cells are directionally tuned to one of the four cardinal directions: front-to-back, back-to-front, upwards, and downwards. Depending on their preferred direction, they terminate in specific sublayers of the lobula plate [45]. Lobula plate tangential cells (LPTC) then integrate direction selectivity outputs from multiple sublayers to generate behaviorally relevant encoding. For example, previous work described ultra-selective LPTCs for detecting the size and speed of a looming target (LPLC2 and LC2). After the detection of the looming threat, it remains an open question which LPTCs may support the active control necessary for the aerodynamically unstable in-flight evasive maneuvers.

Because of this high angular velocity, banked turns during evasive maneuvers are mostly composed of pitch/roll combinations. These rotations generate rotational flow-fields processed mainly by the the vertical system (VS) network [46–48]. There are 20 different VS cells (10 at each compound eye). VS dendrites integrate upward and downward local motion signals from T4 and T5 such that their axons output global motion components of ego-rotation based on the fly's visual experience. Each VS cell has its receptive field centered at a specific rotational axis of the fly's coronal plane (combinations of all pitch and roll rotations). They are numbered VS1-VS10 according to their preferred ego-rotation arranged along the fly's anterior-posterior axis [49]. Recordings from VS cells confirm that they mainly respond to rotational flow fields but not expanding flow fields [50]. If the animal is moving in an open field, translational motion does not generate rotational flow, and the VS network is only sensitive to ego-rotation. VS cells are upstream of both other LPTCs as well as descending neurons connecting to the neck motor center. Because the rotations during evasive maneuvers are mainly pitch and roll combinations (similar to how a fighter jet gains tactical advantage), the VS network is a good candidate for encoding information specific to and most relevant for this behavior.

The VS network has a chain-like wiring architecture [21, 51–53]. Each VS cell only connects with other VS cells with immediately neighboring receptive fields, e.g., VS2 only connects to VS1 and VS3. Meanwhile, the VS1 and VS10 cells show reciprocal inhibition [54]. Previous dual-recording experiments [52] have shown that VS cells connect amongst each other through electrical synapses. Further dye-coupling experiments also have shown that these electrical synapses were gap junctions [53]. In [21], they identified that these gap junctions are located at the axons of VS cells. By having these axonal gap junctions, the VS network implements an electrotonic segregation mechanism between its dendrites and axons: all VS cells show broadened receptive fields at their axons compared to those at their dendrites. These broadened receptive fields improve the encoding robustness of visual motion [21, 55] at the output of the VS network. The output from the VS network arises from subpopulations of adjacent cell triplets, which target different downstream areas [23, 56]. In particular, the VS network connects to the downstream descending motor neurons or neck motor neurons only through the VS 5-6-7 triplet of cells [51, 56], which have dendritic receptive fields located at the center of the fly's field of view.

Recent work also showed that this wiring architecture, coupled with subpopulation readout from the VS 5-6-7 triplet, enables near-optimal encoding of constant rotational motion [57]. Namely, the axonal voltages of both the VS triplets and the entire VS network encode almost all of available information from the input current about a constant instantaneous rotation. However, an encoding scheme optimized for retaining all of the information about the instantaneous input is not necessarily optimized for natural behaviors. These behaviors, like evasive maneuvers, have highly dynamic

ego-motion. This is survival-critical: the animal will maximize its survival chances if it can make predictions about its own future ego-motion and control its wing steering accordingly. Encoding a past ego-motion is only useful if it is informative for the animal's future experience. In order to identify whether this specific wiring architecture of the VS network supports evasive maneuvers, we ask whether it encodes predictive information about future ego-motion on a fast enough timescale.

To explore this hypothesis, ideally one would trace the activity of the VS cells in behaving animals. However, evasive flight maneuvers require untethered flight, which makes population recording from the VS network prohibitive. Furthermore, it is not feasible to block the VS network in flying animals, because they are essential for optomotor responses [50, 58]. Therefore, we use numerical simulations of an in-silico, biophysically realistic compartmental reconstruction of the VS network to investigate how the VS network might encode this kind of fast, predictive information. This compartmental reconstruction of the VS network is highly experimentally constrained [21, 22]. All single-cell [51] and neuronal circuitry parameters [21, 52, 53] are set such that this compartmental reconstruction behaves as does the real VS network when presented with the same current injection [21, 22, 52, 53]. Our computational approach (using a model when large scale recordings from a complete circuit during natural behavior is not possible) is similar to how previous work on electric fish [59] used a synthetic population model of the electrosensory lobe to show how accurate cancellation of self-generated electric signals is achieved.

We first show that the VS network uses axonal gap junctions to output substantial predictive information for evasive maneuvers. Next, we show that this predictive information is near-optimal throughout the duration of evasive maneuvers. We further show that the output circuitry of the VS network (the VS 5-6-7 triplet) to the neck motor center retains all available information about future stimuli, i.e. compressing the readout does not sacrifice how much a downstream pathway knows about the ongoing evasive maneuver. Finally, we identify that the encoding of predictive information is particularly suitable to fine-tuning future motion directions. These results suggest possible novel sensory-motor pathways: either a direct connection between the lobula plate descending neurons to the wing steering muscles [60–62], or an indirect connection from the visual system through the halteres to the wing steering muscle, as proposed in the control loop hypothesis [13, 33].

Results

Visual prediction provides substantial information about motion without delay

We show in Fig.1 that visual prediction contains substantial information about future motion for controlling evasive flight maneuvers. We first use a schematic trace to illustrate the inputs and delays in the fly visual system (Fig.1A). Previous works showed that the haltere outputs reach the wing steering muscles after a 15-20 ms delay [20, 27], towards the second half of the maneuver, right before the active counter-banked turn starts. Visual feedback would arrive too late, coming online only after 30 ms, long after the initial rotation is replaced by the counter rotation through active control [27].

To quantify how much visual prediction encodes about ego-rotation (θ) in the fly's future escape trajectories (Fig.1B), we define this ego-rotation-relevant predictive information in the output voltage from the fly VS network as $I_{future:\theta}$ (Eq.(1)),

$$\begin{aligned} I_{future:\theta}(\Delta t) &= I(V_{past}; \theta_{future}) \\ &= I(V_t; \theta_{t+\Delta t}), \end{aligned} \tag{1}$$

where V_t is the output axonal voltage of the VS network at time t. Δt is the time interval between the past voltage and future ego-rotation. Here, we use intervals of $\Delta t = 10ms, 20ms, 30ms, 40ms$ to obtain the output of the VS network. This is because the maximum firing rate of the descending neuron connecting to the neck motor center is 100 Hz [22], which corresponds to an integration step of at least 10 ms (see Materials and Methods). Throughout this paper, we represent future ego-rotations $\theta_{t+\Delta t}$ by their vector components ($\cos(\theta_{t+\Delta t})$, $\sin(\theta_{t+\Delta t})$). The cosine component corresponds to roll direction/magnitude and the sine component corresponds to pitch direction/magnitude. This vector is within the fly's coronal plane, to which the VS neurons are selectively sensitive. We then estimate $p(\theta_{t+\Delta t})$, i.e. the stimuli distribution and $p(\theta_{t+\Delta t}|V_t)$, i.e. the probability of future ego-rotation conditioned on the past output axonal voltage to obtain $I_{future:\theta}$ (see Materials and Methods). Fig.1B shows that the predictive information $I_{future:\theta}$ in the VS output voltages captures nearly 50% of the entropy of the future escape trajectory. This is because where and how the escape maneuver begins constrains how it will unfold, giving the visual system ample predictive power. This suggests that the predictive information encoded by the VS network is an important information source for evasive flight behaviors in the natural environment.

To evaluate $I_{future:\theta}$, we need to approximate both ego-rotation distributions and the respective output distributions of the VS network. To obtain these ego-rotation distributions, we generate 650,000 samples of visual field motion experience based on behavioral recordings published in [20]. Each visual experience corresponds to one instance of a particular evasive maneuver embedded in a randomly selected set of nature scene images. There are 10,000 samples for each of the 65 evasive flight trajectories with duration of 40 ms (out of the total 92 published trajectories in [20]). Fig.1C shows one exemplar visual experience of a particular evasive maneuver trajectory. Here, we obtain the "cage" of natural images for simulation by randomly selecting six images out of the van Hateren dataset [63] and patch them onto the six faces of a cube. Then we generate a movie mimicking an evasive flight in the natural environment by rotating this natural scene cage according to measured rotations in an evasive flight trajectory Fig.S1 Fig (Because previous work [23] showed that the VS network is not responsive to translation motion, we do not use the translation component of evasive maneuvers in this simulation, also see Materials and Methods). We next project this movie onto a unit sphere that represents the fly's retina, following the protocol introduced in [23, 64]. There are 5,500 local motion detectors located on this unit sphere, whose outputs are local motion estimates based on pixel intensity difference between neighboring photoreceptors.

These outputs are then integrated and become the input current $\mathcal{I}$ to the VS network (Fig.1D). Because brief and fast evasive maneuvers have a quickly varying ego-rotation, we identify prediction as the most important 'relevant' variable in the input to the VS system and set up the corresponding information bottleneck problem (Fig.1E): this method finds the maximal amount of relevant predictive information that the VS network can encode about the future ego-rotation θ_{future} and the input $\mathcal{I}_{future}$ via its axonal voltages at a past time, V_{past}, subject to a constraint on the amount of information V_{past} has about the input in the past. The absolute maximum of the predictability is set by the correlation between past and future of the ego-rotation $\theta_{past}, \theta_{future}$ trace itself, which we take from real fly maneuvers. Note that this is different from asking whether V retains all of the information about an instantaneous ego-rotation. In that case, we search for the optimal V between the instantaneous $\mathcal{I}$ and θ at the same time (without prediction forward in time).

The specific texture of the visual scene significantly impacts the ego-motion-induced input to the VS dendrites. Of course, if the fly rotates against a uniform background, no visual motion information is available. The details of the spatial correlation structure

in the scene also affect the input to the VS system and its subsequent ego-rotation estimate. Previous work [21] compared motion encoding using two different visual backgrounds: an artificial background with regular texture (e.g., random dots) and a natural scene background with irregular texture and an inhomogeneous contrast distribution. Those results showed that with the artificial background, rotational optic flow can be reliably read out at the VS dendrites. Axonal gap junctions do not improve the estimate. However, the irregularly distributed patches of high and low contrast in natural scenes make the motion estimation noisy, even after the dendritic integration. Only once the outputs are filtered through the VS axons with gap junction coupling does robust encoding of the fly's rotational optic flow emerge. By pooling inputs across neighboring VS cells, axonal gap junctions average out the fluctuations in local motion measurements due to irregular textures in natural scenes. Moreover, recent work [57] showed that even with those axonal gap junctions, the VS network can only encode 50% of the motion information in natural scenes, as compared to artificial, regular textures (shown in S1 Fig$C$). Here we are challenging our *in silico* fly with both naturalistic self-motion and complex visual texture. The axonal gap junctions in the VS system may have evolved to solve this particular, behaviorally relevant problem in natural scenes.

