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

Siwei Wang[a,b,d,1], **Alexander Borst**[c], **Idan Segev**[a,b], **and Stephanie Palmer**[d]

[a]Department of Neurobiology, Hebrew University of Jerusalem; [b]The Edmond and Lily Safra Center for Brain Sciences, Hebrew University of Jerusalem; [c]Max Planck Institute Neurobiology; [d]Department of Organismal Biology and Anatomy, University of Chicago

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 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 how the VS network implements this optimal prediction. The axonal gap junctions between the VS cells are crucial for optimal prediction during the short timespan of evasive maneuvers. Its subpopulation output further selectively conveys predictive information about the future visual input to the downstream neck motor center. Our work predicts novel sensory-motor pathways that link prediction to behavior.

Optimality | Neural coding | Information theory | insect vision

Escape from predators is critical 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), fly escape take-offs (2)), to more ongoing behaviors that require continuous sensory processing (e.g. escape from a looming threat during flight (3, 4)). In the latter kind of escape, here called an 'evasive maneuver', the organism must adjust its motor tape of '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. For flying animals, the inflight evasive maneuvers emerged as early as flight itself (5). Such in-flight evasive maneuvers require active control. This is because: 1) Their execution must be tuned according to where the threat appears with respect to the fly; 2) Fly flight is aerodynamically unstable and requires active control; and 3) Active control allows for evasive trajectories that are not 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 stimulus-tuned in-flight evasive maneuvers, at a time scale needed for survival.

Insects, especially diptera, are excellent models for exploring this problem. For the animal models in this work, the blowfly and the drosophila, both precise measurements of the motion executed during their evasive behaviors (7) and mechanistic level understanding of the underlying neuronal circuits (8, 9) are available. Furthermore, insect brain architecture is highly conserved. Flies in particular use similar neural architecture to drive similar behaviors (5). Because the escape behaviors and the corresponding sensory-motor circuits controlling them emerged as early as flight itself, 400 million years ago (5), many modern arthropod species inherited this core sensory-behavioral module. In particular, both blowflies and drosophila use banked turns to change their heading direction, during both stereotypical saccadic flights (10–12) and visually induced turns (7, 13). Despite their size differences (a blowfly is roughly four times larger than a drosophila), recent work has shown that the electrotonic structures and neuronal functions of motion sensitive neurons between these two animals are analogous (14). Meanwhile, the banked turns they make with their bodies are similar in shape (10–12), while drosophila show a slower overall profile of angular velocity (11, 15). In addition, drosophila do not use additional head rotations to stabilize their retinal input, possibly because of the lower acuity in their compound eyes compared to the blowfly. Because the only precise measurements of the fly's evasive maneuver are available in drosophila (7) and the only mechanistic understanding of the neural circuit that processes the visual input during these maneuvers is from the blowfly (8, 9, 16), but the two species are so similar overall (17), we use the behavioral measurements from drosophila to investigate how the blowfly's motion sensing circuit extracts behaviorally relevant information for this survival-critical maneuver. Blowflies have higher angular velocity during their banked turns and higher acuity in their compound eyes, thus we hypothesize that the blowfly must perform evasive maneuvers in a more precise and finely controlled manner than drosophila. Hence, this investigation will show the minimum necessary computation that the blowfly's neural circuit must execute during fast evasive maneuvers.

Drosophila are capable of performing in-flight evasive maneuvers after a sensory-motor delay (e.g., this delay is 60±20ms in (7)). This delay enables the fly to recognize the position of the looming threat and initiates the appropriate evasive ma-

Significance Statement

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.

The authors declare no competing interest.

[2]To whom correspondence should be addressed. E-mail: sepalmer@uchicago.edu

DRAFT

neuvers as banked turns. These evasive maneuvers use a mere 40ms to reorient the animal's heading and accelerate it away from the threat. Previous work showed that such maneuvers consist of visually-guided banked turns: an initial rotation followed immediately by an active counter-rotation. This indicates that even within the brief 40ms time frame of the escape response, sensory-motor circuits in the fly brain continuously transform visual information into the motor commands that control flight.

We set out to explore how visual information from escaping a purely visual threat is used to sculpt the fly's evasive maneuver. Because the fly visual system has a 30ms processing lag (18), it is unlikely that evasive maneuvers use visual information through feedback given their brief, 40ms time span. Previous work (7) hypothesized that either evasive maneuvers are guided by a single feed-forward program or that the fly uses mechanosensory feedback from the halteres to sculpt the active counter-bank turns during the escape. Haltere steering neurons make direct electric synapses with motor neurons of wing steering muscles (19–21) that can cause fast alterations of wing kinematics (22, 23) with only a 15-20ms delay. Thus, the halteres could initiate the active counter-bank turns, about halfway through the fly's escape maneuver. Recent work have shown that these haltere neurons use visual input during flight to regulate their activity (24). This suggests that the feed-forward output of the fly visual system at the beginning of evasive maneuver may be used by the halteres to regulate evasive flight roughly 20ms after the evasive maneuver starts. Since the evasive maneuver is a fast and large repositioning of the fly's body and flight path, this would only be a useful visual input if it contained significant information about the future state of the fly's posture and heading. In this work, we investigate how the fly uses predictive information from the visual system for active flight control.

Specifically, we hypothesize that if the fly can predict the future visual input that it will experience during the evasive maneuver, it can use such information to actively control the banked and counter banked turns that it will execute. Given the exceptionally short time scale of the evasive maneuver, we focus on the bottom-up prediction (i.e., no feedback). To be effective, this prediction exploits the temporal correlations between past and future visual stimuli during evasive maneuvers. Said another way, 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 (7). 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. We show how this bottom-up prediction provides information about future sensory input, subverting delays in the visual input stream. Encoding of optimal predictive information exists in the vertebrate retina (25) and may ensure fluid interaction with the external environment. It may also be important in the formation of long-term memory (26). Here we hypothesize that such bottom-up prediction in the fly visual system enables it to control brief evasive maneuvers that are critical for survival.

Because the banked/counter-banked turns of the evasive maneuvers are combinations of pitch and roll rotations, we focus on how visual prediction emerges in the vertical motion sensing (VS) network of the fly visual system, i.e. the sensory system dedicated to encoding these specific rotation angles (16). The fly visual system is organized in four consecutive layers: retina, lamina, lobula and lobula plate. The VS network consists of 10 lobula plate tangential cells (the VS cells) in each compound eye. It receives retinotopically organized local motion inputs and outputs global motion information in its axonal voltages. Each VS cell has its dendritic receptive field center at a specific rotational axis of the fly's coronal plane. They are numbered VS1-VS10 along the fly's anterior-posterior axis according to their receptive field location. Not only is the VS network essential for generating proper optomotor responses, silencing this network also eliminates the fly's escape response (3). The output from the VS network arises from subpopulations of adjacent cell triplets, which target different downstream areas (16, 27). 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 (27, 28), which have dendritic receptive fields located at the center of the fly's field of view.

The VS network has a chain-like structure (8, 28–30). Each VS cell only connects with other VS cells having immediately neighboring receptive fields. Meanwhile, the VS1 and VS10 cells show reciprocal inhibition (31). Previous dual-recording experiments (29) showed that VS cells connect amongst each other through electrical synapses. Further dye-coupling experiments showed that these electrical synapses were gap junctions (30). In (8), they further 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 motion stimuli (8, 32) at the output of the VS network. 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 speed rotations (33). In this work, we focus on behavioral constraints imposed by fast evasive flights maneuvers, asking whether this same wiring architecture supports visual prediction on a fast enough time scale to efficiently guide evasive maneuvers.

