Axonal gap junctions in the fly visual system enable fast prediction for evasive flight maneuvers

Siwei Wang[1,2,4*], **Alexander Borst**[3], **Idan Segev**[1,2], **Stephanie Palmer**[4*]

***For correspondence:**
siweiw@uchicago.edu (FMS);
sepalmer@uchicago.edu (FMS)

[1]Department of Neurobiology, Hebrew University of Jerusalem; [2]The Edmond and Lily Safra Center for Brain Sciences, Hebrew University of Jerusalem; [3]Max Planck Institute Neurobiology; [4]Department of Organismal Biology and Anatomy, University of Chicago

Abstract

The visual system must make predictions to compensate for inherent delays in its processing, yet little is unknown, mechanistically, about how prediction aids natural behaviors. Here we show that despite a 30ms intrinsic processing delay, the vertical motion sensitive (VS) network of the blowfly can achieve maximally efficient prediction. This prediction enables fine discrimination of input motion direction during evasive flight maneuvers, which last just 40ms. Combining a rich database of behavioral recordings with detailed compartmental modeling of the VS network, we further show that the VS network implements this optimal prediction with a specific wiring architecture; axonal gap junctions between the VS cells are crucial for optimal prediction during the short timespan of the evasive maneuver. Furthermore, a subpopulation output of the VS network can selectively convey predictive information about the future visual input to the downstream neck motor center. Our work links prediction to behavior, via its neural implementation.

Introduction

Escape from approaching predators is critical for survival. Such 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 (***Lopez-Schier, 2019***), fly escape take-offs (***Card and Dickinson, 2008***), to more ongoing behaviors that require continuous sensory processing (e.g. escape from a looming threat during flight (***Schilling and Borst, 2015***; ***Hanlon, 2018***)). In the latter kind of escape, here called an "evasive maneuver", the organism needs to update its motor output dynamically as the escape unfolds. Although most animals steer away from imminent threats, the trajectories of escapes, even for similar threats, are highly variable (***Domenici et al., 2011***). To a certain degree, such variability makes the escape maneuver unpredictable in the eyes of the predator. This feature prevents the predator from learning to anticipate a stereotyped escape pattern and foiling the escape. During ongoing behaviors, generating this kind of escape response requires actively controlled and finely stimulus-tuned maneuvers. Therefore, the escape trajectory is predictable to the animal because it must maintain good control of its motor plant. Here we investigate how fine-controlled variation is instantiated in evasive responses while animals are engaged in ongoing behaviors.

Insects, especially diptera, are excellent models for exploring this problem because precise measurements of their evasive behaviors are available, the neuronal circuits involved are known, and responses can be modeled in rich detail. During an escape, the fly elicits an in-flight evasive maneuver after 60 ms of visual-motor delay. This maneuver adjusts its heading, accelerating the fly

away from the threat via a banked turn followed by a counter-banked turn (***Muijres et al., 2014***). Previous work has shown that banking is the fastest way to change heading (***Dickinson and Muijres, 2016***). The unique challenge in the fly's evasive maneuver is that it lasts only 40 ms, which is roughly the time scale of the visual sensory delay (***Land and Collett, 1974***). Therefore, previous work hypothesized that the dynamic control of the evasive maneuver is only possible through mechanosensory feedback, i.e., via the haltere circuit (***Dickinson and Muijres, 2016***). However, to date, there is no quantitative evidence showing that such feedback alone suffices to guide in-flight escape responses.

Here we explore whether the fly can use an alternate strategy to control its evasive maneuver, based on predictions of its future visual input. We hypothesize that if the fly can accurately predict the future visual input that it will experience during the fluid banked and counter banked turns, it can use such information to actively control evasive flight maneuvers. We focus on how this visual prediction emerges in the vertical motion sensing (VS) network of the fly visual system, which is organized in four consecutive layers: retina, lamina, lobula and lobula plate. The VS network is located in the lobula plate; it contains 10 lobula plate tangential cells (the VS cells) in each hemisphere that integrate local motion signals from upstream processing. The VS network is dedicated to processing global rotational motion, like that incurred in banked turns (***Borst and Weber, 2011***). Not only is the VS network essential for generating proper optomotor responses, silencing this network also eliminates the fly's escape response (***Schilling and Borst, 2015***). This neural circuit has a repeating circuit motif: all the VS cells connect to cells with adjacent receptive fields via axonal gap junctions with a strikingly large ~ 1000 nS conductance (***Hagg and Borst, 1996***). That is 3-4 orders of magnitude stronger than gap junctions found in the vertebrate cortex (***Amsalem et al., 2016***).

The output from the VS network stems from subpopulations of adjacent cell triplets, which target different downstream areas. The VS network connects to the downstream neck motor center only through the VS 5-6-7 triplet of cells (***Hagg and Borst, 1996***; ***Haag et al., 2007***), with receptive fields located at the center fly's field of view. Previous work from our group showed that this wiring architecture, with adjacent cells connected via strong axonal gap junctions, coupled with local readout from the VS 5-6-7 triplet enables near-optimal encoding of constant speed rotations (***Wang et al., 2017***). In this work, we focus on the behavioral constraints imposed by fast evasive flights maneuvers, asking whether this same wiring architecture supports prediction on a fast enough time scale to guide behavior before visual or haltere feedback becomes available.

We also focus on bottom-up prediction (i.e., no feedback), because of the brief time scale of the evasive maneuver. We hypothesize that effective behavior must exploit temporal correlations between the past and future input stimuli to make these predictions. We show how this bottom-up prediction provides information about the future stimulus, subverting delays in the visual input stream. Encoding of optimal predictive information exists in the retina (***Palmer et al., 2015***) and may ensure fluid interaction with the external environment. It may also be important in the formation of long-term memory (***Berman et al., 2016***). Here we hypothesize that the presence of such bottom-up prediction in the VS network enables the fly visual system to control brief evasive maneuvers that are critical for fly survival.

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. Here instead we use numerical simulations of an experimentally constrained, biophysically realistic reconstruction of the VS network to investigate how the VS network might encode this kind of fast, predictive information. Based on analyzing how the output of the VS network encodes predictive information about the visual inputs during recorded evasive maneuvers, we first demonstrate that axonal gap junctions are essential for input prediction in the VS network during the evasive maneuver. Next, we show that this predictive information, present at the VS output, is near-optimal throughout the duration of the evasive maneuver. 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 the future stimulus, i.e., compressing the readout does not sacrifice how much a downstream pathway knows about the ongoing evasive maneuver. Finally, we show that the predictive information of the future stimulus can be used for fine scale discrimination between input motion directions. The encoding of such predictive information is especially beneficial for fine-tuning subsequent motor behavior.

Results

Visual prediction provides substantial information about motion without delay

Figure 1 shows why visual prediction is necessary for controlling the evasive flight maneuvers. ***Figure 1A*** shows a schematic trace of the inputs and delays in the system. Previous work identified that the feedback from the halteres only becomes available after a 20ms delay (***Muijres et al., 2014***; ***Dickinson and Muijres, 2016***), towards the end of the maneuver, during the counter-banked turn. Visual feedback would also be too slow, coming online only after 30ms, long after the banked turn is replaced by the counter-banked turn through active control (***Dickinson and Muijres, 2016***).