Axonal gap junctions enable prediction during evasive maneuvers

Fig.2 shows that the anatomical locations of the gap junctions have significant impact on the VS network's capacity of encoding predictive information. Located in the lobula plate, the VS network does not have direct access to the visual input. Instead, the dendrites of all VS cells receive current inputs resulting from integrating the outputs from hundreds of upstream local motion detectors [52]. Here, we use correlations in the VS input current induced by future ego-rotation as a proxy for correlations among past and future ego-rotations, which themselves can be in turn used to encode predictive information about future ego-rotation. In this encoding scheme, the correlation between the past and the future of the input current itself limits how much predictive information the VS network can encode. This generalized correlation between the past and future of the VS inputs is $I_{future:limit}$ (Eq.(2)).

$$\begin{aligned} I_{future:limit} &= I(\mathcal{I}_{past};\mathcal{I}_{future}) \\ &= I(\mathcal{I}_t;\mathcal{I}_{t+\Delta t}) \end{aligned} \tag{2}$$

This is also the mutual information between the past and future input (the dendritic current) and defines the total interdependence of the current with itself in time.

Similar to $I_{future:limit}$, we also define the information retained by the axonal voltage of the VS network about its future input as $I_{future:in}$ (Eq.(3)).

$$\begin{aligned} I_{future:in} &= I(V_{past};\mathcal{I}_{future}) \\ &= I(V_t;\mathcal{I}_{t+\Delta t}). \end{aligned} \tag{3}$$

This is the predictive information between the output axonal voltage and the future input current, which again we are using as a proxy for future ego-rotation. Causality dictates that the past axonal voltage can only obtain information about the future current from its own past current, therefore $I_{future:limit}$ (shown as the solid line in Fig.2) is an upper bound on $I_{future:in}$. Here, we explore what network wiring features support the maximal transmission of the correlation structure in the input current onto the output axonal voltage of the VS network.

As shown in Fig.2, axonal gap junctions are necessary for the system to encode the maximal amount of predictive information about the input current. Namely, the

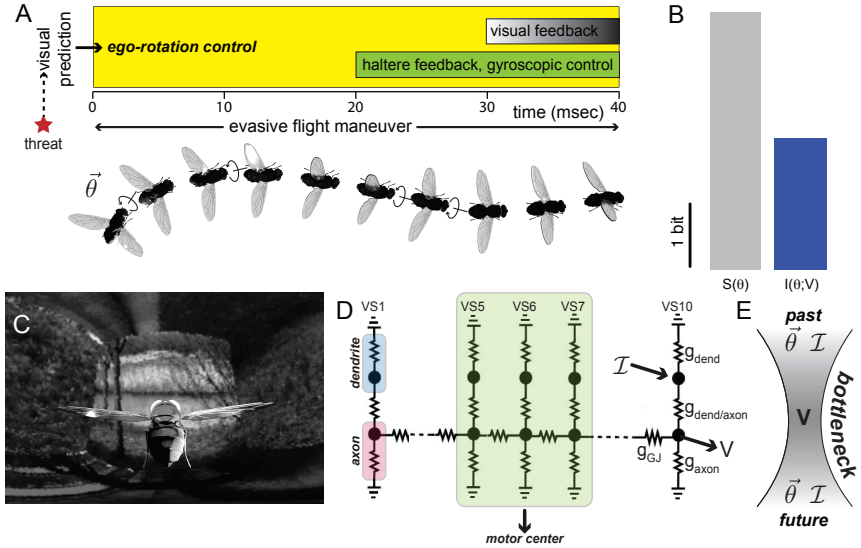


Fig 1. Predictive information is the dominant information source about visual inputs during evasive flight maneuvers. (A) Upon emergence of a threat (shown as the red star), dashed arrow represents the visual-motor delay of 60 ms from the onset of threat to the start of evasive maneuvers. After this sensory-motor delay, the position of the threat is known. The fly performs an evasive maneuver by changing its heading through a banked turn (arrows show a rotation at direction $\vec{\theta}$ and its respective counterrotation). During evasive maneuvers, visual predictions can provide motion information throughout the entire duration, i.e. without delay (shown as the yellow zone), whereas the haltere feedback is only available after 20 ms (shown as the green zone) and the visual feedback is only available after 30 ms (shown as the shaded zone). The arrow leading to the haltere system illustrates how visual information might regulate haltere activity (as recently shown in [13]): because of the 30 ms sensory processing lag, haltere activity must be regulated by visual prediction. (3D fly rendering courtesy of D. Allan Drummond.) (B) This histogram compares how much information the visual prediction (shown in blue) can encode about ego-rotation ($I(\theta, V)$) during evasive maneuvers with their respective entropy (shown as $S(\theta)$ in gray). We use the ego-rotation distribution at $\Delta t = 10ms$ into the evasive maneuver to compute this entropy. Its distribution is shown in S2 Fig$A$. Note that the VS output encodes almost half of the entropy of a future ego-rotation. (C) The Mercator map of a randomly generated natural scene background. To generate this map, we first randomly generate a natural scene environment. We then generate a movie mimicking an evasive flight in the natural environment by rotating this natural scene environment according to the respective measured rotations. We project this movie to a unit sphere that represents the fly's retina, see details in S1 Fig. There are 5,500 local motion detectors (LMD) on this unit sphere as on the fly's retina. The responses of these LMDs are then integrated as the input current $\mathcal{I}$ to the VS network (shown as an arrow to D). (D) A biophysically detailed model of the VS network, based on known neural circuitry [51, 52]. Note that because the soma is decoupled in VS cells (only connecting to the rest of the cell via a soma fiber), we leave out the soma in this VS model. We highlight the outputs to the neck motor center here, the axonal voltages of the VS 5-6-7 triplet. This is the only known readout that directly connects to motor pathways. (E) A cartoon showing how the information bottleneck problem is setup for prediction in this system: using the general correlation between a past input (either the egorotation $\vec{\theta}$ or the corresponding dendritic input $\mathcal{I}$, this information bottleneck finds a compact representation V (i.e. the bottleneck defines how much information about the input is 'squeezed out' when V is generated) of the past (θ_{past} or $\mathcal{I}_{past}$) that retains predictive components about the future (θ_{future} or $\mathcal{I}_{future}$).

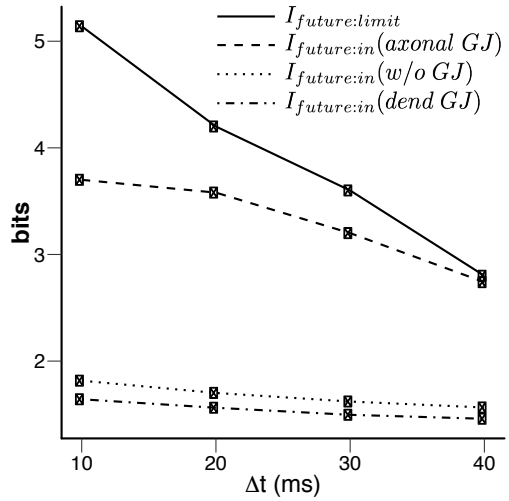


Fig 2. The capacity of the VS network to encode predictive information varies with the anatomical locations of the gap junction between VS cells. The predictive information about the future input current, $I_{future:in}$ encoded in four different schemes: 1) the past dendritic input current (solid line, this is the limit $I_{future:limit}$. It is also the upper bound of $I_{future:in}$), 2) the past axonal voltage when the gap junctions are present between VS axons (dashed line), 3) when the gap junctions are present between VS dendrites (dotted line) and 4) in the absent of gap junctions (dash-dotted line). All gap junctions = 1000 nS for both settings when they are present. Only their locations differ, i.e. axon vs. dendrite. Note that when the gap junctions are present between VS cell axons, the output voltages preserve almost the entire amount of the predictive information available at the inputs (red). (See also Materials and Methods.)

$I_{future:in}$ (shown as dashed line) only approaches $I_{future:limit}$ (shown as solid line) when gap junctions are present between neighboring VS axons. The other two configurations of gap junctions, i.e. no gap junctions or gap junctions at the dendrites (shown as dotted and dashdotted lines respectively), cannot encode as much predictive information. Furthermore, the presence of axonal gap junctions in the VS network improves the efficiency of the encoding of predictive information much more than gap junctions improve the encoding of instantaneous ego-motion (S3 Fig). We observe the same rough level of improvement with gap junction coupling for both the encoding of the future input as well as the future ego-rotation. Next, we investigate whether such improvement approaches the respective physical limits for input current or ego-rotation set by the solution to the information bottleneck equation.

The VS network is near-optimal in predicting its own future input

All of the information encoded by the VS network comes from its sole input current, $\mathcal{I}_{past}$. To quantify the efficiency of encoding, we need to qualify not only the prediction (i.e. the $I_{future:in}$), but also how much the axonal voltages encode from past input. We define this as another mutual information quantity: $I_{past;in}$ (Eq.(4)),

$$\begin{aligned} I_{past:in} &= I(V_{past}; \mathcal{I}_{past}) \\ &= I(V_t; \mathcal{I}_t). \end{aligned} \tag{4}$$

Comparing $I_{past:in}$ and $I_{future:in}$, where the past is at time t and the future at $t + \Delta t$, we can ask formally whether the VS network encodes as much as predictive information as possible, using the information bottleneck framework [65]. Given the amount of information the axonal voltage encodes about the past sensory input, what is the maximal amount of information it can retain about the future input? Such an optimum $I^*_{future:in}(I_{past:in})$ traces out a bound (the dark blue line) in Fig.3 as a function of $I_{past:in}$. It is the maximal possible predictive information at each level of compression, $I_{past:in}$. For encodings with the same $I_{past:in}$, those approaching the bound are optimal.

The known circuitry of the VS network allows us to probe two coupled questions: 1) What is the predictive efficiency (based on the encoding of the past) and 2) what is the predictive capacity (encoding of the past input only to predict the future input) of the VS network, given different readout circuitry (the entire VS network vs. the output VS 5-6-7 triplet)?

The predictive capacity of the VS network for its own future inputs is near-optimal. As shown in Fig.3A, the axonal voltages of the VS network encode $I_{future:in} = 3.49 \pm 0.1$ bits for future inputs at $\Delta t = 10$ ms (the beginning of the banked turn). Considering that optimum is $I^*_{future:in}(I_{past:in}) = 3.59$ bits, using axonal voltages of the entire VS network capture $I_{future:in}/I^*_{future:in} = 97.2\%$ of the optimal predictive information.

Similarly, using only the axonal voltages from the triplets, prediction of the entire VS network's future input is close to optimal as well (the cross in red in Fig.3A). All encodings based on outputs of triplets reach $I_{future:in} = 2.89 \pm 0.36$ bits while their respective physical limits are 3.07 ± 0.24 bits in Fig.3A. This suggests that all triplets achieve $89.8 \pm 1.5\%$ efficiency in encoding predictive information about the inputs $I_{future:in}/I^*_{future:in}$. VS triplets are not as efficient as the entire VS network in encoding the future dendritic input.