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. 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 (8, 9). All single-cell (28) and neuronal circuitry parameters (8, 29, 30) are set such that this compartmental reconstruction behaves as does the real VS network when presented with the same current injection (8, 9, 29, 30). 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 (34) used a synthetic population model of the electrosensory lobe to show how accurate cancellation of self-generated electric signals is achieved.

Based on how the output of the VS network encodes pre-

DRAFT

$$I_{future:stim}(\Delta t) = I(V_{past}; stim_{future}) \\ = I(V_t; stim_{t+\Delta t}) \quad [1]$$

dictive 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 evasive maneuvers. Next, we show that this predictive information, present at the output of the VS network, 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 show that the predictive information about future visual inputs 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. Our results also predict the existence of a novel sensory-motor pathway between the visual system and a visually gated motor neuron of wing steering muscles, namely the second basalare motor neurons (M.b2) (35–38).

Results

Visual prediction provides substantial information about motion without delay. Fig.1 shows 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 system (Fig.1A). Previous works showed that the haltere outputs reach the wing steering muscles after a 15-20ms delay (7, 39), towards the second half of the maneuver, right before the active counter-banked turn starts. Visual feedback would also arrive too late, coming online only after 30ms, long after the banked turn is replaced by the counter-banked turn through active control (39).

To quantify how much visual prediction encodes about the stimulus (the fly's future escape trajectory) (Fig.1B), we define this stimulus-relevant predictive information in the output voltage from the fly VS network as $I_{future:stim}$ (Eq.(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 (9), which corresponds to an integration step of at least 10ms (see Materials and Methods). 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 Materials and Methods). Fig.1B shows that the predictive information $I_{future:stim}$ in the VS output voltages capture nearly 50% of the entropy of the future escape trajectory. 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 $I_{future:stim}$, 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 (7). 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 (7)). 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 (40) 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 (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 (16), 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 (16, 41). There are 5,500 local motion detectors located on this unit sphere, whose responses are integrated as the input current of the VS network (Fig.1D).

The specific texture of the visual scene significantly impact the ego-motion-induced estimates input to the VS dendrites. Of course, if the fly rotates within a uniform background, no visual motion information is available. The details of the correlation structure in the scene also affect this estimate. Previous work (8) 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. They found 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 (33) 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 regular textures (shown in Fig.S1*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 predictive encoding

DRAFT

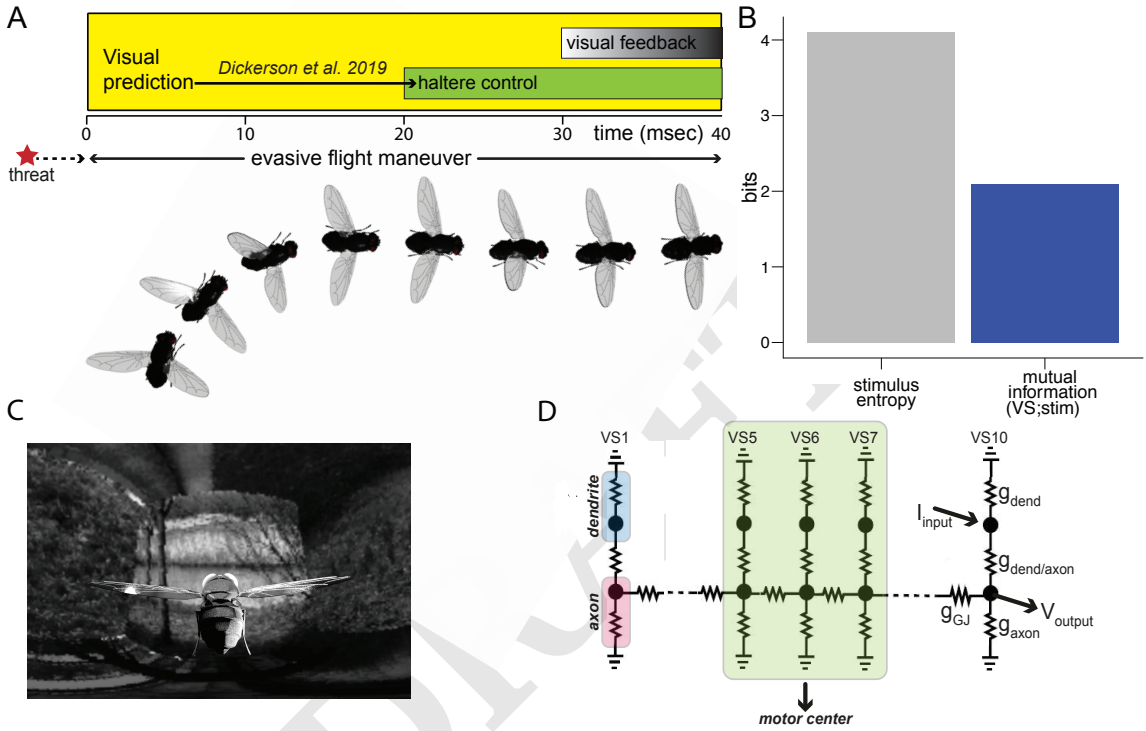

Fig. 1. Predictive information is the dominant information source about visual inputs during evasive flight maneuvers. (A) Upon emergence of a threat (shown as the red star, dashed arrow represents the visual-motor delay of 60ms from the onset of threat to the start of the evasive maneuver. After this sensory-motor delay, the position of the threat is known. Then 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 yellow zone), whereas the haltere feedback is only available after 20ms (shown as the green zone) and the visual feedback is only available after 30ms (shown as the shaded zone). The arrow leading to the haltere system illustrates how visual information might regulate haltere activity (as recently shown in (24)): because of the 30ms 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 the stimulus during the evasive maneuver 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 Fig.S3*A*. Note that the VS output encodes almost half of the entropy of a future stimulus. (C) The mecator 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 Fig.S1. 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 to the VS network (shown as an arrow to D). (D) A biophysically detailed model of the VS network, based on known neural circuitry (28, 29). 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.

$$I_{future:limit} = I(curr_{past}; curr_{future}) \quad [2]$$
$$= I(curr_t; curr_{t+\Delta t})$$

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

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 (29). 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 is $I_{future:limit}$ (Eq.(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 $I_{future:in}$ (Eq.(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 Fig.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. 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 (8). 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

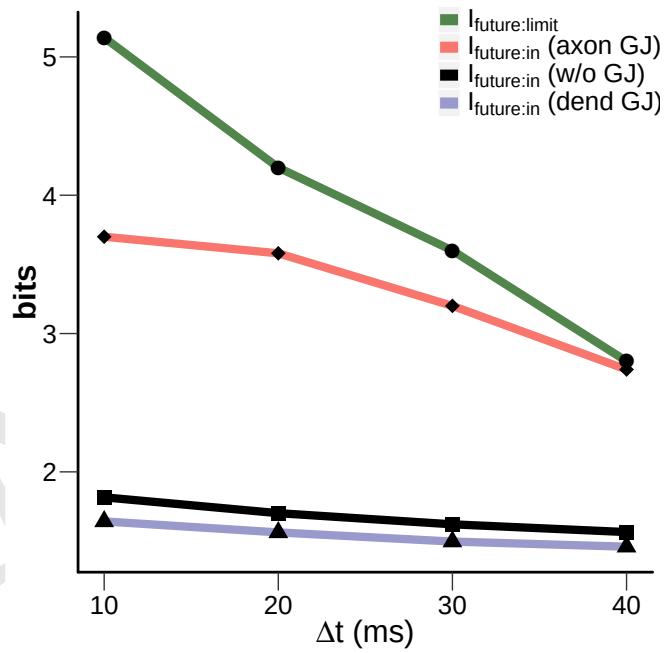

Fig. 2. The capacity of the VS network to encode predictive information varies with the anatomical locations of the gap junction between VS cells. The predictive information about the future input current, $I_{future:in}$ encoded in four different schemes: 1) the past dendritic input current (in green, 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 Materials and Methods.)