Compared to haltere feedback, visual prediction can encode more information about the stimulus (***Figure 1B***). To quantify this, we first estimate how much the halter feedback can encode about the stimulus i.e., the visual input resulting from the banked turn in the natural environment. In (***Fox et al., 2009***), it was shown that the haltere neurons can encode about 4.42 bits/spike about a uniformly distributed motion stimulus embedded in a visual scene composed of white noise. However, compared to white noise, the statistics of natural scenes only allow for the embedding of about half of the motion information, as we showed in (***Wang et al., 2017***). In addition, the distribution of rotational motion in the evasive maneuver is rather skewed (***Figure 1–Figure Supplement 1***), its entropy is only about half of the uniformly sampled motion (i.e., 4.3 bits vs.$\log_2(360) \approx 8.5$ bits for a uniform distributed rotational stimulus with 1° resolution). Therefore, only $4.42 \times 50\% \times 50\% \approx 1.1$ bits per spike are available from the haltere neuron during evasive maneuvers embedded in natural scenes (randomly generated from the van Hateren dataset (***van Hateren, 1992***)). This is roughly 25% of the stimulus entropy (comparing the red bar with the gray bar in ***Figure 1B***). Surprisingly, the predictive information based solely on the past visual input can potentially provide more information about the future stimulus during the evasive maneuver than the haltere feedback (***Figure 1B***, the blue bar versus the red bar). Here we define this stimulus-relevant predictive information contained in the output voltage from the fly VS network as:

$$\begin{aligned} I_{future:stim}(\Delta t) = I(V_{past}; stim_{future}) &= I(V_t; stim_{t+\Delta t}) \\ &= \iint_{V_t, stim_{t+\Delta t}} p(V_t) p(stim_{t+\Delta t}|V_t) \log_2 \frac{p(stim_{t+\Delta t}|V_t)}{p(V_t)}, \end{aligned} \tag{1}$$

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 visual stimulus. 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 100Hz (***Weber et al., 2008***), which corresponds to an integration step of at least 10ms (see Methods and Materials). Throughout this paper, we represent the future rotational stimulus $stim_{t+\Delta t}$, by its vector components ($\cos(stim_{t+\Delta t})$, $\sin(stim_{t+\Delta t})$). 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 $p(stim_{t+\Delta t})$, i.e., the stimuli distribution and $p(stim_{t+\Delta t}|V_t)$, i.e., the probability of the future stimulus conditioned on the past output axonal voltage to obtain $I_{future:stim}$ (see Methods and Materials). ***Figure 1B*** shows that the predictive information $I_{future:stim}$ in the VS output voltage about the future stimulus captures nearly 50% of the entropy of the future motion, much higher than the 25% encoded by the haltere feedback. Although our calculation of the bit rate we expect from the haltere is a back-of-the-envelope estimation, it suggests that the predictive information encoded by the VS network is a dominant information source for evasive flight behaviors, in the natural environment.

To evaluate $I_{future:stim}$, we need to approximate both the stimulus distribution and the respective output distribution of the VS network. To obtain the stimulus distribution, we generate 650,000 samples of visual field motion trajectories based on behavioral recordings published in (***Muijres et al., 2014***). 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 40ms (out of the total 92 published trajectories in (***Muijres et al., 2014***)). ***Figure 1C*** shows one exemplar visual experience of a particular evasive maneuver trajectory with the blue arrow as the instant rotation trajectory. Here, we obtain the "cage" of natural images for simulation by randomly selecting six images out of the van Hateren dataset (***van Hateren, 1992***) and patch them onto the six faces of a cube. Then we generate a movie mimicking the evasive flight in the natural environment by rotating this natural scene cage according to the measured rotations in the evasive flight trajectory (we do not use the translation component of the evasive maneuver in this simulation because previous work showed that the VS network is not sensitive to translation (***Borst and Weber, 2011***), also see Methods and Materials). We next project this movie onto 5,500 local motion detectors, whose responses are integrated as the input current of the VS network (***Figure 1D***). This simulation procedure is the same as that described in (***Trousdale et al., 2014***). Previously, we showed that the VS network can use its own transient response (10ms after the onset of stimuli) to encode the stimulus with constant rotational motion at a relatively high fidelity (***Wang et al., 2017***). Here we use a behaviorally realistic, highly variable visual input and explore its intrinsic correlation structure, to investigate how the fly's brain use this predictive information to make fast and accurate sensory predictions.

Axonal gap junctions enable prediction during evasive maneuvers

Figure 2 shows that the anatomical locations of the gap junctions have significant impact on the predictive encoding capability of the VS network. 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 (***Haag and Borst, 2004***). Here, we use correlations in the VS input current induced by the stimulus as a proxy for the stimulus correlations, themselves, which can be in turn used to encode predictive information about the future stimulus. 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 $I_{future:limit}$, is

$$\begin{aligned} I_{future:limit} &= I(curr_{past}; curr_{future}) = I(curr_t; curr_{t+\Delta t}) \\ &= \iint_{curr_t, curr_{t+\Delta t}} p(curr_t) p(curr_{t+\Delta t} | curr_t) \log_2 \frac{p(curr_{t+\Delta t} | curr_t)}{p(curr_t)}. \end{aligned} \tag{2}$$

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 encoded by the axonal voltage of the VS network from its own input as

$$I_{future:in} = I(V_{past}; curr_{future}) = I(V_t; curr_{t+\Delta t}). \tag{3}$$

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 stimulus. Causality dictates that the past axonal voltage can only obtain information about the future current from the past current, therefore $I_{future:limit}$ 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 ***Figure 2***, axonal gap junctions are necessary for the system to encode the maximal amount of predictive information about the input current. Namely, the $I_{future:in}$ (shown in pink) only approaches $I_{future:limit}$ (shown in green) when gap junctions are present to neighboring VS axons.