In general, triplets also retain less absolute predictive information about future input than the entire VS network throughout the time span of evasive maneuvers (Fig.3B). But the key point is not how well the triplet predicts the VS input, but how well it might help guide behavior by predicting the fly's future ego-rotation. We next explore this directly by asking how much information the triplet output has about the future ego-rotation.

The triplet architecture selectively encodes predictive information about future ego-rotation

Here, we show that the triplet readout architecture retains nearly all of the available predictive information about the future ego-rotation $I_{future:\theta}$ (light bars in Fig.4A) available to the VS network from the upstream input (darker bars in Fig.4A). Because downstream pathways of the VS network only read out from triplets, the VS network appears to use a two-step strategy to optimize this readout: it first efficiently represents correlations within its own past and future input, i.e. $I_{future:in}$, at its output; then it selects components within that output that are relevant for predicting future ego-rotation, i.e. $I_{future:\theta}$. This is possible because correlations coming from events in the visual world, such as the movement of large objects or the full-field background movement, have a different temporal structure (e.g. longer correlation times) than those internal to the brain.

We also observe that all triplets are near-optimal in encoding the predictive information about future ego-rotation (Fig.4B) (such optimality is also present for prediction of the distant future, i.e. $\Delta t > 10$ ms, results not shown). Considering that the VS 5-6-7 triplet encodes nearly as much information about the future ego-rotation as the VS network (Fig.4A), the main benefit of using triplets is compression: the VS triplet readout discards information predictive of the less behaviorally-relevant intrinsic

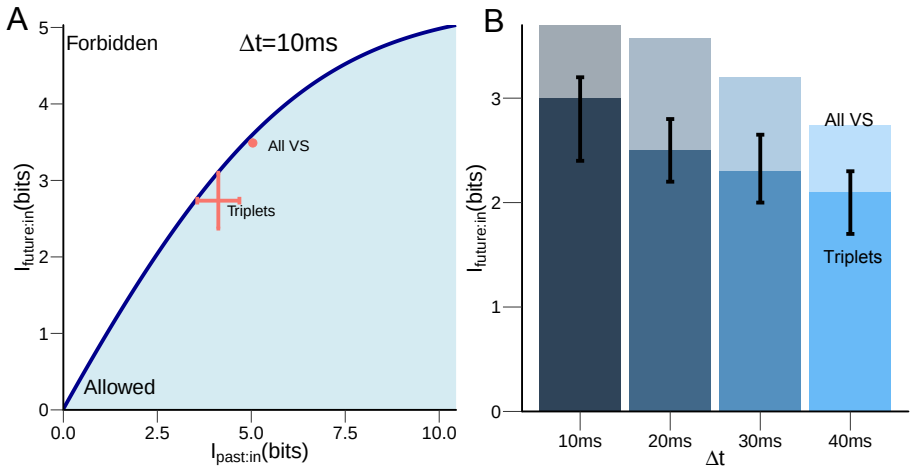


Fig 3. Near-optimal prediction of the input to the VS network. (A) The encoding of predictive information about the future current input to the VS network is near-optimal 10 ms after evasive maneuvers starts ($\Delta t = 10$ ms). Such performance is present for using both the entire VS network and the triplets. The dark blue curve traces out optimum encoding of future input to the VS network given varying amounts of information retained about the past input (also see Materials and Methods). This curve also divides the plane into allowed (blue shaded region) and forbidden regions. No encoding can exist in the forbidden region because it cannot have more information about its future inputs than the input correlation structure allows, given causality and the data processing inequality. In addition, the maximal amount of information (shown as the highest point of the information curve) that is available as predictive information is limited by the correlation structure of the input (current), itself. We then plot the amount of information the axonal voltages of VS network (we show with axonal gap junctions in pink and without gap junctions in black) encode about the future input (the input current at time $t + \Delta t$) versus the information they retain about the past input (the input current at time t) (with all 120 triplets (crosses) and the whole network (circle)). The information efficiency, compared to the bound, contained in a particular encoding scheme corresponds to a single point in this $I_{past:in}$- $I_{future:in}$ plane, which shows how much information it encodes about the past input vs. how much it encodes about the future. A particular VS encoding could occupy any point within the blue shaded region, but those that get close to the bound $I^*_{future:in}(I_{past:in})$ for a particular $I_{past:in}$ are the maximally informative predictors of the future input. (B)$I_{future:in}$ for all VS (light bars) vs. triplets (dark bars, with error bars) throughout the time span of evasive maneuvers ($\Delta t = 10$ ms, 20 ms, 30 ms, 40 ms)

dynamics of the inputs themselves. This compression is close to the niche region where the predictive information just begins to saturate, indicating that the triplet output has a better trade-off than the whole VS network in terms of how much $I_{future:\theta}$ it can encode given its cost $I_{past:in}$.

Although all triplets encode similar amounts of information about the future ego-rotation (the standard deviation of the $I_{future:\theta}$ amongst all 120 triplets is just 0.1 bits), the particular triplet connecting to the neck motor center, the VS 5-6-7, is one of the better choices in terms of how much information about the future ego-rotation it packs into its prediction of the future input, while the VS 1-2-3 triplet is the most efficient. However, if we factor in wiring constraints, linking the output from VS 5-6-7 to a downstream dendritic arbor in the descending neurons for the neck motor center requires a much shorter wiring length compared to the peripheral location of the VS 1-2-3 triplet (VS cells are numbered according to their locations along the anterior-posterior axis; VS 5-6-7 are central in the compound eyes). It is possible that the minimization of wiring length [66] is important in selecting the simultaneously most predictive and most resource-efficient encoding.

Here we show that the VS 5-6-7 triplet that projects to the downstream neck motor center retains nearly all of predictive information about future ego-rotation as is present in the entire VS network. This result also shows that the predictive information encoded by the VS network is compressible: the VS 5-6-7 triplet successfully reformats the predictive information from 20 dendrites/axons (10 VS cells from both compound eyes combined) into six axons (the axons of VS 5-6-7 from both compound eyes combined). In the next section, we investigate how ego-rotation representations vary based on either the entire VS network or the VS 5-6-7 triplet. We hope to understand a) what kind of computation is possible via the encoding of near optimal predictive information, and b) how the VS 5-6-7 triplet reformats this near-optimal prediction.

Predictive information encoded by the VS network provides fine-scale discrimination of nearby stimuli

Because the entire VS network has 20 VS cells and the VS 5-6-7 triplet only has 6 cells, it is difficult to directly compare their encoding of the fly's ego-rotation with different dimensions. However, both encoding schemes encode similar amounts of predictive information for ego-rotation. Because solutions with the same y-value or similar y-values on the information curve share the same dimensionality in their respective compressed representations [65], the difference between these representations shows how the whole VS network differs from the VS 5-6-7 triplet in predictive information encoding. Because solving for these representation is generally intractable, we instead use the variational information bottleneck (VIB) to approximate these optimal representations [67–69]. The VIB is a generative learning framework. It is closely related to variational autoencoders [70]. Given pairs of inputs and outputs, it generates a latent feature space whose dimensions are predictive features from the input to the output (S5 Fig). One can then project the input into this latent feature space to obtain the predictive representation of the output. Here, we obtain these predictive representations in two steps: we first train the VIB to generate a latent feature space that maps the input (the axonal voltages of the VS network) to the future input current. We next project the axonal voltages (that correspond to the same future ego-rotation) onto this latent space. We can label these points in the latent space according to their respective future ego-rotation, and repeat this procedure for multiple ego-rotations. We can visually and computationally examine how overlapping or distinct these maximally predictive clusters of ego-rotations are in the latent space of the VIB. To allow for a direct comparison, we keep the dimension of the latent feature space to be the same while

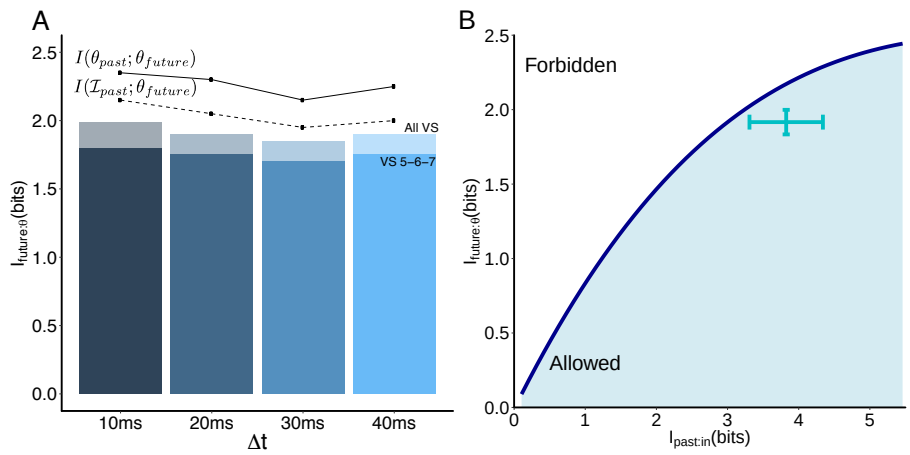


Fig 4. Encodings based on the axonal voltages of triplets are near-optimal in predicting the future ego-rotation. (A) Histogram showing that the triplets (we use the output triplet VS 5-6-7 triplet here) encode nearly as much information *about the future ego-rotation* (shown in dark bars) vs. the entire VS network (shown in light bars), throughout evasive maneuvers. The solid line shows the mutual information within ego-rotation themselves: between the prior heading upon the detection of visual threat θ_{past} and different future ego-rotation throughout evasive maneuvers. The dashed line shows the mutual information between past input to the VS network and different future ego-rotation, all using the same past input $\mathcal{I}_{past}$ corresponding to the previous θ_{past}. This is also the limit of prediction in the information bottleneck framework. (B) The encoding of predictive information for the θ_{future} at 10 ms after the start of evasive maneuvers ($\Delta t = 10$ ms). The dark blue curve traces out the optimum encoding for the future ego-rotation ($I(V_{past;}\theta_{future})$) given varying amounts of information retained about the past input ($I(V_{past;}\mathcal{I}_{past})$). The cyan cross corresponds to how much information each of all possible 120 triplets encodes about the future ego-rotation vs. how much information they retain from the past input.

changing the input, using either the axonal voltages of the entire VS network, or those of the VS 5-6-7 triplet. S6 Fig shows that using two dimensional latent space already enables these predictive representations to encode substantial predictive information.