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

time in the past). We define this as another mutual information quantity: $I_{past;in}$ (Eq.(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 (42). 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 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 encoding architectures?

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 $\Delta t = 10ms$ (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 Fig.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 Fig.3A and Fig.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 Fig.3A. This suggests that all triplets achieve $89.8 \pm 1.5\%$ efficiency in encoding predictive information about the inputs $I_{future:in}/I^*_{future:in}$.

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 Fig.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 Fig.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 future stimuli, i.e.,$I_{future:stim}$. 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.

Fig.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 (Fig.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 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 (Fig.S4). 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 (43) 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

DRAFT

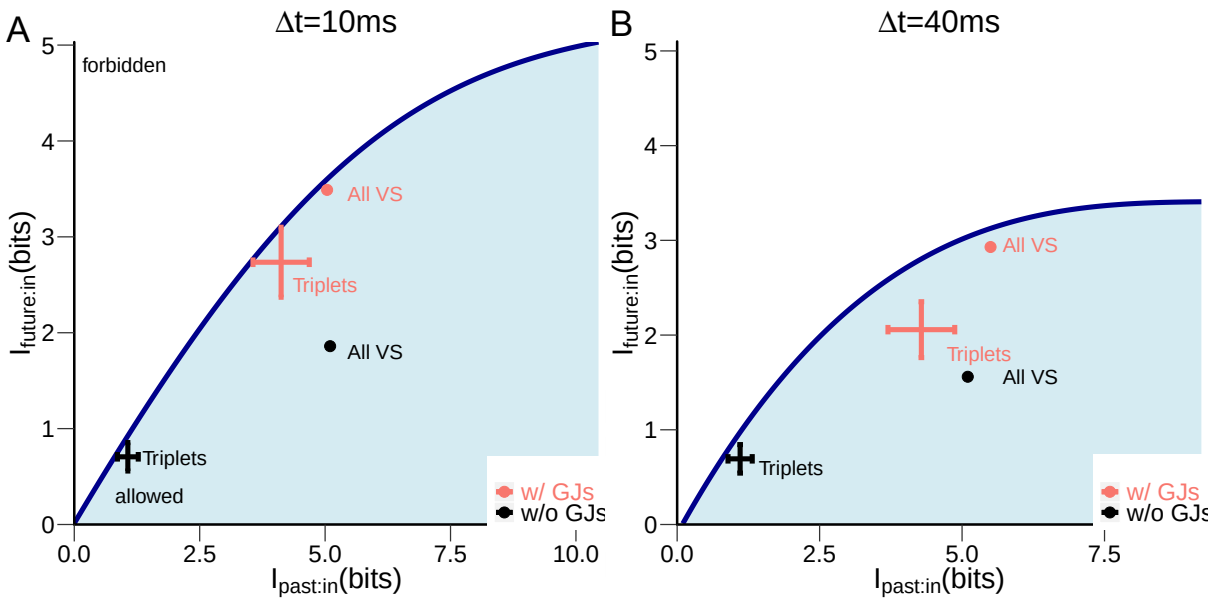

Fig. 3. Near-optimal prediction of the input to the VS network. (A) The encoding of predictive information about the future current input to the VS network is near-optimal 10ms after the evasive maneuver starts ($\Delta t = 10ms$). Such performance is present for using both the entire VS network and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

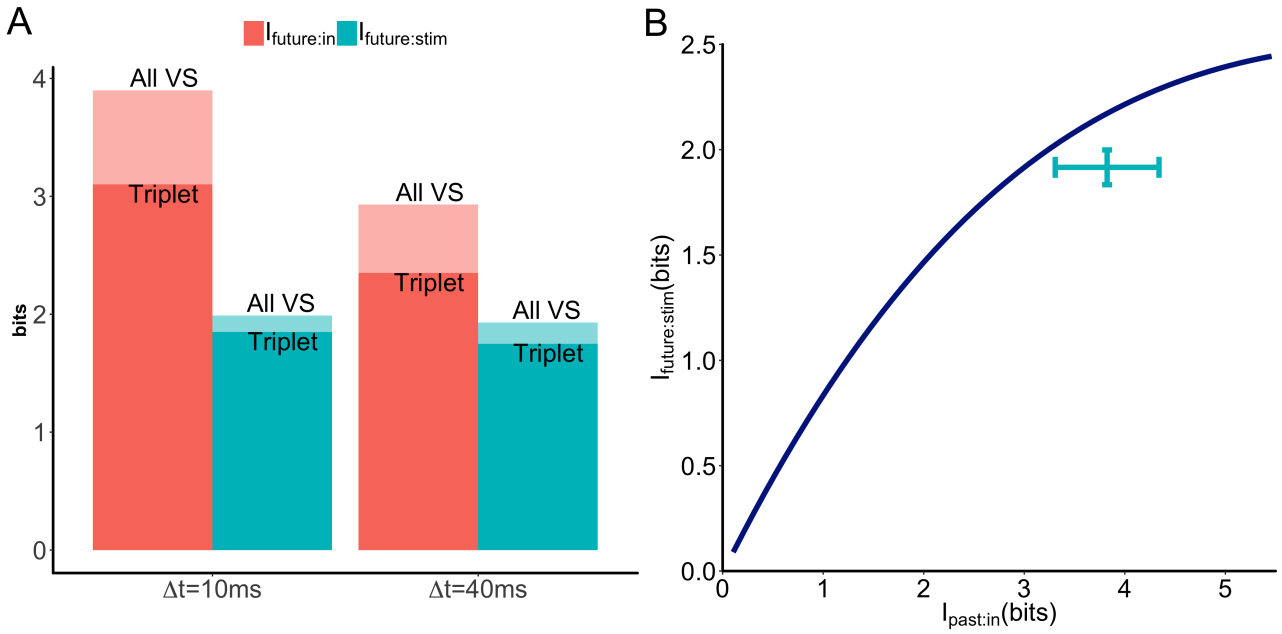

Fig. 4. Encodings based on the axonal voltages of triplets are near-optimal in predicting the future stimulus. (A) Histogram showing that although the VS triplets encode less predictive information compared to a combination of all VS axonal voltages (shown in red), the triplets (we use the VS 5-6-7 triplet 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 both the immediate future stimulus after the onset of evasive maneuvers, $\Delta t = 10ms$, and the future stimulus at the end of evasive manevuers, i.e.,$\Delta t = 40ms$. Because banked turns consist of initial rotations (at $\Delta t = 10, 20$ms) and active counter rotations that almost reverse the respective initial rotation (at $\Delta t = 30, 40$ms), we observe similar predictive information throughout the entire evasive maneuver (as shown in Fig.S2). (B) Similar to Fig.3A: The encoding of predictive information for the future stimulus 10ms after the start of evasive maneuvers ($\Delta t = 10ms$). 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. Analogous to A), similar relationships also hold for other future stimuli, e.g. $\Delta t > 10ms$.

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 the 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 (42). 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 (44–46) to the information bottleneck (VIB) problem. It has been shown that this VIB approximation is closely related to the loss function used to train variational autoencoders (47). 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 (Fig.S5). 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 Methods). 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 (Fig.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 the axonal voltages that correspond to the same future stimulus onto this latent space. We can label these points in the latent space by their future stimuli values, and repeat this procedure for several different stimuli 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 representations of future stimuli, we can understand what is being computed during the evasive maneuver. The Fig.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 Fig.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 (Fig.5C for using the entire VS network as input and Fig.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 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. 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 (Fig.4) that both the VS network and the VS 5-6-7 triplet, as well as their respective VIB approximations (Fig.S8), retain almost the same amount of predictive information about future stimuli.