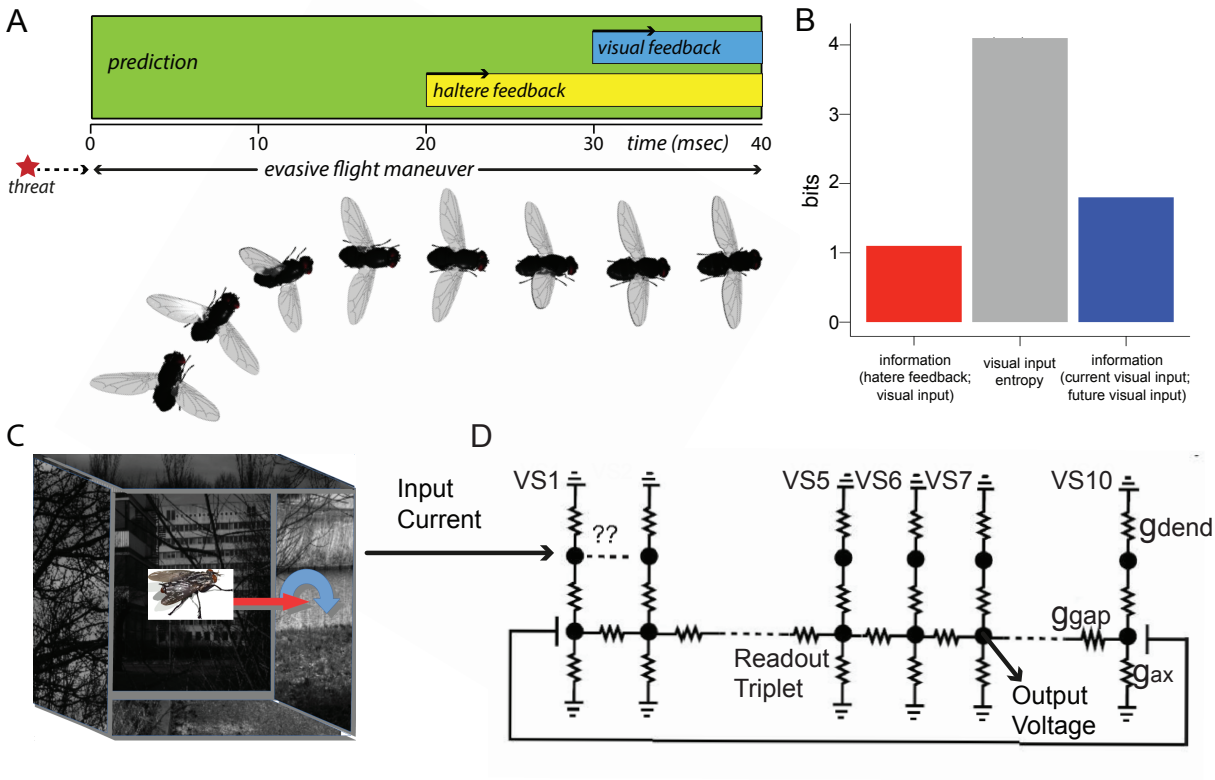

Figure 1. Predictive information is the dominant information source about visual inputs during evasive flight maneuvers. (A) Upon emergence of a threat (shown as the red star, dashed line represents the visual-motor delay of 60ms from the onset of threat to the start of the evasive maneuver), the fly performs an evasive maneuver by changing its heading through a banked turn. During the evasive maneuver, visual predictions can provide motion information throughout the entire duration, i.e., without delay (shown as the green zone), whereas the haltere feedback is only available after 20ms (shown as the yellow zone) and the visual feedback is only available after 30ms (shown as the blue zone). (B) This histogram compares how much information the haltere feedback (shown in red) and the visual prediction (shown in blue) can encode about the stimulus during the evasive maneuver, compared with the stimulus entropy (shown in gray). We use the stimulus distribution at $\Delta t = 10ms$ into the evasive maneuver to compute this entropy. Its distribution is shown in ***Figure 1–Figure Supplement 1A***. Note that the VS output contains more information about the future stimulus than the haltere feedback. (C) Schematic depiction of the visual stimuli for the simulation. 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 (***van Hateren, 1992***); 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 ~5,500 local motion detectors (LMD), which are randomly distributed on the fly's retina. The responses of these LMDs are then integrated as the input current to the VS network (shown as arrow to D). (D) A biophysically detailed model of the VS network, based on known neural circuitry (***Hagg and Borst, 1996***). 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.

Figure 1–Figure supplement 1. Stimulus distributions for different time steps during the evasive maneuver.

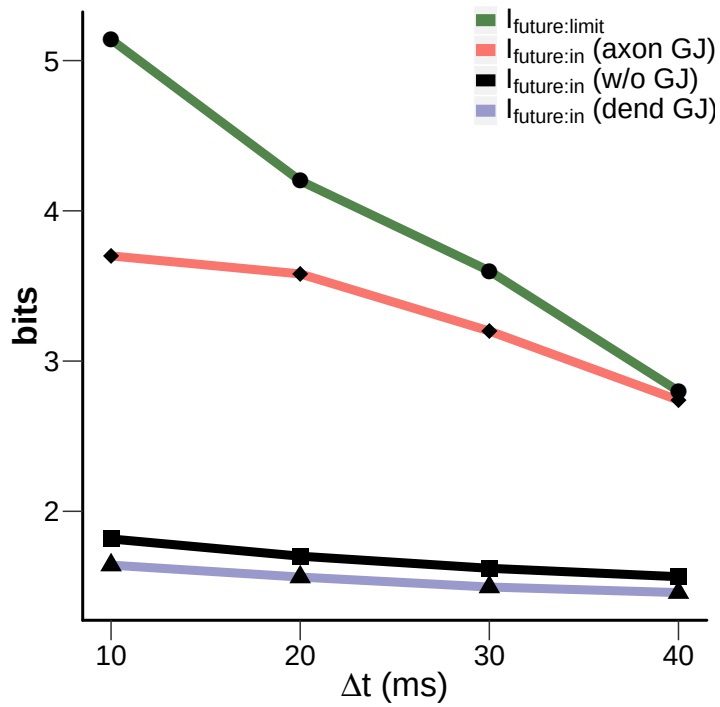

Figure 2. The capacity of the VS network to encode predictive information varies with the anatomical locations of the gap junction between VS cells. The predictive information about the future input current, $I_{future:in}$ encoded in four different schemes: 1) the past dendritic input current (in green, 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 (pink), 3) when the gap junctions are present between VS dendrites (purple) and 4) in the absent of gap junctions (black). 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 details in Methods and Materials.)

The other two configurations of gap junctions, i.e., no gap junctions or gap junctions at the dendrites (shown in black and purple, respectively), cannot encode as much predictive information. Previous work had shown that the axonal gap junctions implement an electrotonic segregation that results in the broadening of receptive fields at the axonal terminal of the VS cells, for constant rotational motion (***Cuntz et al., 2007***). Here, we further determine that such electrotonic segregation supports maximal predictive encoding in a realistic behavioral context.

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, $curr_{past}$. To quantify the efficiency of encoding, we not only need to quantify the benefit (i.e., the $I_{future:in}$), we also need to quantify the cost, which is how much the axonal voltage encodes about its input (at the same time in the past). We define this as another mutual information quantity,

$$I_{past:in} = I(V_{past}; curr_{past}) = I(V_t; curr_t). \tag{4}$$

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 (***Tishby et al., 2000***). Given the amount of information the axonal voltage encodes about the past sensory input, what is the maximal amount of information it can encode about the future input? Such an optimum $I^*_{future:in}(I_{past:in})$ traces out a bound (the dark blue line) in ***Figure 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 past encoding) 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 encoding architectures?

The predictive capacity of the VS network for its own future inputs is near-optimal, as shown in ***Figure 3A***, the axonal voltages of the VS network encode $I_{future:in} = 3.49 \pm 0.1$ bits for future inputs at $\Delta t = 10ms$ (the beginning of the banked turn). Considering that optimum is $I^*_{future:in}(I_{past:in}) = 3.59$ bits, using axonal voltages of all VS cells capture $I_{future:in}/I^*_{future:in} = 97.2\%$ of the optimal predictive information. Such optimality is also present for predicting visual inputs at $\Delta t = 40ms$ (the counter banked turn segment, corresponding to the end of the evasive maneuver, as shown in ***Figure 3B***).

Similarly, using only the axonal voltages from the triplets, prediction of the entire VS network's future input is also close to optimal. We show this as the cross in red, for predicting both the future right after the start of the evasive maneuver and towards the end of the evasive maneuver in ***Figure 3A*** and ***Figure 3B***, respectively. For example, 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 ***Figure 3A***. This suggests that all triplets achieve $89.8 \pm 1.5\%$ efficiency in encoding predictive information about the inputs $I_{future:in}/I^*_{future:in}$.

However, the near-optimality of the triplets are still inferior to that obtained by having the axonal voltages of all VS neurons in the network $(89.8\% \ vs. \ 97.2\%)$. All encodings based on triplets contain somewhat less absolute predictive information. The VS network achieves its best trade-off efficiency, the middle shoulder section between the quickly-rising phase and the diminishing-returns-phase of the bound $I^*_{future:in}(I_{past:in})$ in ***Figure 3B***, for encoding stimuli towards the end of the flight maneuver. Because the prediction of its own input is only a proxy of prediction for future stimulus, inferior optimality in predicting its own input does not necessarily map to the same degradation in the prediction of the future stimulus. In the next section, we explore how much this efficiency for predicting its own future input affects the encoding of future visual stimulus.

The triplet architecture selectively encodes predictive information about the future stimulus.

The triplet readout encoding architecture retains close to all of the available predictive information about the future stimulus available to the VS network at its input. Using the VS 5-6-7 triplet as an example (darker color bars in ***Figure 4A***), we can see that triplets capture most of the available predictive information about the future stimulus ($I_{future:stim}$, defined in Section 1 of the Results). Because downstream pathways of the VS network only readout 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 selects components within that output that are relevant for predicting the future stimulus, $I_{future:stim}$. 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.

Figure 4B shows that all triplets are near-optimal in encoding the predictive information about the future stimulus. Interestingly, such optimality is close to the niche region where the predictive information just begins to saturate (such optimality is also present for prediction of the distant future, i.e., $\Delta t > 10ms$, results not shown). Considering that the VS 5-6-7 triplet encodes nearly the same information about the future stimulus compared to the VS network (***Figure 4A***), the main benefit of using triplet is compression: despite encoding less predictive information about its own input, the VS triplet readout encoding retains nearly as much as possible about the future statistical structure of the ongoing maneuver, discarding information predictive of the less-behaviorally-relevant intrinsic dynamics of the inputs, themselves.