The VIB-generated representations of future ego-rotation (Fig.5) show that the predictive information encoded by the VS network supports fine-scale discrimination of the input motion direction. For ego-rotations with different degrees of clockwise roll and up-tilt pitch (but the same direction), the encoded predictive information puts those close in their azimuths far apart, i.e. a pair of ego-rotations (56° and 67°, shown in S8 Fig$A$ and S8 Fig$B$) with 10 degrees difference are mapped to distinct, well-separated clusters in the latent space of the VIB, whereas another pair (37° and 56°) that are farther apart share some overlap (S8 Fig$C$). The VS 5-6-7 triplet preserves this fine scale discrimination (and S8 Fig$B$ S8 Fig$D$) while compressing the readout. The same fine-scale discrimination is also present for ego-rotations combining counter-clockwise roll and up tilt, i.e. corresponding to vectors within the 4th quadrant of the fly's coronal plane (Fig.5B and Fig.5D). However, these predictive representations cannot discriminate ego-rotations with vastly different roll or pitch directions, i.e. belonging to different quadrants: there is substantial overlap if we overlay these predictive representations, e.g. the cluster corresponding to 270° (shown in magenta in Fig.5C) will entirely cover the cluster corresponding to 19° (also shown in magenta, but in Fig.5A). The same overlap is also present in Fig.5B and Fig.5D as well.

Even with a representation that retains all of ego-rotation-relevant information in the input to the VS network, one cannot use information available at the VS network input to discriminate ego-rotations of wide difference. We construct such a representation based on the instantaneous input current of the present ego-rotation. These input currents contain 2.44 bits relevant to the fly's instantaneous ego-rotation i.e. without prediction forward in time. This information is higher than that available via predictive information from the past input current (2.1 bits, shown as the red bar in S6 Fig). The first two principal components (PC) of the input current retain nearly all available ego-rotation relevant information, so we ask how ego-rotations disentangle in this representation of the first two instantaneous PCs. By projecting all VS input currents into these first 2 PCs, we find that there still exists substantial overlap between ego-rotation (S7 Fig), e.g. the cluster of 19° in magenta almost covers the entire cluster of 247° in light green (S7 Fig$A$). This shows that the input to the VS network can only support fine-scale discrimination, whether an instantaneous readout or predictive readout.

Ego-rotation representations are different if the system is optimized for prediction as opposed to optimized for encoding the present input. For example, the VIB-based predictive representation not only separates nearby ego-rotations (e.g., 56° and 67°) into distinct clusters, but also inserts a cluster of another ego-rotation angle (e.g, 19°) between them. By contrast, the disentanglement implemented by a circuit that maximally encodes the instantaneous inputs projects all distinguishable ego-rotations into adjacent clusters according to their azimuthal angle, i.e. 56° and 67° are next to each other. Because predictive representations contain less absolute information about the ego-rotation than the instantaneous-optimal representation, this difference suggests that the predictive information might preferentially support fine-scale discrimination of nearby ego-rotations. Information is not necessarily lost, however. Downstream, at the descending motor neurons [71–74] or the neck motor center [75, 76], information from other pathways (e.g. the haltere and prosternal organs [24]) is integrated. Thus, the discrimination of larger ego-rotation angles may be supported, while allowing the VS system to serve a specialized role in fine discrimination.

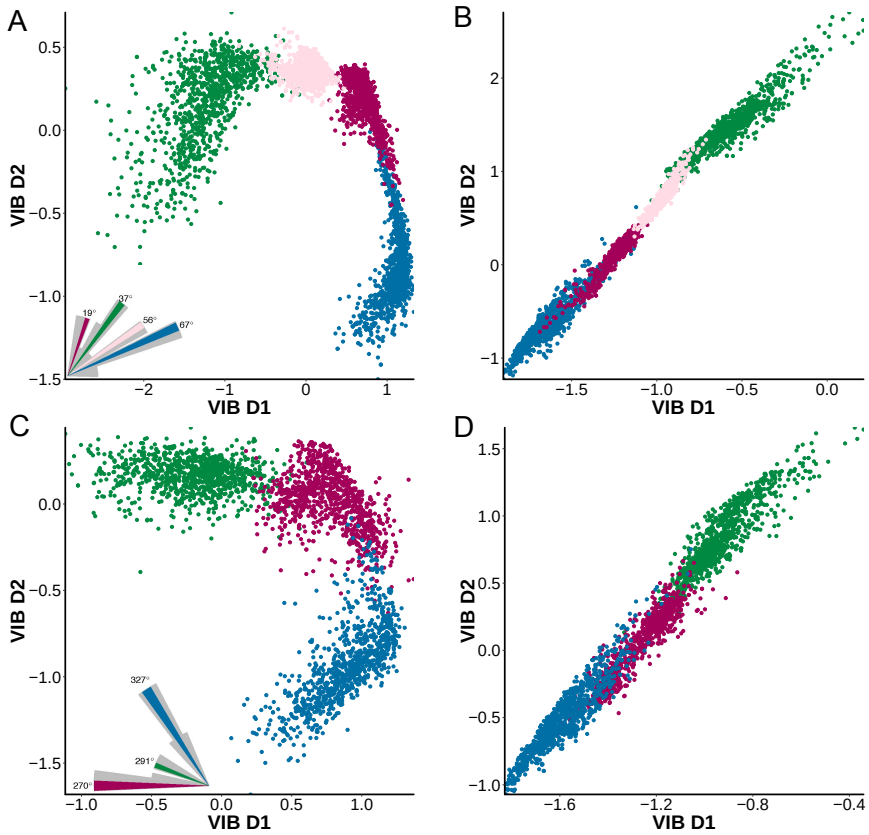


Fig 5. The predictive information encoded by the VS network supports fine scale discrimination of future ego-rotation. (A) The predictive representation of four future ego-rotations in the same quadrant of roll and pitch, e.g. an up-tilt and a clockwise roll. This representation maps the axonal voltage of the entire VS network to future ego-rotation through a latent feature space. The dimensions in this latent feature space (shown as VIB D1 and VIB D2) are VIB-learned predictive features based on the output of the VS network. All ego-rotation correspond to vectors within the 1st quadrant of the fly's coronal plane. The inset shows a polar histogram in grey and the four selected ego-rotations in color. (B) Similar to A but using the axonal voltages of the VS 5-6-7 triplet. (C) Similar to A, but ego-rotation are all counter-clockwise roll and up-tilt, corresponding to vectors in the 4th quadrant (between 270° and 360°) of the fly's coronal plane. (D) Similar to C, but obtained using the axonal voltages of the VS 5-6-7 triplet as the VIB input.

Discussion

Here, by focusing our analysis of the fly's neural code for a key survival strategy, the evasive flight maneuver, we have shown that the VS network in the early fly visual system encodes behaviorally relevant predictive information near-optimally. A subpopulation readout mechanism, based on triplets of VS cells, further compresses the representation of that predictive information. This compression trades off the prediction of its own input with the prediction of future ego-rotation: while it encodes less absolute information about the future input, it retains the predictive information about what a fly will experience during evasive maneuvers at higher fidelity, in the outputs. The encoding of predictive information has a concrete behavioral goal: it enables fine-tuning of ego-rotation during evasive maneuvers.

Combining these observations, the fly brain satisfies an overarching computational goal of effectively guiding evasive flight trajectories through visual prediction at both levels of input filtering (via axonal gap junctions) and output reformatting (via subpopulation readout based on triplets). By next identifying that the predictive representations of future ego-rotation are best at enabling fine-scale discrimination of nearby ego-rotations, we have shown how structure maps to function in this motion sensing system. In addition, we have shown that behaviorally relevant features of ego-rotation are faithfully encoded via circuit mechanisms at the sensory end of the arc from sensation to action. This suggests that behavioral goals sculpt neural encoding even at the earliest stages of sensory processing.

These results predict a novel sensory-motor pathway between the visual system and a visually gated motor neuron of wing steering muscles. Evasive maneuvers consist of rapid initial rotations and counter rotations. These rapid turns require elevated wing kinematics, including stroke amplitude and frequency [61]. Previous work showed that a motor neuron of wing steering muscles, e.g., the second basalare motor neurons (M.b2) in both the blowfly and *Drosophila*, are responsible for initiating these elevated wing kinematics [60, 77]. However, little is known about these neurons other than that they can be switched on and off by visual input [62, 78]. Our work shows that visual prediction can reach these motor neurons at the beginning of evasive maneuvers. For rapid evasive maneuvers, this pathway is likely implemented by the halteres, following the control loop hypothesis [33]. When the predictive visual information tunes the strength of haltere mechanosensory feedback [13], it indirectly activates these motor neurons via the haltere's connections to the second basalare motor neurons (M.b2) through electric synapses.

Our work also suggests that the encoding of predictive information is a key functional role of the lobula plate tangential neurons in dipterans. This might not be obvious given that different species have very different behavioral repertoires and selective pressures that sculpt their tangential neuron computations [24]. However, all dipterans use their lobula plate tangential neurons to encode motion information [24, 73]. All of these neurons have long processing lag, with respect to their behavioral timescale (e.g. the reaction time of robber flies are even faster for prey capture, around 10-30 ms, and their sensory processing delay is around 18-28 ms) [79]. The common theme that they overcome processing lag in their global motion sensitive neurons suggests that they may all use prediction to satisfy various selective pressure from different survival-critical behaviors.

Gap junctions are prevalent throughout the brain in many species [80, 81]. In vertebrate visual systems, the retina also encodes predictive information near-optimally to potentially circumvent sensory processing delays [38, 82]. Initial evidence supports the notion that gap junctions are a key circuit element in improving signal transmission in retina: for example, gap junctions between directionally selective ganglion cells in the mouse retina result in lag-normalization [83], and the gap junctions present in cones and

bipolar cells improve the signal-to-noise ratio in their respective outputs [84]. Gap junctions can also rapidly regulate chemical synapses and improve sensitivity to correlated signals [85]. When processing stimuli with correlations between the past and the future (e.g. predictable motion), these mechanisms can support prediction to compensate for delays. In the central nervous system, gap junctions are versatile enough to support flexible hierarchical information processing in cortical circuits, as hypothesized in [86]. The ubiquitous evolutionary pressure to perform efficient prediction may shape nervous systems through this common circuit motif.

The brain carries out flexible, robust, and efficient computations at every moment as an organism explores and interacts with the external world. These computations are only possible through versatile mechanisms that operate under realistic behavioral constraints. We have shown that optimizing the transmission of predictive information in sensing systems is a useful way to interrogate the neural code. Given the presence of predictive information in sensory systems that evolved independently [38], our work supports the idea that predictive information may very well be a fundamental design principle that underlies neural circuit evolution. While we have dug into the specific mechanisms and representations that support this kind of efficient prediction for fast, natural and behaviorally critical motion processing in the fly visual system, the lessons learned may apply to a much larger class of neural sensing systems.

Materials and Methods

Data Availability

The paper is a theoretical work and does not contain experimental data. All the parameters and open source software packages required to reproduce our simulation and results are specified in this Materials and Methods section.