Interestingly, such discrimination preferentially disentangles nearby stimuli. For example, a pair of stimuli (56° and 67°, shown in Fig.S8*A* and Fig.S8*B*) that are just 10 degrees apart are mapped to distinct, well-separated clusters in the latent space of the VIB. Conversely, another pair (19° and 37°) that are farther apart share some overlap (Fig.5A and Fig.5C). The VS 5-6-7 triplet preserves this fine scale discrimination (Fig.5B and Fig.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 (Fig.S6), 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 Fig.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 Fig.S6). 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 (Fig.S7), e.g. the cluster of 19° in magenta almost covers the entire cluster of 247° in light green (Fig.S7*A*). 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 descending motor neurons (48–51) or the neck motor center(52, 53) combines information from other pathways ((VS network, haltere and prosternal organs (17)) 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. This suggests that behavioral goals sculpt neural encoding even at the earliest stages of sensory processing.

Our work predicts a novel sensory-motor pathway between the visual system and a visually gated motor neuron of wing steering muscles. Evasive maneuvers consist of rapid banked and counter-banked turns. These rapid turns require elevated wing kinematics, including stroke amplitude and frequency (36). Previous work showed that a motor neuron of wing steering muscles, e.g., the second basalare motor neurons (M.b2) in both blowfly and drosophila, are responsible for initiating these elevated wing kinematics (35, 54). However, little is known about these neurons other than that they can be switched on and off by visual input (37, 38). Our work shows that visual prediction can reach these motor neurons at the beginning of evasive maneuvers. This makes visual prediction a suitable input source that may activate these motor neurons at the beginning of rapid banked turns.

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 (17). However, all dipterans use their lobula plate tangential neurons to encode motion information (17, 50). All

DRAFT

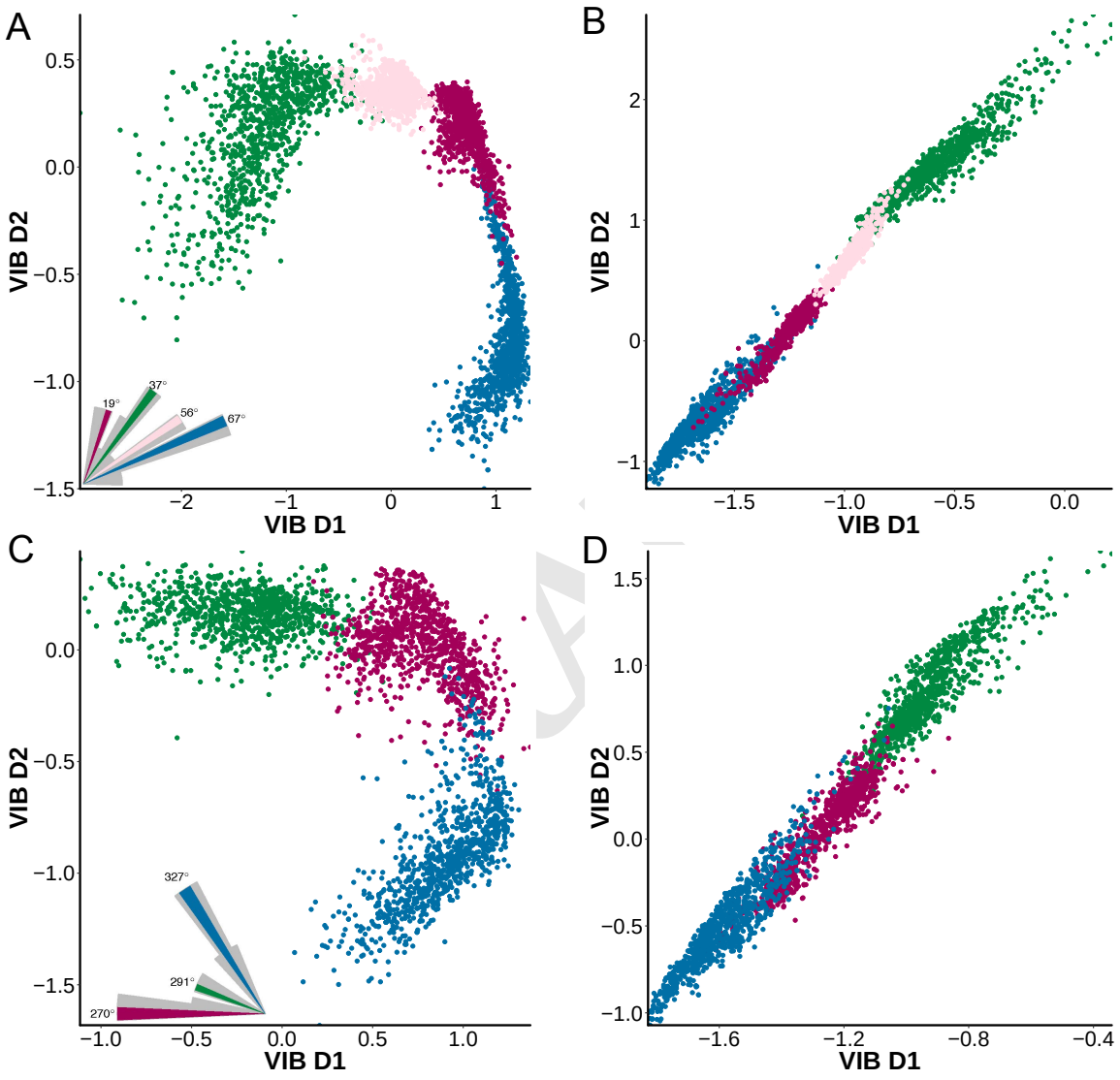

Fig. 5. The predictive information encoded by the VS network supports fine scale discrimination of the future stimulus. (A) The predictive representation of four future stimuli in the same quadrant of roll and pitch, e.g. an up-tilt and a clockwise roll. This representation maps the axonal voltage of 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.

of these neurons have long processing lags, with respect to their behavioral time scales (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) (55). 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 (56, 57). In vertebrate visual systems, the retina also encodes predictive information near-optimally to potentially circumvent sensory processing delays (25, 58). 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 (59), and the gap junctions present in cones and bipolar cells improve the signal-to-noise ratio in their respective outputs (60). Gap junctions can also rapidly regulate chemical synapses and improve sensitivity to correlated signals (61). 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 (62). 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 (25), 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 required to reproduce our simulation and results are specified in this Materials and Methods section.

Self-motion stimuli for evasive flight maneuvers. We obtain self-motion stimuli from a large dataset of evasive flight maneuvers in drosophila published in (7). 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 (7), 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 40ms evasive maneuver is the best drosphila can do. 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. Our simulation uses a biophysically realistic simplified model of the VS network based on a reconstruction introduced in (8). 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 (29, 30); b) the negative conductance between the VS1 and the VS10 to account for the reciprocal innhibition (31).

Compared to the detailed reconstruction, the simplified, biophysically realistic model introduced in (9) 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 (9) so that this simplified model behaves roughly the same as the real VS network when given the same current injection (29, 30).

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 (16)). 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 (9). Based on experimental findings from (28), we vary the magnitude of the GJ conductance between 0 and 1 μS.