Although all triplets encode similar amounts of information about the future stimulus (the standard deviation of the $I_{future:stim}$ amongst all 120 triplets is just 0.1 bits), the particular triplet

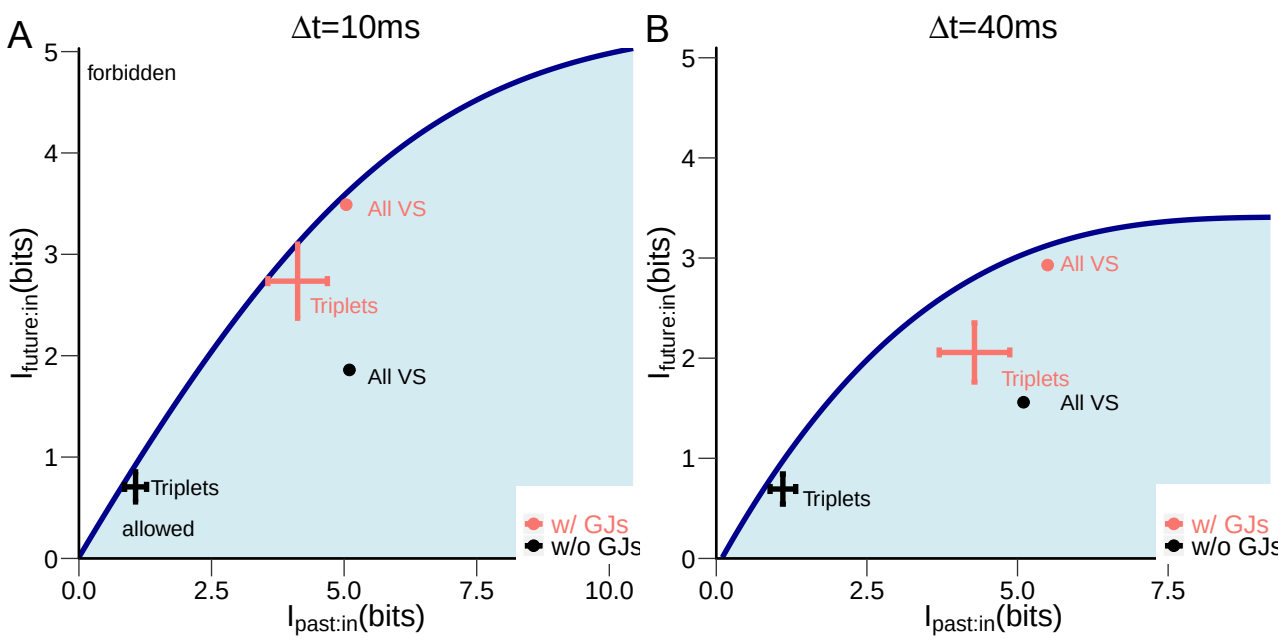

Figure 3. Near-optimal prediction of the input to the VS network. (A) The encoding of predictive information about the future current input to the VS network is near-optimal 10ms after the evasive maneuver starts ($\Delta t = 10ms$). Such performance is present for using both the entire VS network and 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 Methods and Materials). 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) Similar to A, but for prediction of the distant future: $\Delta t = 40ms$, corresponding to the stimulus at the end of evasive maneuver.

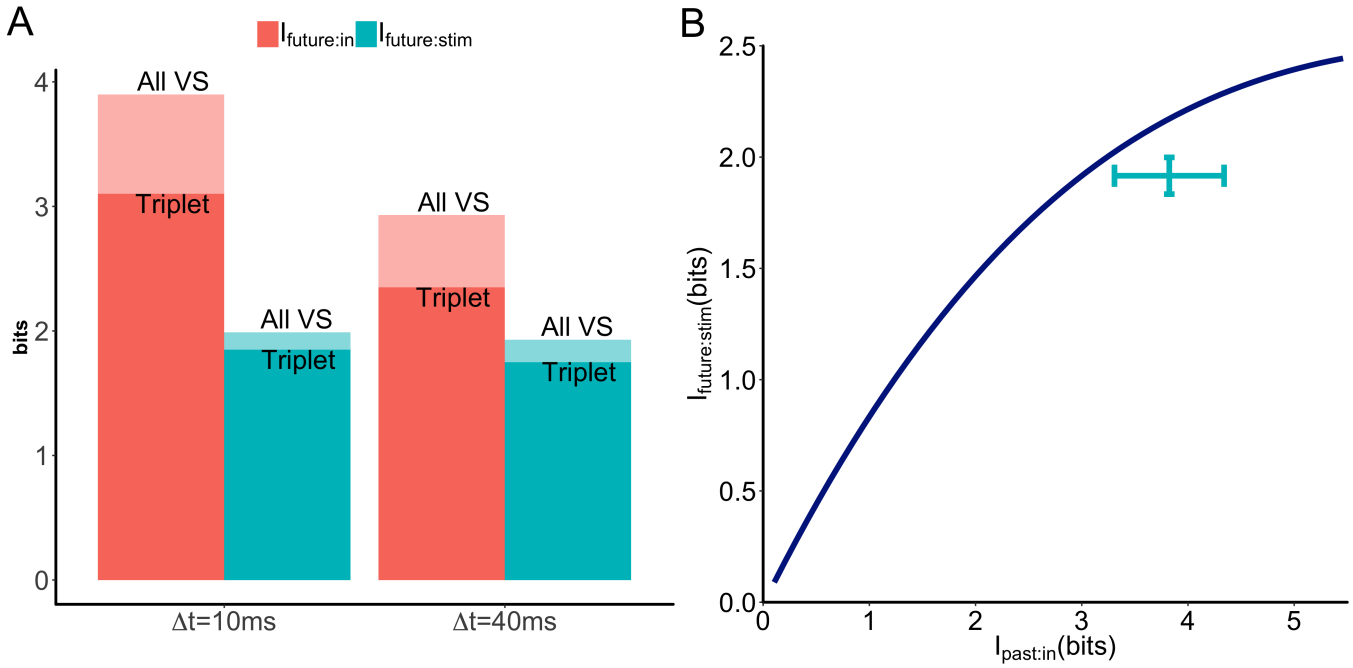

Figure 4. Encodings based on the axonal voltages of triplets are near-optimal in predicting the future stimulus. (A) Histogram showing that although the VS triplets encode less predictive information compared to a combination of all VS axonal voltages (shown in red), the triplets (we use the VS 5-6-7 triplet as an example here) encode nearly as much information *about the future stimulus* (shown in cyan) as the whole VS population. Here we show the comparison for the immediate future stimulus after the onset of the evasive maneuver, $\Delta t = 10ms$. Similar relationships also hold for the more distant future stimuli, e.g. $\Delta t > 10ms$ (not shown). (B) Similar to ***Figure 3A***: The encoding of predictive information for the future stimulus 10ms after the start of the evasive maneuver ($\Delta t = 10ms$). The dark blue curve traces out the optimum encoding of the future stimulus given varying amounts of information retained about the past input. The cyan cross corresponds to how much information each of all possible 120 triplets encode about the future stimulus vs. how much information they retain from the past input.

Figure 4–Figure supplement 1. How much a triplet readout 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.

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 stimulus it packs into its prediction of the future input (***Figure 4–Figure Supplement 1***. The most efficient triplet is VS 1-2-3. 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 (***Cuntz et al., 2009***) is important in selecting the simultaneously most predictive and most resource-efficient encoding.

Here we show that the VS 5-6-7 triplet is successful in retaining nearly all of the predictive information about the future stimulus compared to that encoded by the entire VS network. This result also clarifies 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 the stimulus representations vary based on either the entire VS network or the VS 5-6-7 triplet. We do this 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.