Ego-rotation for evasive flight maneuvers

We obtain ego-rotation stimuli from a large dataset of evasive flight maneuvers in *Drosophila* published in [20]. This dataset contains 82 traces of evasive trajectories when the flies face looming targets from all possible angles in their left visual field. All traces contain motion information (e.g., direction, velocity, etc.) from the emergence of the threat to the end of the evasive maneuver. In this dataset, the evasive flight trajectories are aligned at the beginning of the maneuver. In [20], they showed that both the speed and the expansion sizes of looming threats do not change the respective escape time courses and dynamics, i.e. this 40 ms evasive maneuver is the best drosphila can do. The duration of the evasive trajectories vary between 10-40 ms, with 65 out of 82 flights as long as 40 ms. We chose this dataset for two reasons: a) its sample rate (7500 fps) allows us to trace the activity of the VS network at the millisecond scale; b) it contains threats approaching the fly from angles spanning a full 180°, providing a well-sampled collection of the fly's behavioral repertoire.

Simulation of the in silico VS network

Our simulation uses a biophysically realistic simplified model of the VS network based on a reconstruction introduced in [21] (we used the modelDB python package introduced in [64]). This reconstruction models each VS cell with hundreds of dendritic compartments based on image stacks obtained by two-photon microscopy. Meanwhile, it implements the chain-like circuitry of the VS network by using both a) resistances connecting neighboring cells as axonal gap junctions [52, 53]; b) the negative

conductance between the VS1 and the VS10 to account for the reciprocal innhibition [54].

Compared to the detailed reconstruction, the simplified, biophysically realistic model introduced in [22] reduces all dendritic compartments into a single compartment while keeping other components intact. In the simplified model, an individual VS cell is represented by one dendritic compartment and one axonal compartment, respectively. All its parameters were determined by a genetic algorithm [22] so that this simplified model behaves roughly the same as the real VS network when given the same current injection [52, 53].

Both the dendritic and axonal compartments have their own conductances (g_{dend} and g_{axon}, respectively) and a connection conductance between them (shown as the $g_{dend/axon}$). This VS network model defines the receptive field (RF) of these dendritic compartments as a 2-D Gaussian with $\sigma_{azimuth} = 15°$ and $\sigma_{elevation} = 60°$, tiling along the anterior-posterior axis (e.g., the centers of these receptive fields are located at $10°, 26°, \cdots, 154°$ for VS1-10 respectively, see detailed configuration in [23]). Input from local motion detectors within the receptive field of an individual dendrite are integrated into the input current (shown as the arrow between Fig.1C and Fig.1D)). The neighboring axonal compartments of different VS cells are connected by gap junctions (shown as g_{GJ}), whereas VS1 and VS10 are connected by inhibitory chemical synapses. In our simulation, we set all conductance magnitudes using the same method as in [22]. Based on experimental findings from [51], we vary the magnitude of the GJ conductance between 0 and 1 μS.

For each of the 65 evasive traces that lasted a full 40 ms, we simulated 10,000 randomly generated natural scenes to obtain samples of the input (current arriving at dendrites) and output (axonal voltages) for subsequent analysis. In every simulation, we first generate the pseudo-3D visual "cube" (S1 Fig$A$), representing the environment to which our model fly visual system responds, by randomly selecting six images from the van Hateren dataset. We then rotate this natural scene cube according to the rotational motion during evasive maneuvers recorded in [20] (we sample the rotational motion at a $\Delta t = 1$ ms interval, and integrate the VS response at a smaller time step of 0.01 ms to guarantee numerical accuracy). This yields the optic flow pattern which we then project onto a unit sphere that represents the fly's retina, following the protocol introduced in [23, 64]. There are 5,500 local motion detectors (LMD) evenly distributed on that unit sphere. Each LMD contains two subunits that differ by 2° in elevation. Each VS dendrite takes as input the output of the LMDs that fall into its respective field to generate the input current to the model VS network. We then use the temporal average of the resulting axonal voltage $V_{past} = 1/T \int V_{past}(t)dt$. For the voltages just before the start of evasive maneuvers, we use the average from $t = -10 \sim 0$ ms, i.e. 0 ms is the start of evasive maneuvers.

Efficient encoding of predictive information

To predict the future input motion, the only input the VS network has is its dendritic input at past times up to the present, i.e. $\mathcal{I}_{past}$. Ideally, the VS network output represents the future motion in a specific form, Z, following the optimal encoding dictated by the solution to our information bottleneck problem. The bottleneck minimizes how much the representation retains about the past input $I(Z; \mathcal{I}_{past})$ and maximizes how much it encodes about the future input i.e. $I(Z; \mathcal{I}_{future})$. Formally, such encoding Z solves the following variational problem, prediction of its own input:

$$\mathcal{L}_{p(Z|\mathcal{I}_{past}),\beta} = I_{past:in} - \beta I_{future:in} \tag{5}$$

where β is the trade-off parameter between compression of information about the past, and retention of information about the future sensory input (we switch to $I_{future:\theta}$ when we look at the prediction of future ego-rotation, as shown in Section 4 of the Result). For each $I_{past:in}$, there exists an optimal $I^*_{future:in}(I_{past:in})$ which is the maximum $I_{future:in}$ possible for a specified $I_{past:in}$, determined by the statistics of the sensory input, i.e. $\mathcal{I}_{past}$, itself.

We use the following iterative (the Blahut-Arimoto algorithm [87]) algorithm (the MATLAB implementation is available at: `https://www.mathworks.com/matlabcentral/fileexchange/65937-information-bottleneck-iterative-algorithm` to find Z that optimizes Eq.(5): (we use $X = \mathcal{I}_{past}$ and $Y = \mathcal{I}_{future}$ here)

$$p_t(Z|X) = \frac{p_t(Z)}{Z(X,\beta)} \exp[-\beta \sum_Y p(Y|X)) \log \frac{p(Y|X)}{p_t(Y|z)}] \tag{6}$$

$$p_{t+1}(Z) = \sum_X p(X) p_t(z|X) \tag{7}$$

$$p_{t+1}(Y|Z) = \sum_X p(Y|X) p_t(X|Z) \tag{8}$$

Mutual information estimation

We use the k-nearest neighbor approach described in [88] (its open source software MILCA is available at `https://www.ucl.ac.uk/ion/milca-0`) to obtain mutual information estimates of $I_{future:in}$, $I_{future:limit}$, $I_{future:\theta}$ and $I_{past:in}$. Here, the mutual information is approximated via its corresponding complete gamma function:

$$I(X;Y) \quad = \psi(K) - < \psi(n_x + 1) + \psi(n_y + 1) > + \psi(N) \tag{9}$$

with N being the sample size, here $N = 650,000$. To choose k, we run blahut-arimoto algorithm with a large β (i.e. β=100) to estimate the upper bound of mutual information, based on the observation that the information bottleneck reaches its optimum at $I(X;Y)$. Each blahut-arimoto estimation takes up to one week, so we only obtained the upper bounds for $I_{future:in}$, $I_{future:\theta}$ and $I_{past:In}$ for the entire VS network and the VS5-6-7 triplet. We then use these upper bounds to determine k. In general, we use $k = 10, \cdots, 15$ (or $k = 1000, \cdots, 1100$ for those bootstrapped stimuli distributions) and calculate the mean as the estimate in our analysis. We omitted the standard deviations when we plotted Fig.2 and Fig.4 because of their small magnitudes (< 0.2).

Variational approximation of optimal encoding of the predictive information (VIB)

We use the variational approximation introduced in [67]. We first rewrite Eq.(5) as Eq.(10).

$$\mathcal{L}'_{p(z|\mathcal{I}_{past}),\beta'} = \quad I_{future:in} - \beta' I_{past:in} \tag{10}$$

The minimization of Eq.(5) is equivalent to the maximization of Eq.(10) (i.e. when $\beta' = \frac{1}{\beta}$, Eq.(10) is the same as Eq.(5)).

$$\mathcal{L}'_{p(z|\mathcal{I}_{past}),\beta} - \beta' H_{\theta_{future}} \geq \tag{11}$$
$$\mathcal{L}_{VIB} = \int dydzp(\mathcal{I}_{future}, Z) \log q(\mathcal{I}_{future}|Z)$$
$$-\beta' \int dcurr_{past} dzp(\mathcal{I}_{past}) p(Z|\mathcal{I}_{past}) \log \frac{p(Z|\mathcal{I}_{past})}{r(Z)}$$

Next, we minimize the variational lower bound Eq.(11) of Eq.(10). The advantage of using this variational approximation Eq.(11) is that we can constrain the distribution of Z to a particular form (i.e. a 2-D Gaussian) while letting the distributions of x and y to be arbitrary. This provides us with a latent feature representation of the lower bound for the optimal encoding of predictive information.

In this work, we would like to understand the structure of the optimal encoding for future ego-rotation given the input (the dendritic current, the VS axonal voltages, or the triplet voltages). Therefore, we obtain the respective solutions of $\mathcal{L}_{VIB}$ with fixed $\beta' = 40$. This is the value that falls into the diminishing return part of the IB curves in both Fig.3 and Fig.4. We also limit the dimension of Z to be 2 for direct comparison of inputs having different dimensions.

Acknowledgements

SW and IS and were supported by a grant from the Gatsby Charitable Foundation and by the Max Planck Hebrew University Center for Sensory Processing of the Brain in Action. The latter grant also supported AB. SEP was also supported by the National Science Foundation, both via CAREER award 1652617, and through the Center for the Physics of Biological Function (PHY-1734030). This work was also supported by NIH grant R01EB026943 (SEP). We thank D. Allan Drummond for providing the blowfly rendering in Figure 1.

References

1. Lopez-Schier H. Neuroplasticity in the acoustic startle reflex in larval zebrafish. Current opinion in Neurobiology. 2019;54:134–139.

2. Card G, Dickinson M. Visually mediated motor planning in the escape response of Drosophila. Current Biology. 2008;18(17):1300–7.

3. Schilling T, Borst A. Local motion detectors are required for the computation of expansion flow-fields. Biol Open. 2015;4(9):1105–8.

4. Hanlon R. Cephalopod behaviour. Cambridge University Press; 2018.

5. Dickinson MH. Death Valley,Drosophila, and the Devonian Toolkit. Annual Review of Entomology. 2014;59(1):51–72. doi:10.1146/annurev-ento-011613-162041.

6. Catania KC. Tentacled snakes turn C-starts to their advantage and predict future prey behavior. Proceedings of the National Academy of Sciences. 2009;106(27):11183–11187. doi:10.1073/pnas.0905183106.

7. Muijres FT, Elzinga MJ, Iwasaki NA, Dickinson MH. Body saccades of Drosophila consist of stereotyped banked turns. The Journal of experimental biology. 2015;218:864–875. doi:10.1242/jeb.114280.

8. Schilstra, Hateren. Blowfly flight and optic flow. I. Thorax kinematics and flight dynamics. The Journal of experimental biology. 1999;202 (Pt 11):1481–1490.

9. Hateren, Schilstra. Blowfly flight and optic flow. II. Head movements during flight. The Journal of experimental biology. 1999;202 (Pt 11):1491–1500.