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. In every simulation, we first generate the pseudo-3D visual "cube" (Fig.S1*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 (7) (we sample the rotational motion at a $\Delta t = 1ms$ interval, and integrate the VS response at a smaller time step of 0.01ms 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 (16, 41). There are 5,500 local motion

DRAFT

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

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 $t = -10 \sim 0ms$, i.e., 0ms 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., $curr_{past}$. 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: 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 (63)) algorithm to find Z that optimizes Eq.(5):

Mutual information estimation. We use the k-nearest neighbor approach described in (64) 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: with N being the sample size, here $N = 650,000$ or ($N = 650,00$ for bootstrapped stimuli distributions). Given the skewed stimulus distributions shown in Fig.S3, we vary $k = 10, \cdots, 15$ (or $k = 1000, \cdots, 1100$ for bootstrapped stimuli distributions) and use the mean as the estimate in our analysis. We omitted the standard deviations when we plotted Fig.4 and Fig.S2 because of their small magnitudes (< 0.1).

Variational approximation of optimal encoding of the predictive information (VIB). We use the variational approximation introduced in (44). We first rewrite Eq.(5) as Eq.(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)).

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 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 Fig.3 and Fig.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). We thank D. Allan Drummond for providing the blowfly rendering in Figure 1. This work was also supported by NIH grant R01EB026943 (SEP).

1. H Lopez-Schier, Neuroplasticity in the acoustic startle reflex in larval zebrafish. *Curr. opinion Neurobiol.* **54**, 134–139 (2019).
2. G Card, M Dickinson, Visually mediated motor planning in the escape response of Drosophila. *Curr. Biol.* **18**, 1300–7 (2008).
3. T Schilling, A Borst, Local motion detectors are required for the computation of expansion flow-fields. *Biol Open* **4**, 1105–8 (2015).
4. R Hanlon, *Cephalopod behaviour.* (Cambridge University Press), (2018).
5. MH Dickinson, Death valley,drosophila, and the devonian toolkit. *Annu. Rev. Entomol.* **59**, 51–72 (2014).
6. KC Catania, Tentacled snakes turn c-starts to their advantage and predict future prey behavior. *Proc. Natl. Acad. Sci.* **106**, 11183–11187 (2009).
7. F Muijres, M Elzinga, J Melis, M Dickinson, Flies evade looming targets by executing rapid visually directed banked turns. *Science* **344**, 172–177 (2014).
8. H Cuntz, J Haag, F Forstner, I Segev, A Borst, Robust coding of flow-field parameters by axo-axonal gap junctions between fly visual interneurons. *Proc Natl Acad Sci USA* **104**, 10229–10233 (2007).
9. F Weber, H Eichner, H Cuntz, A Borst, Eigenanalysis of a neural network for optic flow processing. *New J. Phys.* **10**, 015–013 (2008).
10. FT Muijres, MJ Elzinga, NA Iwasaki, MH Dickinson, Body saccades of drosophila consist of stereotyped banked turns. *The J. experimental biology* **218**, 864–875 (2015).
11. Schilstra, Hateren, Blowfly flight and optic flow. i. thorax kinematics and flight dynamics. *The J. experimental biology* **202 (Pt 11)**, 1481–1490 (1999).
12. Hateren, Schilstra, Blowfly flight and optic flow. ii. head movements during flight. *The J. experimental biology* **202 (Pt 11)**, 1491–1500 (1999).
13. CN Balint, MH Dickinson, The correlation between wing kinematics and steering muscle activity in the blowfly calliphora vicina. *The J. experimental biology* **204**, 4213–4226 (2001).
14. H Cuntz, et al., Preserving neural function under extreme scaling. *PLoS ONE* **8**, e71540 (2013).
15. F van Breugel, MH Dickinson, The visual control of landing and obstacle avoidance in the fruit fly drosophila melanogaster. *The J. experimental biology* **215**, 1783–1798 (2012).
16. A Borst, F Weber, Neural action fields for optic flow based navigation: a simulation study of the fly lobula plate network. *PLoS One.* **6**, e16303 (2011).
17. K Buschbeck, N Strausfeld, The relevance of neural architecture to visual performance: phylogenetic conservation and variation in Dipteran visual systems. *J Comp Neurol.* **383**, 282–304 (1997).
18. M Land, T Collett, Chasing behaviour of houseflies (fannia canicularis). *J. Comp.* **89**, 331–357 (1974).
19. G.Heide, Neural mechanisms of flight control in Diptera, Technical report (1983).
20. A Fayyazuddin, MH Dickinson, Haltere afferents provide direct, electrotonic input to a steering motor neuron in the blowfly, calliphora. *The J. neuroscience : official journal Soc. for Neurosci.* **16**, 5225–5232 (1996).
21. A Fayyazuddin, MH Dickinson, Convergent mechanosensory input structures the firing phase of a steering motor neuron in the blowfly, calliphora. *J. neurophysiology* **82**, 1916–1926 (1999).
22. AJ Bergou, L Ristroph, J Guckenheimer, I Cohen, ZJ Wang, Fruit flies modulate passive wing pitching to generate in-flight turns. *Phys. review letters* **104**, 148101 (2010).
23. L Ristroph, et al., Discovering the flight autostabilizer of fruit flies by inducing aerial stumbles. *Proc. Natl. Acad. Sci.* **107**, 4820–4824 (2010).
24. BH Dickerson, AM de Souza, A Huda, MH Dickinson, Flies regulate wing motion via active control of a dual-function gyroscope. *Curr. Biol.* **29**, 3517–3524 (2019).
25. S Palmer, O Marre, M Berry, W Bialek, Predictive information in a sensory population. *Proc Natl Acad Sci USA* **112**, 6908–6913 (2015).
26. G Berman, W Bialek, J Shaevitz, Predictability and hierarchy in Drosophila behavior. *Proc Natl Acad Sci USA* **113**, 11943–11948 (2016).
27. J Haag, A Wertz, A Borst, Integration of lobula plate output signals by DNOVS1, an identified premotor descending neuron. *J. Neurosci.* **27**, 1992–2000 (2007).