Comparing the stimulus representations encoded by the entire VS network and the VS 5-6-7 triplet is challenging. No direct comparison can be made between representations of different dimensions (20-D based on the entire VS network and 6-D based on the VS 5-6-7 triplet). We can, however, make use of our information plane analysis and fact that both schemes encode similar amount of predictive information. Points at the same y-value or similar y-values in the information plane have the same dimensionality in their compressed representations, Z (***Tishby et al., 2000***). Therefore, to compare encoding schemes, we can explore the structure of the encoding Z. While we were able to compute information quantities in Z as shown above, it is more complicated to derive the structure of of the mapping to Z, itself. Thankfully, recent work in machine learning and computational neuroscience guides the way forward. We can approximate the structure of the optimal encoding by finding a variational approximation (***Alemi et al., 2016***; ***Higgins et al., 2017***; ***Chalk et al., 2016***) to the information bottleneck (VIB) problem. It has been shown that this approximation is closely related to the loss function use to train variational autoencoders (***Kingma and Welling, 2013***). The VIB is a generative learning framework. Given pairs of the inputs and outputs, it generates a latent feature space whose dimensions are predictive features from the input to the output (***Figure 5–Figure Supplement 1***). One can then project the input into this latent feature space to obtain the predictive representation of the output. Therefore, by using the axonal voltage as input and the future input current as output (During the evasive maneuver, the VS network does not have direct access to the visual input. Instead, it uses the correlations between its past and future inputs induced by the stimulus as a proxy for the stimulus correlations, themselves) in training a VIB, we can explore the representation of the future stimulus encoded by the optimally predictive VS network at a fixed level compression (see Materials and Method). To allow for a direct comparison, we keep the dimension ($D = 2$) of the latent feature space to be the same while changing the input, using either the axonal voltages of the entire VS network, or those of the VS 5-6-7 triplet.

The representations of the future stimulus generated by the VIB (***Figure 5***), reveal that the predictive information encoded by the VS network supports fine-scale discrimination of the input motion direction. We obtain these predictive representations in two steps: first we 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. Next, we project input voltages that correspond to the same future stimulus onto this latent space. We can label these points in the latent space by their future stimulus value, and repeat this procedure for several different stimulus values. We can visually and computationally examine how overlapping or distinct these maximally predictive stimulus clusters are in the latent space of the VIB. Based on these predictive stimulus representations, we can understand what is being computed during the evasive maneuver. The ***Figure 5A*** shows a predictive representation for stimuli with different degrees of clockwise roll and up-tilt pitch (i.e., their corresponding directional vectors are located in the 1st quadrant in the fly's coronal plane). The ***Figure 5B*** shows a similar predictive representation using the axonal voltages of the VS 5-6-7 triplet as input. In both predictive representations, the clusters corresponding to different stimuli are distinguishable. Such discrimination also applies to stimuli combining counter-clockwise roll and up tilt, i.e., corresponding to vectors within the 4th quadrant of the fly's coronal plane (***Figure 5C*** for using the entire VS network as input and ***Figure 5D*** for using the VS 5-6-7 triplet as input, respectively). However, these predictive representations cannot discriminate stimuli 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 ***Figure 5C***) will entirely cover the cluster corresponding to 19° (also shown in magenta, but in ***Figure 5A***). The same overlap is also present in ***Figure 5B*** and ***Figure 5D***. Therefore, both the VS network and the VS 5-6-7 triplet support a similar fine-scale discrimination between closely-related stimuli. This similarity agrees with our previous result (***Figure 4***) that both the VS network and the VS

5-6-7 triplet, as well as their respective VIB approximations (***Figure 5–Figure Supplement 4***), retain almost the same amount of predictive information about the future stimulus.

Interestingly, such discrimination preferentially disentangles nearby stimuli. For example, a pair of stimuli (56° and 67°, shown in ***Figure 5–Figure Supplement 4A*** and ***Figure 5–Figure Supplement 4B***) that are just 10 degrees apart are mapped to distinct, well-separated clusters in the latent space of the VIB. Conversly, another pair (19° and 37°) that are farther apart share some overlap (***Figure 5A*** and ***Figure 5C***). The VS 5-6-7 triplet preserves this fine scale discrimination (***Figure 5B*** and ***Figure 5D***) while compressing the readout. During the evasive maneuver, this fine-scale discrimination can aid in fine tuning of banking, allowing the fly to carefully adjust its heading change based on the input threat. We hypothesize that the predictive information encoded by the VS network can help the downstream neck motor center to actively control the evasive maneuver. Because the predictive representations generated by VIB contain about 75% of the predictive information in their inputs (***Figure 5–Figure Supplement 2***), this result provides a lower estimate of what the biological networks can encode: because both the VS network and the VS 5-6-7 triplet encode more predictive information than their respective VIB approximations, they should be able to perform at or better in the fine-scale discrimination achieved by the VIB's, as shown in ***Figure 5***.

Even with a representation that retains all of the stimulus information in the input to the VS network, one cannot discriminate stimuli in vastly different directions. We construct such a representation based on the instantaneous input current of the present stimuli. These input currents contain 2.44 bits of the stimulus information at the same point in time, i.e. without prediction forward in time. This information is higher than that available via predictive information encoding of the past input current (2.1 bits, shown as the red bar in ***Figure 5–Figure Supplement 2***). The first two principal components (PC) of the input current retain nearly all available stimulus information, so we ask whether disparate stimuli can be disentangled in this subspace of the instantaneous stimulus representation. Hence, we obtain a representation retaining all stimulus information by projecting all VS input currents into these first 2 PCs. We find that there still exists substantial overlaps between stimuli (***Figure 5–Figure Supplement 3***), e.g. the cluster of 19° in magenta almost covers the entire cluster of 247° in light green (***Figure 5–Figure Supplement 3A***). This shows that the input to the VS network can only support fine-scale discrimination, whether an instantaneous readout or predictive readout. This also means that based on prediction, the VS network can only perform fine-scale discrimination. Therefore, it is possible that the integration at the neck motor center from multiple sources (VS network, haltere and prosternal organs (***Buschbeck and Strausfeld, 1997***) combines information from other pathways to discriminate stimuli with larger direction differences.

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 can encode 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 local input prediction with the prediction of the future stimulus: while it encodes the future input somewhat sub-optimally, it retains the more behaviorally-important predictive stimulus information at higher fidelity, in all triplets. The encoding of predictive information has a concrete behavioral goal: it enables fine-tuning of motion discrimination during the evasive maneuver.

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 stimuli are best at enabling fine-scale discrimination of nearby stimuli, we have shown how structure maps to function in this motion sensing system. In addition, we have shown that behaviorally relevant features of the stimulus are faithfully encoded via circuit mechanisms at the sensory end of the arc from sensation to action.