10. Pringle JWS. The gyroscopic mechanism of the halteres of Diptera. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences. 1948;233(602):347–384. doi:10.1098/rstb.1948.0007.

11. Lehmann FO, Dickinson MH. The control of wing kinematics and flight forces in fruit flies (Drosophila spp.). Journal of Experimental Biology. 1998;201(3):385–401.

12. Cuntz H, Forstner F, Schnell B, Ammer G, Raghu SV, Borst A. Preserving Neural Function under Extreme Scaling. PLoS ONE. 2013;8(8):e71540. doi:10.1371/journal.pone.0071540.

13. Dickerson BH, de Souza AM, Huda A, Dickinson MH. Flies Regulate Wing Motion via Active Control of a Dual-Function Gyroscope. Current Biology. 2019;29(20):3517–3524.

14. Gnatzy W, Grunert U, Bender M. Campaniform sensilla of Calliphora vicina (Insecta, Diptera). Zoomorphology. 1987;106(5):312–319. doi:10.1007/bf00312005.

15. Chevalier RL. The fine structure of campaniform sensilla on the halteres ofDrosophila melanogaster. Journal of Morphology. 1969;128(4):443–463. doi:10.1002/jmor.1051280405.

16. Smith DS. The fine structure of haltere sensilla in the blowfly Calliphora erythrocephala (Meig.), with scanning electron microscopic observations on the haltere surface. Tissue and Cell. 1969;1(3):443–484. doi:10.1016/s0040-8166(69)80016-9.

17. Agrawal S, Grimaldi D, Fox JL. Haltere morphology and campaniform sensilla arrangement across Diptera. Arthropod Structure & Development. 2017;46(2):215–229. doi:10.1016/j.asd.2017.01.005.

18. Toh Y. Structure of campaniform sensilla on the haltere ofDrosophila prepared by cryofixation. Journal of Ultrastructure Research. 1985;93(1-2):92–100. doi:10.1016/0889-1605(85)90089-8.

19. Chan WP, Dickinson MH. Position-specific central projections of mechanosensory neurons on the haltere of the blow fly, Calliphora vicina. Journal of Comparative Neurology. 1996;369(3):405–418.

20. Muijres F, Elzinga M, Melis J, Dickinson M. Flies evade looming targets by executing rapid visually directed banked turns. Science. 2014;344:172–177.

21. Cuntz H, Haag J, Forstner F, Segev I, Borst A. Robust coding of flow-field parameters by axo-axonal gap junctions between fly visual interneurons. Proc Natl Acad Sci USA. 2007;104:10229–10233.

22. Weber F, Eichner H, Cuntz H, Borst A. Eigenanalysis of a neural network for optic flow processing. New Journal of Physics. 2008;10:015–013.

23. Borst A, Weber F. Neural action fields for optic flow based navigation: a simulation study of the fly lobula plate network. PLoS One. 2011;6(1):e16303.

24. Buschbeck K, Strausfeld N. The relevance of neural architecture to visual performance: phylogenetic conservation and variation in Dipteran visual systems. J Comp Neurol. 1997;383(3):282–304.

25. Fitzgerald JE, Clark DA. Nonlinear circuits for naturalistic visual motion estimation. eLife. 2015;4. doi:10.7554/elife.09123.

26. Kim AJ, Fenk LM, Lyu C, Maimon G. Quantitative Predictions Orchestrate Visual Signaling in Drosophila. Cell. 2017;168(1-2):280–294.e12. doi:10.1016/j.cell.2016.12.005.

27. Dickinson M, Muijres F. the aerodynamics and control of free flight manoeuvres in Drosophila. Phil Trans R Soc. 2016;371.

28. Klapoetke NC, Nern A, Peek MY, Rogers EM, Breads P, Rubin GM, et al. Ultra-selective looming detection from radial motion opponency. Nature. 2017;551(7679):237–241.

29. Ristroph L, Bergou AJ, Ristroph G, Coumes K, Berman GJ, Guckenheimer J, et al. Discovering the flight autostabilizer of fruit flies by inducing aerial stumbles. Proceedings of the National Academy of Sciences. 2010;107(11):4820–4824. doi:10.1073/pnas.1000615107.

30. Bergou AJ, Ristroph L, Guckenheimer J, Cohen I, Wang ZJ. Fruit flies modulate passive wing pitching to generate in-flight turns. Physical review letters. 2010;104:148101. doi:10.1103/PhysRevLett.104.148101.

31. v Buddenbrock W. Die vermutliche Losung der Halternfrage. Pfugers Arch. 1919;175:125–164.

32. W D. Physico-Theology. W.&J.Inny; 1714.

33. Chan WP, Prete F, Dickinson MH. Visual input to the efferent control system of a fly's "gyroscope". Science (New York, NY). 1998;280:289–292. doi:10.1126/science.280.5361.289.

34. Hengstenberg R. Mechanosensory control of compensatory head roll during flight in the blowflyCalliphora erythrocephala Meig. Journal of Comparative Physiology A. 1988;163(2):151–165. doi:10.1007/bf00612425.

35. Sherman A. A comparison of visual and haltere-mediated equilibrium reflexes in the fruit fly Drosophila melanogaster. Journal of Experimental Biology. 2003;206(2):295–302. doi:10.1242/jeb.00075.

36. Sherman A. Summation of visual and mechanosensory feedback in Drosophila flight control. Journal of Experimental Biology. 2004;207(1):133–142. doi:10.1242/jeb.00731.

37. Land M, Collett T. Chasing behaviour of houseflies (fannia canicularis). J Compara. 1974;89:331–357.

38. Palmer S, Marre O, Berry M, Bialek W. Predictive information in a sensory population. Proc Natl Acad Sci USA. 2015;112:6908–6913.

39. Berman G, Bialek W, Shaevitz J. Predictability and hierarchy in Drosophila behavior. Proc Natl Acad Sci USA. 2016;113:11943–11948.

40. Borst A, Helmstaedter M. Common circuit design in fly and mammalian motion vision. Nature neuroscience. 2015;18:1067–1076.

41. Clark DA, Demb JB. Parallel Computations in Insect and Mammalian Visual Motion Processing. Current biology : CB. 2016;26:R1062–R1072. doi:10.1016/j.cub.2016.08.003.

42. Joesch M, Schnell B, Raghu SV, Reiff DF, Borst A. ON and OFF pathways in Drosophila motion vision. Nature. 2010;468:300–304. doi:10.1038/nature09545.

43. Hassenstein B RW. Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. Z Naturforsch B. 1956;11:513–524.

44. Haag J, Denk W, Borst A. Fly motion vision is based on Reichardt detectors regardless of the signal-to-noise ratio. Proceedings of the National Academy of Sciences. 2004;101(46):16333–16338. doi:10.1073/pnas.0407368101.

45. Maisak MS, Haag J, Ammer G, Serbe E, Meier M, Leonhardt A, et al. A directional tuning map of Drosophila elementary motion detectors. Nature. 2013;500(7461):212–216.

46. Krapp HG, Hengstenberg R. Estimation of self-motion by optic flow processing in single visual interneurons. Nature. 1996;384(6608):463–466. doi:10.1038/384463a0.

47. Haag J, Borst A. Dendro-dendritic interactions between motion-sensitive large-field neurons in the fly. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2002;22:3227–3233. doi:20026316.

48. Mauss AS, Meier M, Serbe E, Borst A. Optogenetic and pharmacologic dissection of feedforward inhibition in Drosophila motion vision. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2014;34:2254–2263. doi:10.1523/JNEUROSCI.3938-13.2014.

49. Wertz A, Gaub B, Plett J, Haag J, Borst A. Robust coding of ego-motion in descending neurons of the fly. Journal of Neuroscience. 2009;29(47):14993–15000.

50. Borst A, Haag J, Mauss AS. How fly neurons compute the direction of visual motion. Journal of Comparative Physiology A. 2019; p. 1–16.

51. Borst A, Haag J. The intrinsic electrophysiological characteristics of fly lobula plate tangential cells: I. Passive membrane properties. Journal of computational neuroscience. 1996;3:313–336. doi:10.1007/bf00161091.

52. Haag J, Borst A. Neural mechanism underlying complex receptive field properties of motion sensitive interneurons. Nat Neurosci. 2004;7:628–634.

53. Haag J, Borst A. Dye-coupling visualizes networks of large-field motion-sensitive neurons in the fly. Journal of Comparative Physiology A. 2005;191(5):445–454. doi:10.1007/s00359-005-0605-0.

54. Haag J, Borst A. Reciprocal inhibitory connections within a neural network for rotational optic-flow processing. Frontiers in neuroscience. 2007;1:111–121. doi:10.3389/neuro.01.1.1.008.2007.

55. Elyada Y, Haag J, Borst A. Different receptive fields in axons and dendrites underlie robust coding in motion-sensitive neurons. Nat Neurosci. 2009;12(3):327–332.

56. Haag J, Wertz A, Borst A. Integration of Lobula Plate Output Signals by DNOVS1, an Identified Premotor Descending Neuron. Journal of Neuroscience. 2007;27(8):1992–2000. doi:10.1523/jneurosci.4393-06.2007.

57. Wang S, Borst A, Zaslavsky N, Tishby N, Segev I. Efficient encoding of motion is mediated by gap junctions in the fly visual system. Plos Comp Bio, vol 13, no 12, p e1005846. 2017;13:e1005846.

58. Borst A. Fly visual course control: behaviour, algorithms and circuits. Nature Reviews Neuroscience. 2014;15(9):590–599. doi:10.1038/nrn3799.

59. Kennedy A, Wayne G, Kaifosh P, Alviña K, Abbott L, Sawtell NB. A temporal basis for predicting the sensory consequences of motor commands in an electric fish. Nature neuroscience. 2014;17(3):416.

60. Tu MS, Dickinson MH. The control of wing kinematics by two steering muscles of the blowfly (Calliphora vicina). Journal of comparative physiology A, Sensory, neural, and behavioral physiology. 1996;178:813–830. doi:10.1007/bf00225830.

61. Heide G, Götz KG. Optomotor control of course and altitude in Drosophila melanogaster is correlated with distinct activities of at least three pairs of flight steering muscles. The Journal of experimental biology. 1996;199:1711–1726.

62. Lindsay T, Sustar A, Dickinson M. The Function and Organization of the Motor System Controlling Flight Maneuvers in Flies. Current Biology. 2017;27(3):345–358. doi:10.1016/j.cub.2016.12.018.

63. van Hateren J. A theory of maximizing sensory information. Biol Cybern. 1992;68:23–29.

64. Trousdale J, Carroll S, Gabbiani F, Josi K. Near-optimal decoding of transient stimuli from coupled neuronal subpopulations. JNeurosci. 2014;34:12206–12222.