DRAFT

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

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

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

$$I(X;Y) = \psi(K) - < \psi(n_x+1) + \psi(n_y+1) > + \psi(N), \quad [9]$$

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

28. A Borst, J Haag, The intrinsic electrophysiological characteristics of fly lobula plate tangential cells: I. passive membrane properties. *J. computational neuroscience* **3**, 313–336 (1996).
29. J Haag, A Borst, Neural mechanism underlying complex receptive field properties of motion sensitive interneurons. *Nat Neurosci.* **7**, 628–634 (2004).
30. J Haag, A Borst, Dye-coupling visualizes networks of large-field motion-sensitive neurons in the fly. *J. Comp. Physiol. A* **191**, 445–454 (2005).
31. J Haag, A Borst, Reciprocal inhibitory connections within a neural network for rotational optic-flow processing. *Front. neuroscience* **1**, 111–121 (2007).
32. Y Elyada, J Haag, A Borst, Different receptive fields in axons and dendrites underlie robust coding in motion-sensitive neurons. *Nat Neurosci.* **12**, 327–332 (2009).
33. S Wang, A Borst, N Zaslavsky, N Tishby, I Segev, 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 (2017).
34. A Kennedy, et al., A temporal basis for predicting the sensory consequences of motor commands in an electric fish. *Nat. neuroscience* **17**, 416 (2014).
35. MS Tu, MH Dickinson, The control of wing kinematics by two steering muscles of the blowfly (calliphora vicina). *J. comparative physiology. A, Sensory, neural, behavioral physiology* **178**, 813–830 (1996).
36. G Heide, KG Götz, Optomotor control of course and altitude in drosophila melanogaster is correlated with distinct activities of at least three pairs of flight steering muscles. *The J. experimental biology* **199**, 1711–1726 (1996).
37. W Dickson, A Straw, C Poelma, M Dickinson, An integrative model of insect flight control (invited) in *44th AIAA Aerospace Sciences Meeting and Exhibit.* (American Institute of Aeronautics and Astronautics), (2006).
38. T Lindsay, A Sustar, M Dickinson, The function and organization of the motor system controlling flight maneuvers in flies. *Curr. Biol.* **27**, 345–358 (2017).
39. M Dickinson, F Muijres, the aerodynamics and control of free flight manoeuvres in Drosophila. *Phil. Trans. R. Soc* **371** (2016).
40. J van Hateren, A theory of maximizing sensory information. *Biol Cybern* **68**, 23–29 (1992).
41. J Trousdale, S Carroll, F Gabbiani, K Josi, Near-optimal decoding of transient stimuli from coupled neuronal subpopulations. *J.Neurosci.* **34**, 12206–12222 (2014).
42. N Tishby, FC Pereira, W Bialek, The Information Bottleneck Method (1999).
43. H Cuntz, A Borst, I Segev, Optimization principles of dendritic structure. *Theor Biol Med Model.* **8**, 4–21 (2009).
44. A Alemi, I Fischer, J Dillon, K Murphy, Deep variational information bottleneck int in *Conf. on Learning Representations.* (2017).
45. I Higgins, et al., beta-vae: Learning basic visual concepts with a constrained variational framework in *International Conference on Learning Representations (ICLR).* p. ICLR (2017).
46. M Chalk, O Marre, G Tkacik, Relevant sparse codes with variational information bottleneck (2016).
47. DP Kingma, M Welling, Auto-Encoding Variational Bayes (2013).
48. W Gronenberg, NJ Strausfeld, Descending pathways connecting the male-specific visual system of flies to the neck and flight motor. *J. comparative physiology. A, Sensory, neural, behavioral physiology* **169**, 413–426 (1991).
49. W Gronenberg, NJ Strausfeld, Premotor descending neurons responding selectively to local visual stimuli in flies. *The J. Comp. Neurol.* **316**, 87–103 (1992).
50. JK Douglass, NJ Strausfeld, Anatomical organization of retinotopic motion-sensitive pathways in the optic lobes of flies. *Microsc. Res. Tech.* **62**, 132–150 (2003).
51. N Strausfeld, U Bassemir, The organization of giant horizontal-motion-sensitive neurons and their synaptic relationships in the lateral deutocerebrum of calliphora erythrocephala and musca domestica. *Cell Tissue Res.* **242** (1985).
52. N Strausfeld, H Seyan, Convergence of visual, haltere, and prosternai inputs at neck motor neurons of calliphora erythrocephala. *Cell Tissue Res.* **240**, 601–615 (1985).
53. NJ Strausfeld, HS Seyan, JJ Milde, The neck motor system of the flyCalliphora erythrocephala. *J. Comp. Physiol. A* **160**, 205–224 (1987).
54. F Lehmann, K Gotz, Activation phase ensures kinematic efficacy in flight-steering muscles of drosophila melanogaster. *J. Comp. Physiol. A* **179** (1996).
55. ST Fabian, ME Sumner, TJ Wardill, S Rossoni, PT Gonzalez-Bellido, Interception by two predatory fly species is explained by a proportional navigation feedback controller. *J. The Royal Soc. Interface* **15**, 20180466 (2018).
56. B Connors, Synchrony and so Much More: Diverse Roles for electrical Synapses in Neural Circuits. *Dev Neurobiol.* **77**, 610–624 (2017).
57. E Marder, Electrical synapses: Beyond speed and synchrony to computation. *Curr. Biol.* **8**, R795–R797 (1998).
58. A Sederberg, J MacLean, S Palmer, Learning to make external sensory stimulus predictions using internal correlations in populations of neurons. *Proc Natl Acad Sci USA* **115**, 1105–1110 (2018).
59. S Trenholm, DJ Schwab, V Balasubramanian, GB Awatramani, Lag normalization in an electrically coupled neural network. *Nat. neuroscience* **16**, 154–156 (2013).
60. P Ala-Laurila, M Greschner, E Chichilnisky, F Rieke, Cone photoreceptor contributions to noise and correlation in the retinal output. *Nat Neurosci* **14**, 1309–1316 (2011).
61. J Jacoby, A Nath, Z Jessen, G Schwartz, A self-regulating gap junction network of amacrine cells controls nitric oxide release in the retina. *Neuron* **100**, 1149–1162 (2018).
62. D Heeger, Theory of cortical function. *Proc Natl Acad Sci USA* **114**, 1773–1782 (2017).
63. R Blahut, Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inf. Theor.* **18**, 460–473 (1972).
64. A Kraskov, H Stogbauer, P Grassberger, Estimating mutual information. *Phy. Rev. E* **69**, 066138 (2004).

DRAFT

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

DRAFT

Supplementary Result

References

1. S Wang, A Borst, N Zaslavsky, N Tishby, I Segev, 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 (2017).
2. J van Hateren, A theory of maximizing sensory information. *Biol Cybern* **68**, 23–29 (1992).
3. A Borst, F Weber, Neural action fields for optic flow based navigation: a simulation study of the fly lobula plate network. *PLoS One.* **6**, e16303 (2011).
4. J Trousdale, S Carroll, F Gabbiani, K Josi, Near-optimal decoding of transient stimuli from coupled neuronal subpopulations. *J.Neurosci.* **34**, 12206–12222 (2014).

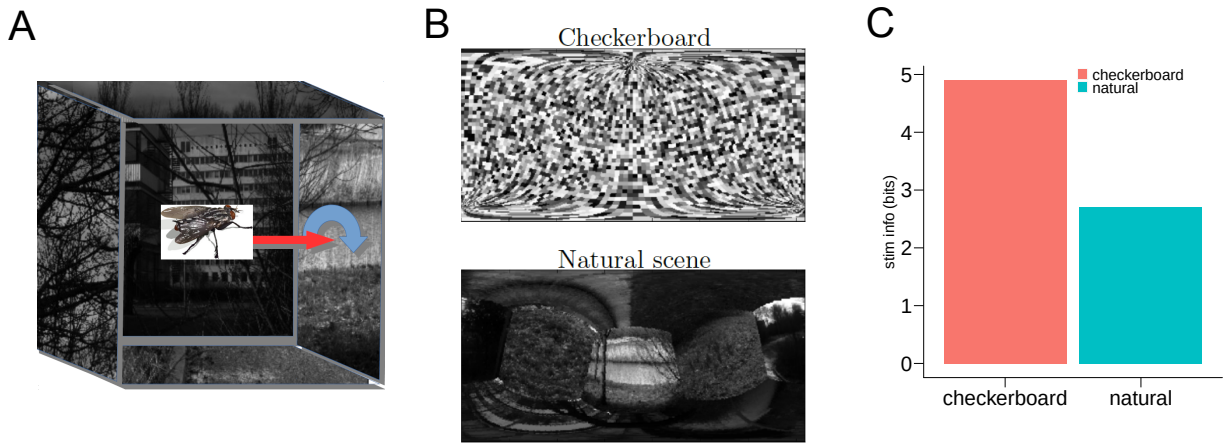

Fig. S1. A) Schematic depiction of the visual stimuli for the simulation, recompiled from (1). Six natural images (five are shown here, with one excluded to reveal the fly's viewing perspective) were randomly selected from the van Hateren dataset (2); each image was patched onto a different face of a 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 (3, 4). 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 $\Sigma_{elevation} = 60^\circ$, 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 $500^\circ/s$ from (1). 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.