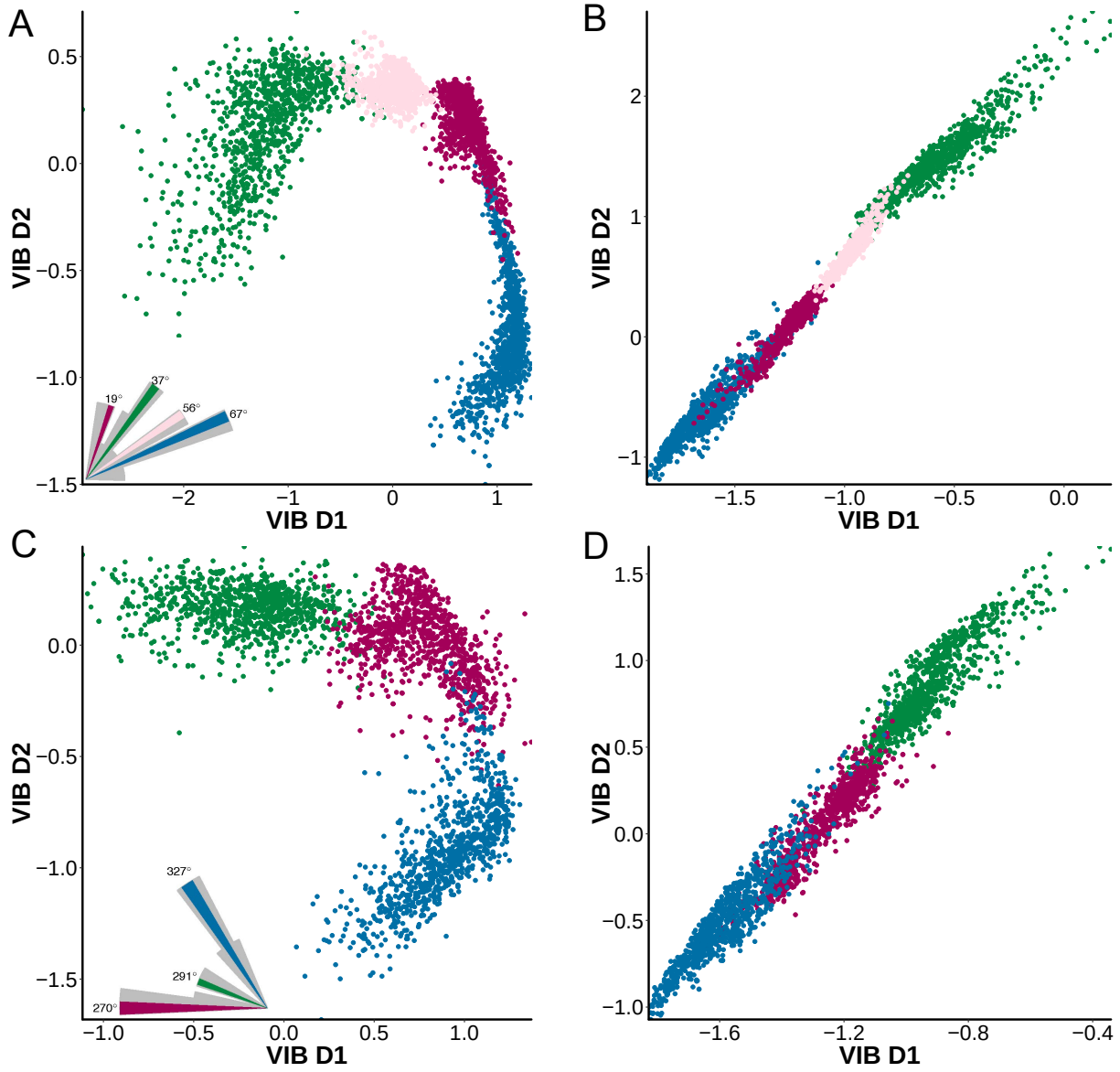

Figure 5. The predictive information encoded by the VS network supports fine scale discrimination of the future stimulus. (A) The predictive representation of four future stimuli in the same quadrant of roll and pitch, e.g. an up-tilt and a clockwise roll. This representation maps the axonal voltage of the entire VS network to the future stimulus 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 stimuli 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 stimuli in color. (B) Similar to A but using the axonal voltages of the VS 5-6-7 triplet. (C) Similar to A, but the motion stimuli 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.

Figure 5–Figure supplement 1. Network schematic for the variational approximation of the information bottleneck solution (VIB)

Figure 5–Figure supplement 2. Predictive information for the future stimulus 10ms after the evasive maneuver starts ($\Delta t = 10ms$)

Figure 5–Figure supplement 3. The input to the VS network only supports local discrimination.

Figure 5–Figure supplement 4. The predictive information encoded by the VS network preferentially discriminates nearby stimuli.

This suggests that behavioral goals sculpt neural encoding even at the earliest stages of sensory processing.

Although our results show that the near-optimal prediction is present in the VS network of the blowfly, further investigation is necessary to identify whether this generalizes to other dipterans. First, not all dipterans use VS cells for flight control, e.g. predatory flies like *Holcocephala* and *Efferia* do not have VS cells. Instead, they use extra, complementary horizontal motion sensitive (HS) cells to compute rotational motion (***Buschbeck and Strausfeld, 1997***); second, other dipterans may have flight behaviors non-exist in blowflies, i.e., stationary hovering or ballistic interception. These maneuvers provide drastically different selective pressure on their respective tangential neurons. However, all dipterans' lobula plate tangential neurons are strongly lagged with respect to their behavioral timescales (e.g. the reaction time of robber flies are even faster for prey capture, around 10-30ms, and their sensory processing delay is around 18-28ms) (***Fabian et al., 2018***).

Gap junctions are prevalent throughout the brain in many species (***Connors, 2017***; ***Marder, 1998***). In vertebrate visual systems, the retina also encodes predictive information near-optimally to potentially circumvent sensory processing delays (***Palmer et al., 2015***; ***Sederberg et al., 2018***). 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 (***Trenholm et al., 2013***), and the gap junctions present in cones and bipolar cells improve the signal-to-noise ratio in their respective outputs (***Ala-Laurila et al., 2011***). Gap junctions can also rapidly regulate chemical synapses and improve sensitivity to correlated signals (***Jacoby et al., 2018***). 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 (***Heeger, 2017***). 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 (***Palmer et al., 2015***), 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.

Methods and Materials

Self-motion stimuli for evasive flight maneuvers

We obtain self-motion stimuli from a large dataset of evasive flight maneuvers in drosophila published in (***Muijres et al., 2014***). 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. The duration of the evasive trajectories vary between 10-40ms, with 65 out of 82 flights as long as 40ms. 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 model VS network

We use the model VS network (***Figure 1D***) first introduced in (***Weber et al., 2008***) to simulate how the blowfly visual system encodes motion during evasive flight maneuvers. This model uses two compartments to describe an individual VS cell. A previous study (***Weber et al., 2008***) showed that by having the same input current, the modeled VS cells yield the same axonal voltage output as real VS neurons. Both the dendritic and axonal compartments have their own conductances (g_{dend} and g_{ax}, respectively) and a connection conductance between them (shown as the g_{de-ax}). 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. Input from local motion detectors within the receptive field of an individual dendrite are integrated into the input current (shown as the arrow between ***Figure 1C*** and ***Figure 1D***)). The neighboring axonal compartments of different VS cells are connected by gap junctions (shown as g_{gap}), whereas VS1 and VS10 are connected by inhibitory chemical synapses. In our simulation, we set all conductance magnitudes using the same method as in (***Weber et al., 2008***). Based on (***Hagg and Borst, 1996***), we vary the magnitude of the GJ conductance between 0 and 1 μS.

In every simulation, we first generate the pseudo-3D visual "cube" (***Figure 1C***) representing the environment to which our model fly visual system responds, by randomly selecting six images from the van Hateren dataset. Next, we rotate this cube according to the rotational motion during evasive maneuvers recorded in (***Muijres et al., 2014***). We sample the rotational motion at a $\Delta t = 1ms$ interval. This yields the optic flow pattern that we then project into the 5000 local motion detectors (LMD) in our model visual system. Each LMD contains two subunits that differ by 2° in elevation. They are randomly distributed in a sphere mimicking the visual range of the fly. 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 voltage just before the start of the evasive flight maneuver, we use $t = -10 \sim 0ms$, because 0ms is the start of evasive maneuver. For each of the 65 evasive traces that lasted a full 40ms, 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.

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., $curr_{past}$. 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 $I(Z; curr_{past})$ and maximizes how much it encodes about the future input i.e., $I(Z; curr_{future})$. Formally, such encoding Z solves the following variational problem, prediction of its own input:

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

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 $I_{future:stim}$ when we look at the prediction of the future stimulus, 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., $curr_{past}$, itself.

We use the following iterative (the Blahut-Arimoto algorithm (***Blahut, 1972***)) algorithm to find Z that optimizes ***Equation 5***:

$$p_t(Z|curr_{past}) = \frac{p_t(Z)}{Z(curr_{past}, \beta)} \exp[-\beta \sum_{curr_{future}} p(curr_{future}|curr_{past}) \log \frac{p(curr_{future}|curr_{past})}{p_t(curr_{future}|z)}] \tag{6}$$

$$p_{t+1}(Z) = \sum_{curr_{past}} p(curr_{past}) p_t(z|curr_{past}) \tag{7}$$

$$p_{t+1}(curr_{future}|Z) = \sum_{curr_{past}} p(curr_{future}|curr_{past})p_t(curr_{past}|Z) \tag{8}$$

Mutual information estimation

We use the k-nearest neighbor approach described in (***Kraskov et al., 2004***) to obtain mutual information estimates of $I_{future:in}$, $I_{future:limit}$, $I_{future:stim}$ and $I_{past:in}$. Here, the mutual information is approximated via its corresponding complete gamma function:

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

with N being the sample size, here $N = 650,000$. Given the skewed stimulus distributions shown in ***Figure 1–Figure Supplement 1***, we vary $k = 10, \cdots, 15$ and use the mean as the estimate in our analysis.