65. Tishby N, Pereira FC, Bialek W. The Information Bottleneck Method; 1999.

66. Cuntz H, Borst A, Segev I. Optimization principles of dendritic structure. Theor Biol Med Model. 2009;8:4–21.

67. Alemi A, Fischer I, Dillon J, Murphy K. Deep variational information bottleneck Int. In: Conf. on Learning Representations; 2017.

68. Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, et al. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. In: International Conference on Learning Representations (ICLR); 2017. p. ICLR.

69. Chalk M, Marre O, Tkacik G. Relevant sparse codes with variational information bottleneck; 2016.

70. Kingma DP, Welling M. Auto-Encoding Variational Bayes; 2013.

71. Gronenberg W, Strausfeld NJ. Descending pathways connecting the male-specific visual system of flies to the neck and flight motor. Journal of comparative physiology A, Sensory, neural, and behavioral physiology. 1991;169:413–426. doi:10.1007/bf00197654.

72. Gronenberg W, Strausfeld NJ. Premotor descending neurons responding selectively to local visual stimuli in flies. The Journal of Comparative Neurology. 1992;316(1):87–103. doi:10.1002/cne.903160108.

73. Douglass JK, Strausfeld NJ. Anatomical organization of retinotopic motion-sensitive pathways in the optic lobes of flies. Microscopy Research and Technique. 2003;62(2):132–150. doi:10.1002/jemt.10367.

74. Strausfeld NJ, Bassemir UK. The organization of giant horizontal-motion-sensitive neurons and their synaptic relationships in the lateral deutocerebrum of Calliphora erythrocephala and Musca domestica. Cell and Tissue Research. 1985;242(3). doi:10.1007/bf00225419.

75. Strausfeld NJ, Seyan HS. Convergence of visual, haltere, and prosternai inputs at neck motor neurons of Calliphora erythrocephala. Cell and Tissue Research. 1985;240(3):601–615. doi:10.1007/bf00216350.

76. Strausfeld NJ, Seyan HS, Milde JJ. The neck motor system of the flyCalliphora erythrocephala. Journal of Comparative Physiology A. 1987;160(2):205–224. doi:10.1007/bf00609727.

77. Lehmann FO, Gotz KG. Activation phase ensures kinematic efficacy in flight-steering muscles of Drosophila melanogaster. Journal of Comparative Physiology A. 1996;179(3). doi:10.1007/bf00194985.

78. Dickson W, Straw A, Poelma C, Dickinson M. An Integrative Model of Insect Flight Control (Invited). In: 44th AIAA Aerospace Sciences Meeting and Exhibit. American Institute of Aeronautics and Astronautics; 2006.

79. Fabian ST, Sumner ME, Wardill TJ, Rossoni S, Gonzalez-Bellido PT. Interception by two predatory fly species is explained by a proportional navigation feedback controller. Journal of The Royal Society Interface. 2018;15(147):20180466. doi:10.1098/rsif.2018.0466.

80. Connors B. Synchrony and so Much More: Diverse Roles for electrical Synapses in Neural Circuits. Dev Neurobiol. 2017;77(5):610–624.

81. Marder E. Electrical synapses: Beyond speed and synchrony to computation. Current Biology. 1998;8:R795–R797.

82. Sederberg A, MacLean J, Palmer S. Learning to make external sensory stimulus predictions using internal correlations in populations of neurons. Proc Natl Acad Sci USA. 2018;115(5):1105–1110.

83. Trenholm S, Schwab DJ, Balasubramanian V, Awatramani GB. Lag normalization in an electrically coupled neural network. Nature neuroscience. 2013;16:154–156. doi:10.1038/nn.3308.

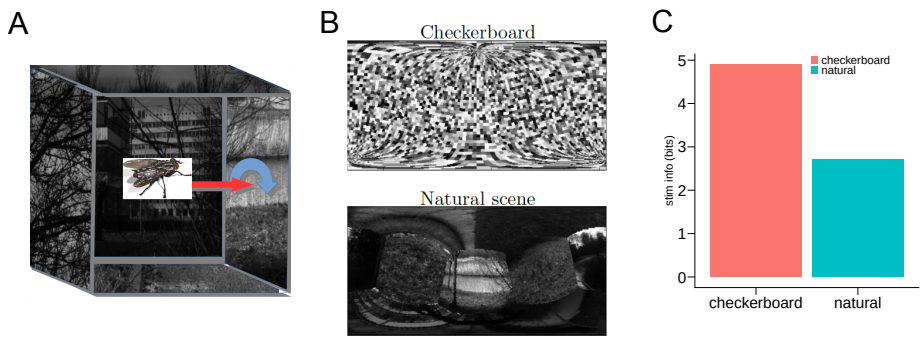
84. Ala-Laurila P, Greschner M, Chichilnisky E, Rieke F. Cone photoreceptor contributions to noise and correlation in the retinal output. Nat Neurosci. 2011;14:1309–1316.

85. Jacoby J, Nath A, Jessen Z, Schwartz G. A Self-Regulating Gap Junction Network of Amacrine Cells Controls Nitric Oxide Release in the Retina. Neuron. 2018;100(5):1149–1162.

86. Heeger D. Theory of cortical function. Proc Natl Acad Sci USA. 2017;114(8):1773–1782.

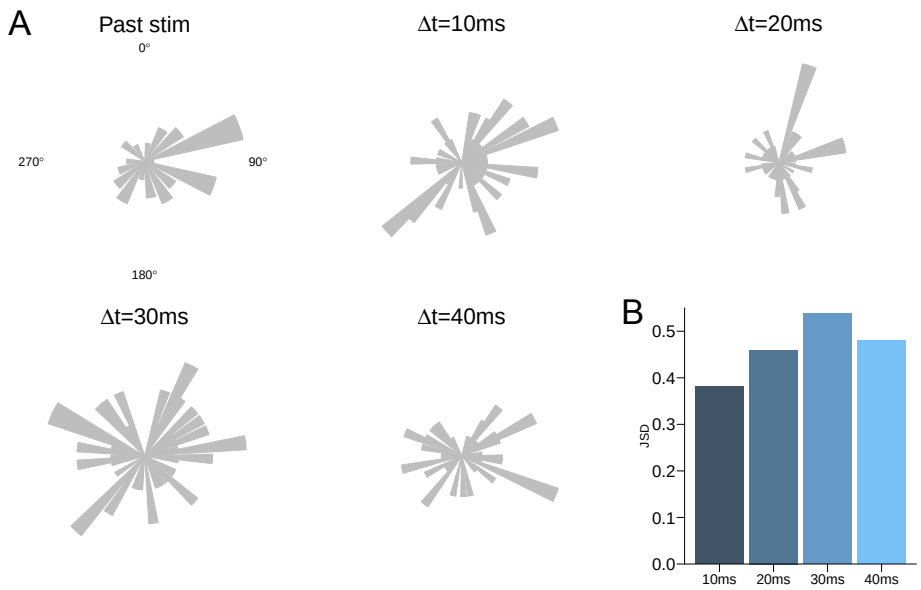
87. Blahut R. Computation of Channel Capacity and Rate-distortion Functions. IEEE Trans Inf Theor. 1972;18(4):460–473. doi:10.1109/TIT.1972.1054855.

88. Kraskov A, Stogbauer H, Grassberger P. Estimating mutual information. Phy Rev E. 2004;69:066138.

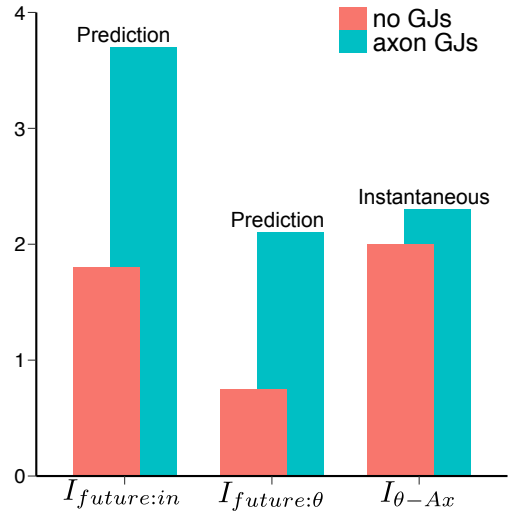


S1 Fig. A) Schematic depiction of the visual stimuli for the simulation, recompiled from [57]. Six natural images (five are shown here, with one excluded to reveal the fly's viewing perspective) were randomly selected from the van Hateren dataset [63]; each image was patched onto a different face of a cube. Assuming that the fly is located in the center of this cube, we obtain the visual experience of the fly's ego-rotational motion by rotating this cage around a particular motion direction shown by the dark blue arrow. We then project the moving natural scene cage to a unit sphere that represents the fly's retina, following the protocol introduced in [23, 64]. There are ∼5,500 local motion detectors (LMD) evenly distributed on this unit sphere. The responses of those LMDs whose locations are within a VS cell's dendritic receptive field ($\Sigma_{azimuth} = 15$ and $\Sigma_{elevation} = 60°$, tiling along the fly's anterior-posterior axis, see details in supplementary Materials and Methods) are then integrated as the input current to this particular VS cell. B) Mecator maps with both checkerbord and natural scene backgrounds, at 1° resolution in spherical coordinates. C) ego-motion information inferable in checkerboard and natural scene backgrounds. The stimulus is a constant rotation of $500°/s$ from [57]. Note that there is only half of the information about this motion stimulus using the background of natural scene textures compared to the checkerboard background.