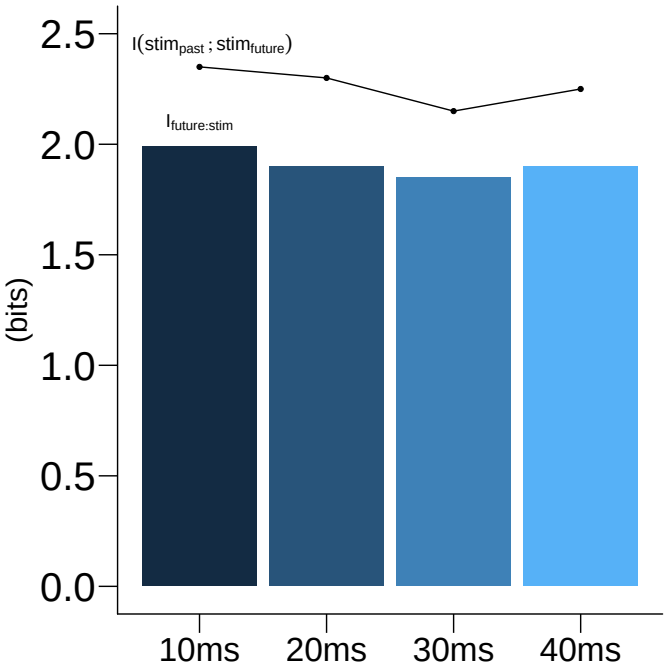

Fig. S2. The VS network output captures most of the predictive information in the fly's escape trajectory. The fly's ego-rotational angles are the visual stimulus, and are inferred by the VS motion sensing system. Black line: Predictive information $I(stim_{past}; stim_{future})$ between the past stimulus ($\Delta t = -10ms$) before the start of the evasive maneuver and future stimuli at time steps $\Delta t = 10, 20, 30, 40$ within the evasive maneuver. Because banked turns consist of initial rotations (at $\Delta t = 10, 20$ms) and active counter rotations that almost reverse the respective initial rotation (at $\Delta t = 30, 40$ms), we observe similar predictive information throughout the entire evasive maneuver. Blue bars: Predictive information about the future stimulus as encoded by the axonal outputs of the entire VS network (defined as $I_{future:stim}$ (defined as Eq.(1) in Results) at each time step within the evasive maneuver.

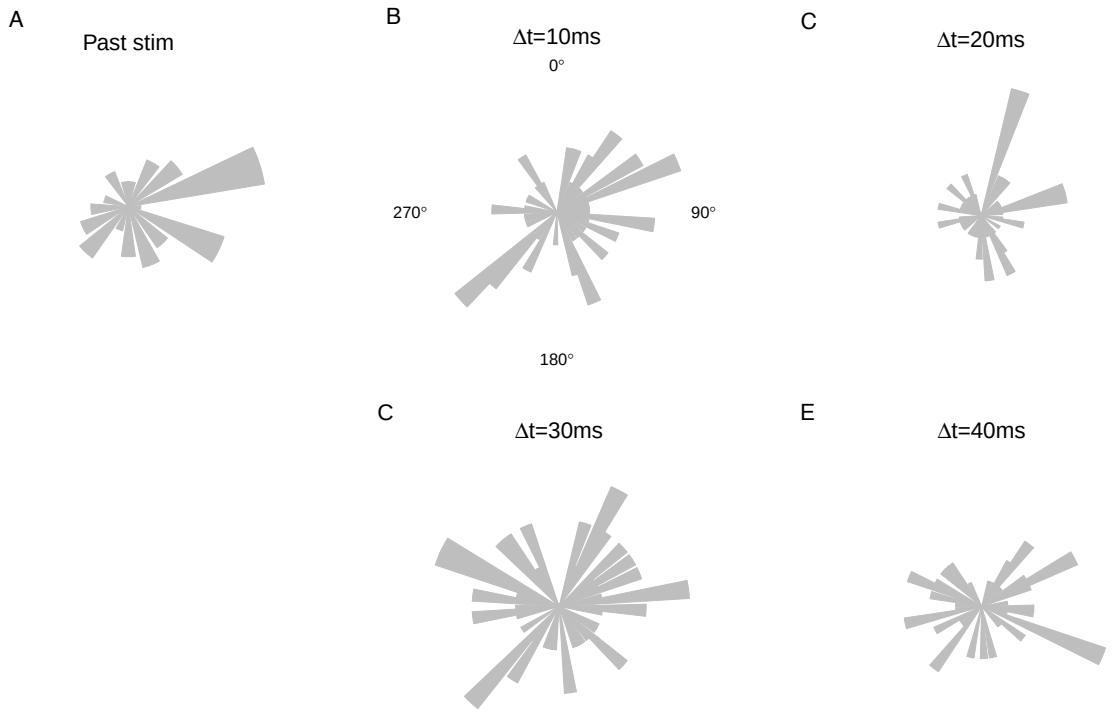

Fig. S3. Stimulus distributions for different time steps during the evasive maneuver. Stimulus distributions for different time steps during the evasive maneuver. Here we focus on the stimuli to which the VS network is sensitive. Because the VS network is only responsive to combinations of roll and pitch 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 before the onset of evasive maneuvers. B) Similar to A, but for the stimulus at 10ms after the initiation of evasive maneuvers. B) Similar to A, but for the stimulus at 20ms within 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 within the evasive maneuver. This motion corresponds to the start of the counter-banked turn. D) Similar to A, but for the stimulus 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 stimulus distributions have comparable entropy $\sim 4 - 4.3$ bits.

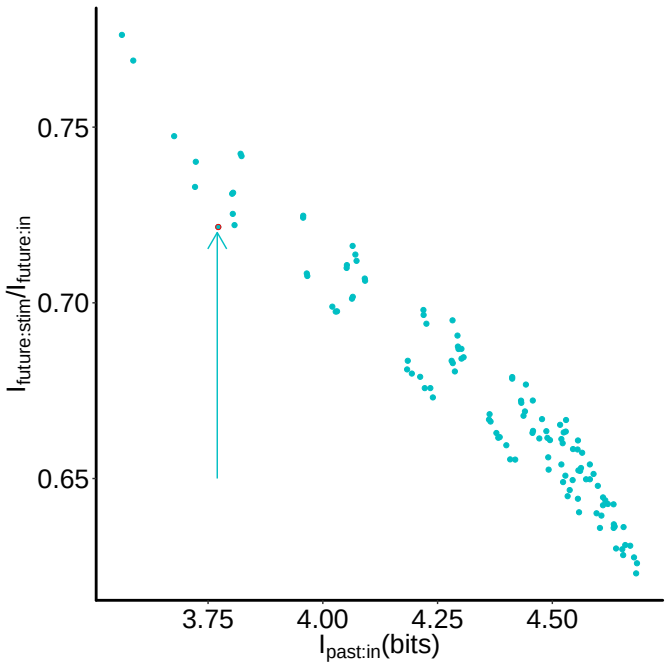

Fig. S4. How much a triplet based encoding retains from the past input vs. how much that information is about the future stimulus (out of the information about their own future input), for all 120 possible triplets. The particular VS 5,6,7 triplet (shown by the red circle and the arrow) that connects 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.