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

We use the variational approximation introduced in (***Alemi et al., 2016***). We first rewrite ***Equation 5*** as:

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

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

Next, we minimize the following variational lower bound of ***Equation 10***:

$$\begin{aligned} \mathcal{L}'_{p(z|curr_{past}),\beta} - \beta' H_{stim_{future}} \geq \mathcal{L}_{VIB} = \int dy dz p(curr_{future}, Z) \log q(curr_{future}|Z) \\ -\beta' \int dcurr_{past} dz p(curr_{past}) p(Z|curr_{past}) \log \frac{p(Z|curr_{past})}{r(Z)} \end{aligned} \tag{11}$$

The advantage of using this variational approximation of ***Equation 10*** 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 the future stimulus 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 ***Figure 3*** and ***Figure 4***. We also limit the dimension of Z to be 2 for direct comparison of inputs having different dimensions.

Acknowledgement

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).

References

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

Alemi, A. A., Fischer, I., Dillon, J. V., and Murphy, K. (2016). Deep Variational Information Bottleneck. *arXiv e-prints*, page arXiv:1612.00410.

Amsalem, O., Geit, W., Muller, E., Markram, H., and Segev, I. (2016). From neuron biophysics to orientation selectivity in electrically-coupled networks of neocortical L2/3 large basket cells. *Cerebral Cortex*.

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

Blahut, R. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inf. Theor.*, 18(4):460–473.

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

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

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

Chalk, M., Marre, O., and Tkacik, G. (2016). Relevant sparse codes with variational information bottleneck. *arXiv e-prints*, page arXiv:1605.07332.

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

Cuntz, H., Borst, A., and Segev, I. (2009). Optimization principles of dendritic structure. *Theor Biol Med Model.*, 8:4–21.

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

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

Domenici, P., Blagburn, J., and JP., B. (2011). Animal escapology I: theoretical issues and emerging trends in escape trajectories. *J Exp Biol.*, 214:2463–73.

Fabian, S., Sumner, M., Wardill, T., Rossoni, S., and Gonzalez-Bellido, P. (2018). Interception by two predatory fly species is explained by a proportional navigation feedback controller. *J R Soc Interface.*

Fox, J., Fairhall, A., and Daniel, T. (2009). Encoding properties of haltere neurons enable motion feature detection in a biological gyroscope. *Proc Natl Acad Sci USA*, 107:3840–3845.

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

Haag, J., Wertz, A., and Borst, A. (2007). Integration of lobula plate output signals by DNOVS1, an identified premotor descending neuron. *J Neurosci.*

Hagg, J. and Borst, A. (1996). The intrinsic electrophysiological characteristics of fly lobula plate tangential cells. *J Comput Neurosci.*

Hanlon, R. (2018). *Cephalopod behaviour*. Cambridge University Press.

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

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A. (2017). beta-vae: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations (ICLR)*, page ICLR.

Jacoby, J., Nath, A., Jessen, Z., and Schwartz, G. (2018). A self-regulating gap junction network of amacrine cells controls nitric oxide release in the retina. *Neuron*, 100(5):1149–1162.

Kingma, D. P. and Welling, M. (2013). Auto-Encoding Variational Bayes. *arXiv e-prints*, page arXiv:1312.6114.

Kraskov, A., Stogbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Phy. Rev. E*, 69:066138.

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

Lopez-Schier, H. (2019). Neuroplasticity in the acoustic startle reflex in larval zebrafish. *Current opinion in Neurobiology*, 54:134–139.

Marder, E. (1998). Electrical synapses: Beyond speed and synchrony to computation. *Current Biology*, 8:R795–R797.

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

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

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

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

Tishby, N., Pereira, F. C., and Bialek, W. (2000). The information bottleneck method. *arXiv e-prints*, page physics/0004057.

Trenholm, S., Schwab, D., Balasubramanian, V., and Awatramani., G. (2013). Lag normalization in an electrically coupled neural network. *Nat Neurosci.*

Trousdale, J., Carroll, S., Gabbiani, F., and Josi, K. (2014). Near-optimal decoding of transient stimuli from coupled neuronal subpopulations. *J.Neurosci.*, 34:12206–12222.

van Hateren, J. (1992). A theory of maximizing sensory information. *Biol Cybern*, 68:23–29.

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

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

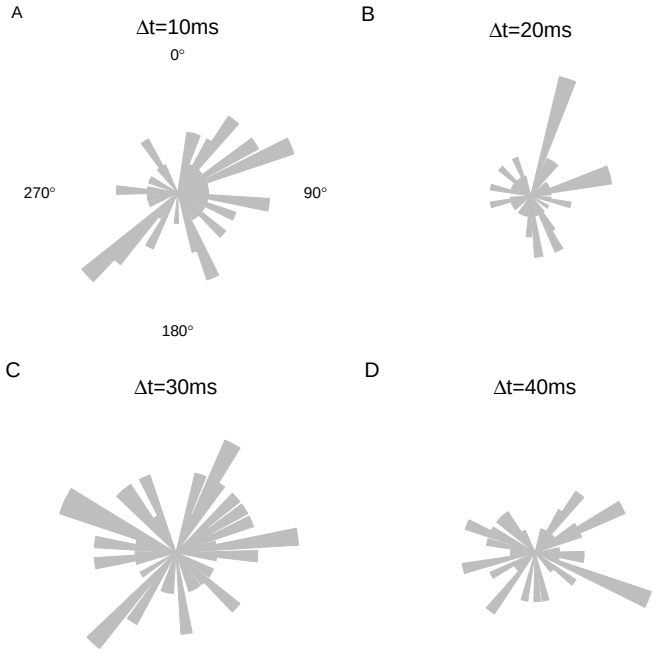

Figure 1–Figure supplement 1. Stimulus distributions for different time steps during the evasive maneuver. Here we focus on the stimuli to which the VS network is sensitive. Because the VS network is only responsive to combinations of roll and pitch motions, i.e., motions within the fly's coronal plane, we represent all stimuli with their corresponding vectors in this plane. A) The stimulus distribution at 10ms after the initiation of the evasive maneuver. B) Similar to A, but for the stimulus at 20ms after the start of the evasive maneuver. Here, most of the banked turns slow down and counter banked turns start. C) Similar to A, but for the stimulus at 30ms after the start of the evasive maneuver. This motion corresponds to the start of the counter-banked turn. D) Similar to A, but for the stimulus at 40ms after the start of the evasive maneuver. This motion corresponds to the slowing down of counter-banked turn and the completion of evasive maneuver. All of these stimulus distributions have entropy $\sim 4 - 4.3$ bits.

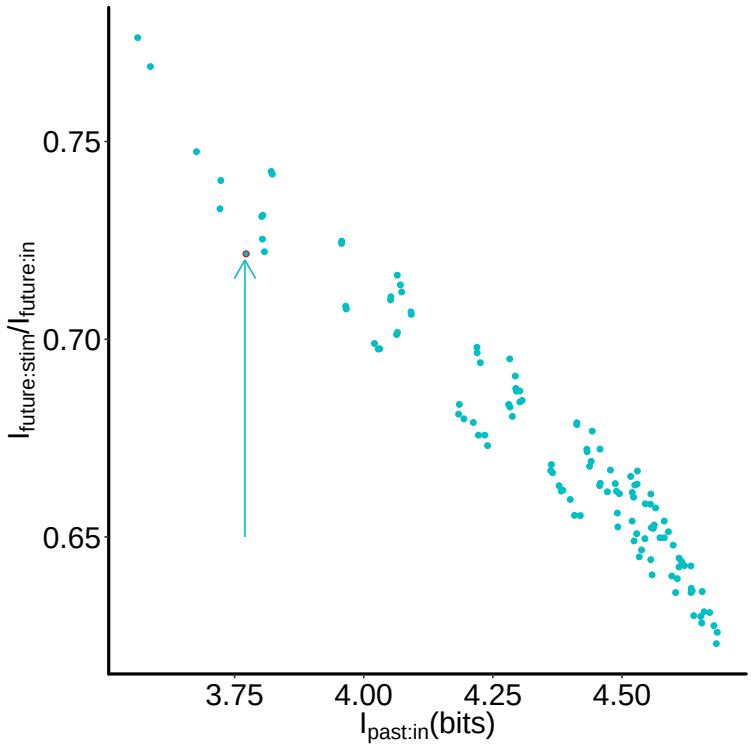

Figure 4–Figure supplement 1. How much a triplet based encoding retains from the past input vs. how much that information is about the future stimulus (out of the information about their own future input), for all 120 possible triplets. The particular VS 5,6,7 triplet (shown by the red circle and the 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.