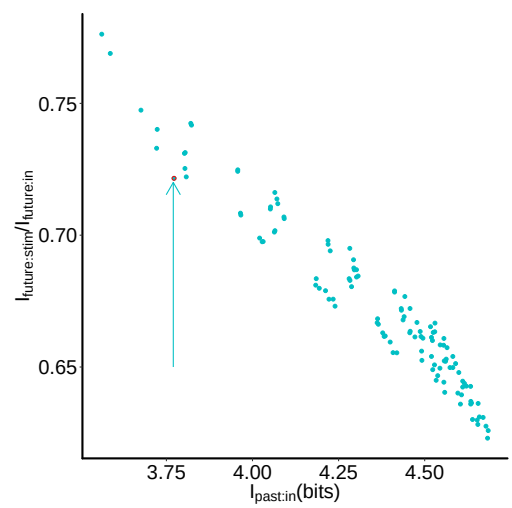
Supporting Information



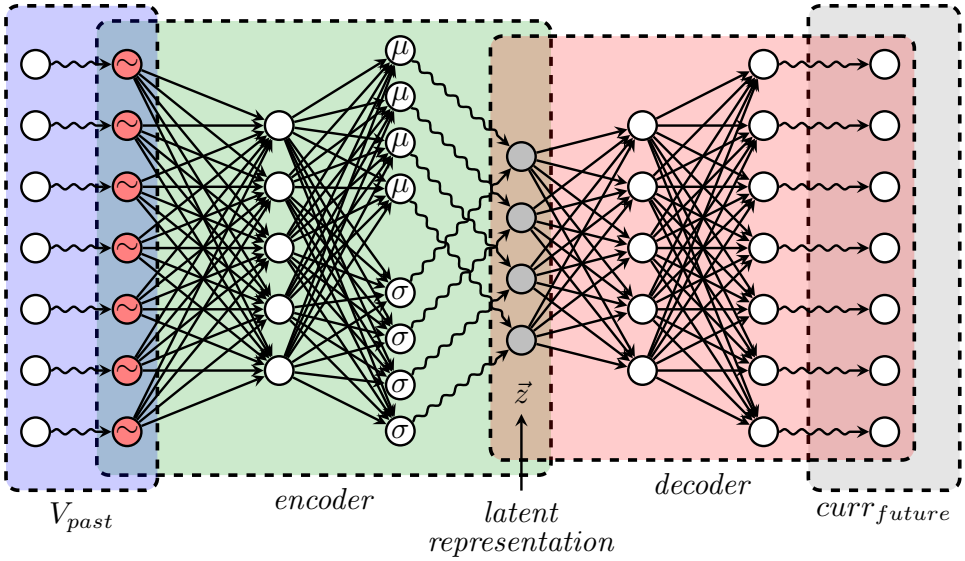
S2 Fig. Egorotation distributions for different time steps during the evasive maneuver. Egorotation distributions for different time steps during the evasive maneuver. Here we focus on the egorotations to which the VS network is sensitive. Because the VS network is only responsive to combinations of roll and pitch motions, i.e. motions within the fly's coronal plane, we represent all stimuli with their corresponding vectors in this plane. A) The egorotation distribution at 10ms before the onset of evasive maneuvers. B) Similar to A, but for the stimulus at 10ms after the initiation of evasive maneuvers. C) Similar to A, but for the egorotation at 20ms within the evasive maneuver. Here, most of the banked turns slow down and counter banked turns start.) D) Similar to A, but for the egorotation at 30ms within the evasive maneuver. This motion corresponds to the start of the counter-banked turn. E) Similar to A, but for egorotations a fly would experience at the end of the evasive maneuver. This motion corresponds to the slowing down of counter-banked turn and the completion of evasive maneuver. All of these egorotation distributions have comparable entropy $\sim 4 - 4.3$ bits. F) The Jensen–Shannon divergence between the past egorotation distribution and the egorotations at $\Delta t = 10, 20, 30, 40ms$ of evasive maneuvers, respectively



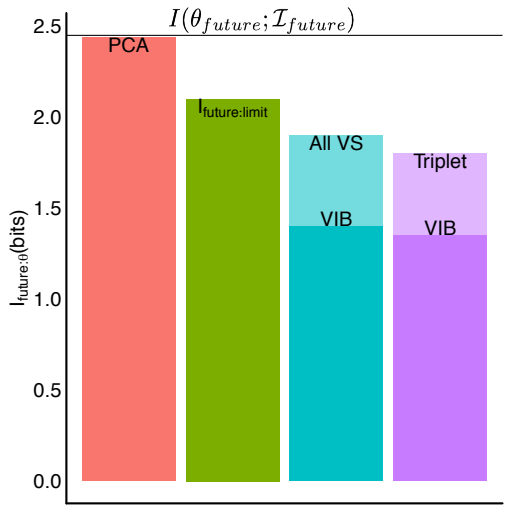
S3 Fig. The presence of axonal GJs improves the encoding of predictive information more than instantaneous input. We compare how the VS network encodes the predictive information in two scenarios (i.e. $I_{future:in}$ and $I_{future:\theta}$ with $\Delta t = 10$ ms) and the instantaneous egomotion θ with axonal GJs (cyan bars) and without GJs (red bars). The encoding of instantaneous constant egomotion $I_{\theta-Ax}$ (without prediction forward in time) is compiled from the previous work [57]. $I_{\theta-Ax}$ was defined as the mutual information between a constant rotation θ and the transient axonal voltages of the VS network (integrated for $\Delta t = 10$ ms).



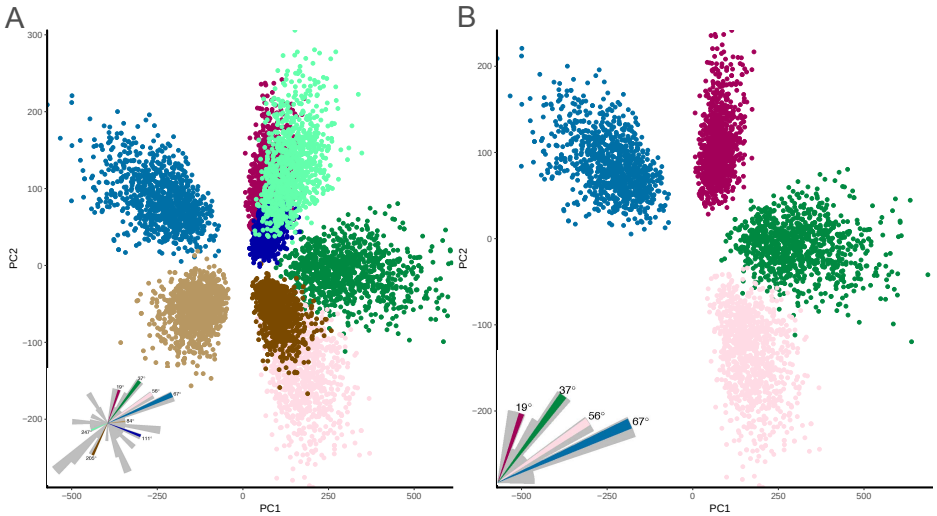
S4 Fig. How much a triplet based encoding retains from the past input vs. how much that information is about the future stimulus (out of the information about their own future input), for all 120 possible triplets. The particular VS 5,6,7 triplet (shown by the red circle and the arrow) that connects with the neck motor center, is one of the most efficient in terms of how much fraction its prediction of its own input is about the future stimulus, while its encoding cost $I_{past:in}$ is modest.



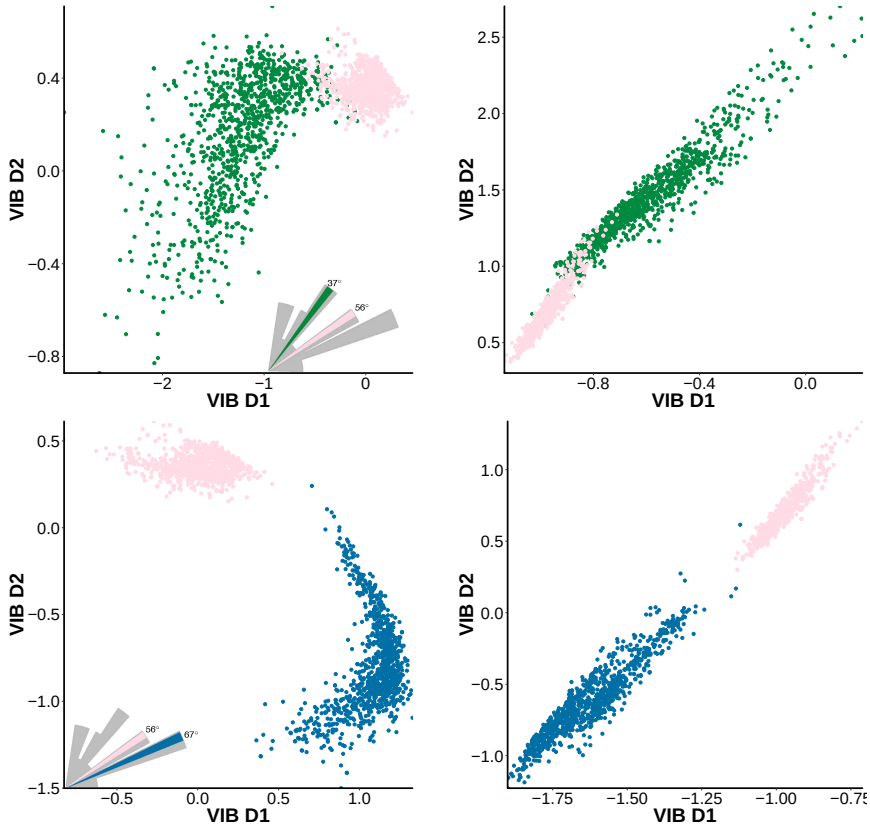
S5 Fig. Network schematic for the variational approximation of the information bottleneck solution (VIB). By constructing a variational approximation, the encoder learned a latent representation $\vec{z}$ from the past VS voltages. For training the encoder, we first project the axonal voltages of 20 VS cells to 200 intermediate filters, followed by a batch normalization layer. We then learn a latent representation with $z = 2$ for easy visualization. Then a decoder of the same structure as the encoder generates samples from $\vec{z}$ and reads them out as the future input current to the VS network. Note the VS network does not have direct access to the stimulus, it uses the correlations between its past and future inputs induced by the stimulus as a proxy for the stimulus correlations, themselves. $\vec{z}$ follows a Gaussian distribution, with parameters as μ and Σ. During training for this VIB, the mean μ and covariance matrix Σ of $\vec{z}$ map the axonal voltages of VS to the future input. When the VIB succeeds, we obtain the predictive representation of the future stimulus by projecting their respective axonal voltages into the latent feature space of $\vec{z}$.



S6 Fig. Predictive information for the future stimulus 10ms after the evasive maneuver starts ($\Delta t = 10ms$). The red bar shows that the PCA projection of the first 2PCs from the input current contains almost all of the stimulus information available at the input current itself. We use this PCA projection to understand whether it is possible to disentangle input stimuli from different quadrants using prediction in Fig.5. The green bar shows the limit on prediction information, based on the information bottleneck method. It corresponds to the point on information curve at the given compression in Fig.4B. The cyan bar corresponds to the predictive information about the future stimulus using outputs from all VS cells. The darker-colored region shows how much information the corresponding VIB captures about the future stimulus. The purple bar is similar to the cyan bar, for predictive encodings of the VS 5-6-7 triplet vs. their respective VIB solution.



S7 Fig. The input to the VS network only supports local discrimination. A) The representation of 8 randomly selected stimuli within the plane whose dimensions are the first two principal components of the input currents. Note that there are substantial overlaps between clusters: e.g. the light-green cluster is almost on top of the dark-red/dark-blue clusters. B) The subset of 4 stimuli from A. The only difference, as compared to A, is that all these stimuli have the same pitch/roll directions (clockwise roll and up tilt pitch, i.e. they are all within the 1st quadrant of the fly's coronal plane).



S8 Fig. The predictive information encoded by the VS network preferentially discriminates nearby egorotations. A) The predictive representation of stimuli at 37° and 56° obtained by mapping the respective axonal voltages of the entire VS network to the latent feature space generated by the VIB. B) Similar to A, but using the VS 5-6-7 triplet as input. C) The predictive representation of two stimuli that are much closer in stimulus space: 56° and 67°, respectively. Note that there is no overlap between these two nearby stimuli whereas there is some overlap for stimuli that are farther apart (shown in A). D) Similar to C, but using the VS 5-6-7 triplet as input.