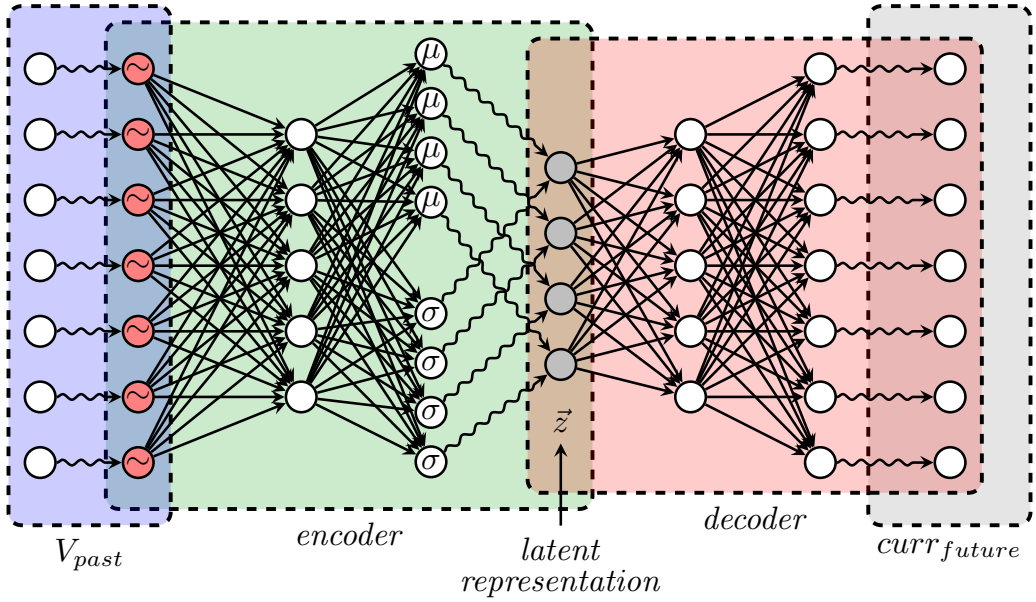

Figure 5–Figure supplement 1. Network schematic for the variational approximation of the information bottleneck solution (VIB). By constructing a variational approximation, the encoder learned a latent representation $\vec{z}$ from the past VS voltages. W generates samples from $\vec{z}$ and reads them 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 $\vec{z}$.

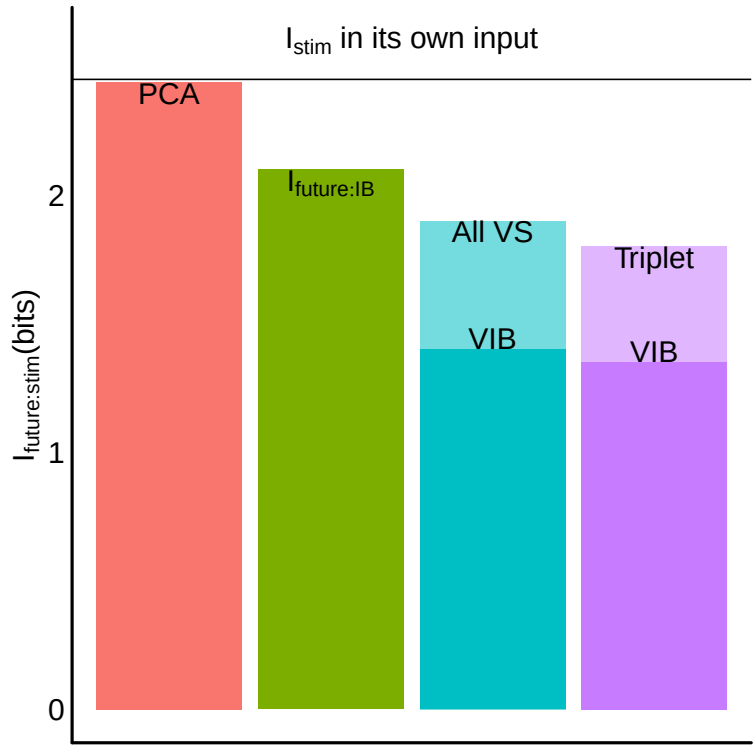

Figure 5–Figure supplement 2. Predictive information for the future stimulus 10ms after the evasive maneuver starts ($\Delta t = 10ms$). The red bar shows that the PCA projection of the first 2PCs from the input current contains almost all of the stimulus information available at the input current itself. We use this PCA projection to understand whether it is possible to disentangle input stimuli from different quadrants using prediction in **??**. 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 **Figure 4*B***. 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.

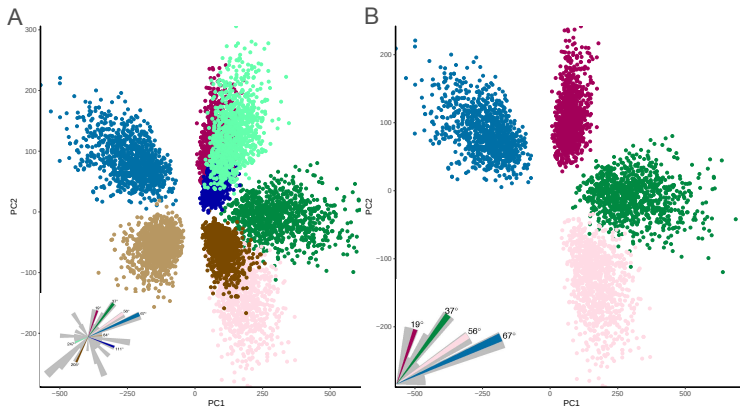

Figure 5–Figure supplement 3. The input to the VS network only supports local discrimination. A) The representation of 8 randomly selected stimuli within the plane whose dimensions are the first two principal components of the input currents. Note that there are substantial overlaps between clusters: 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).

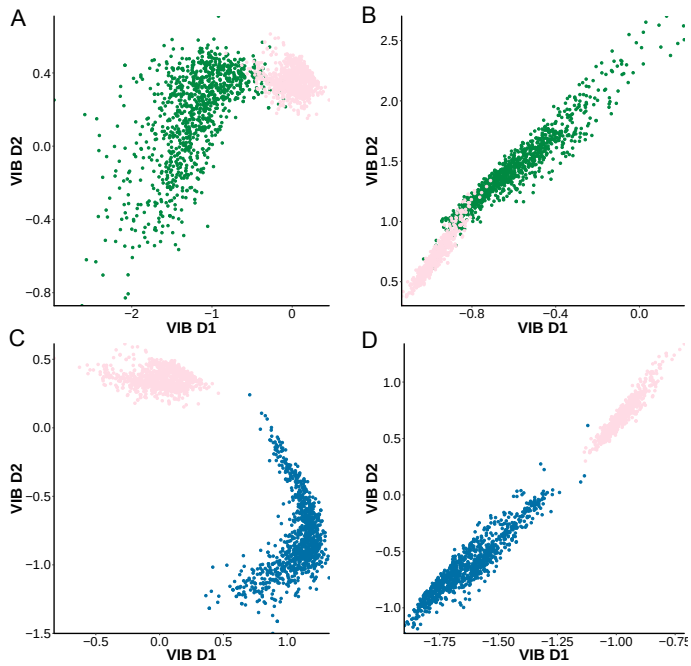

Figure 5–Figure supplement 4. The predictive information encoded by the VS network preferentially discriminates nearby stimuli. A) The predictive representation of stimuli at 19° and 37° obtained by mapping the respective axonal voltages of the entire VS network to the latent feature space generated 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.