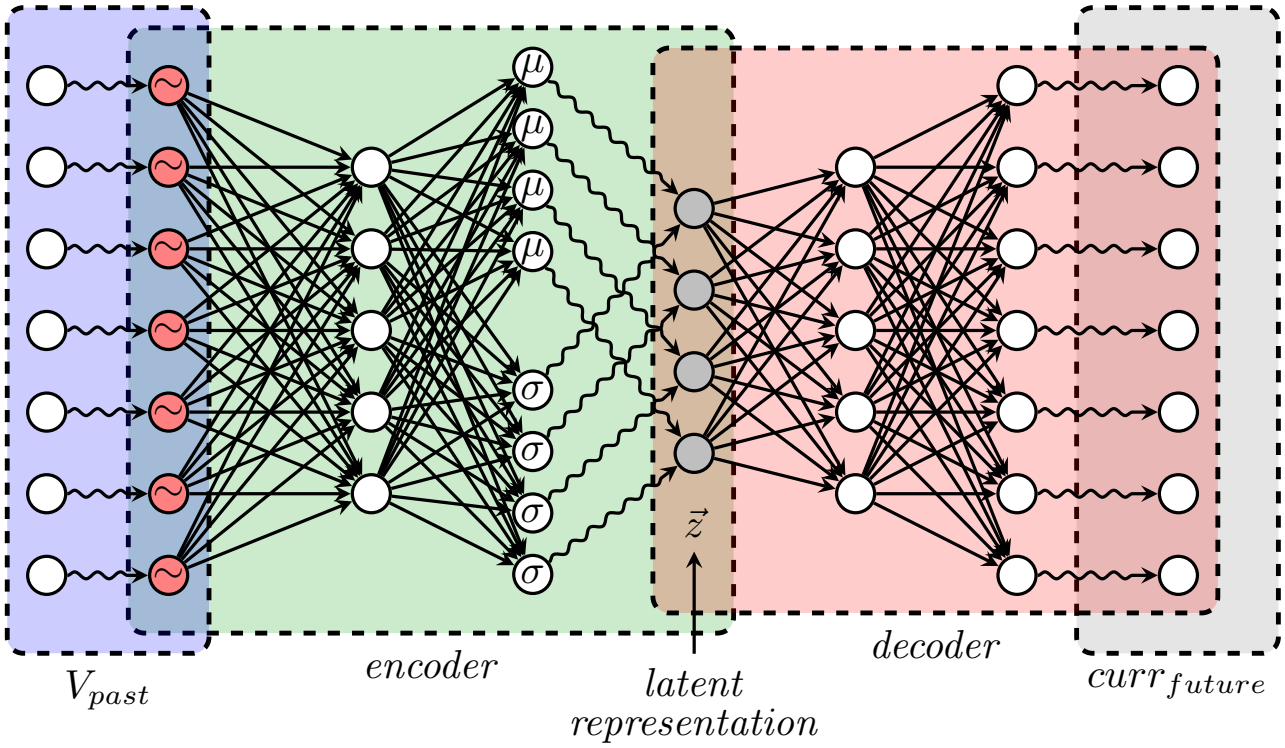

Fig. S5. Network schematic for the variational approximation of the information bottleneck solution (VIB). By constructing a variational approximation, the encoder learned a latent representation $\vec{z}$ from the past VS voltages. W generates samples from $\vec{z}$ and reads them out as the future 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 $\vec{z}$.

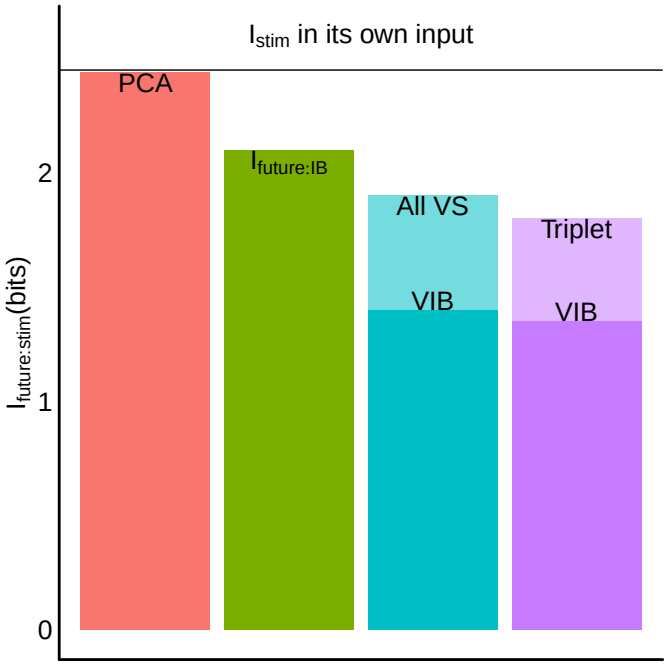

Fig. S6. Predictive information for the future stimulus 10ms after the evasive maneuver starts ($\Delta t = 10ms$). The red bar shows that the PCA projection of the first 2PCs from the input current contains almost all of the stimulus information available at the input current itself. We use this PCA 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.

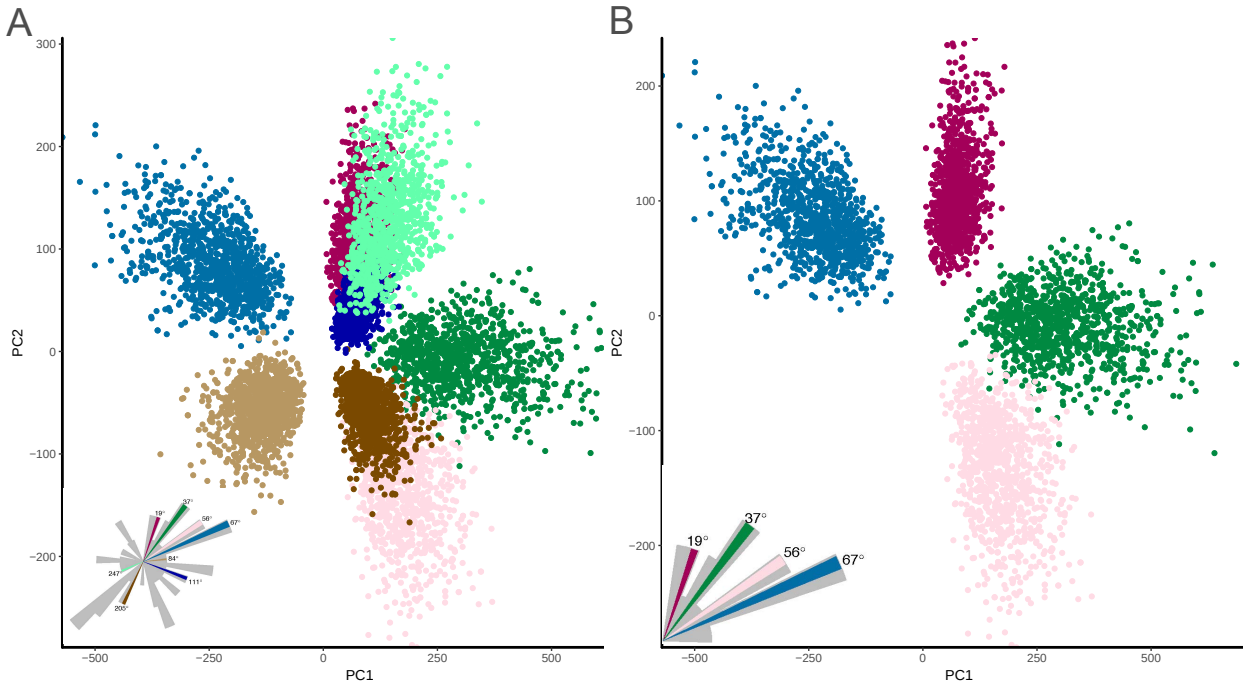

Fig. S7. The input to the VS network only supports local discrimination. A) The representation of 8 randomly selected stimuli within the plane whose dimensions are the first two principal components of the input currents. Note that there are substantial overlaps between clusters: e.g. the light-green 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).

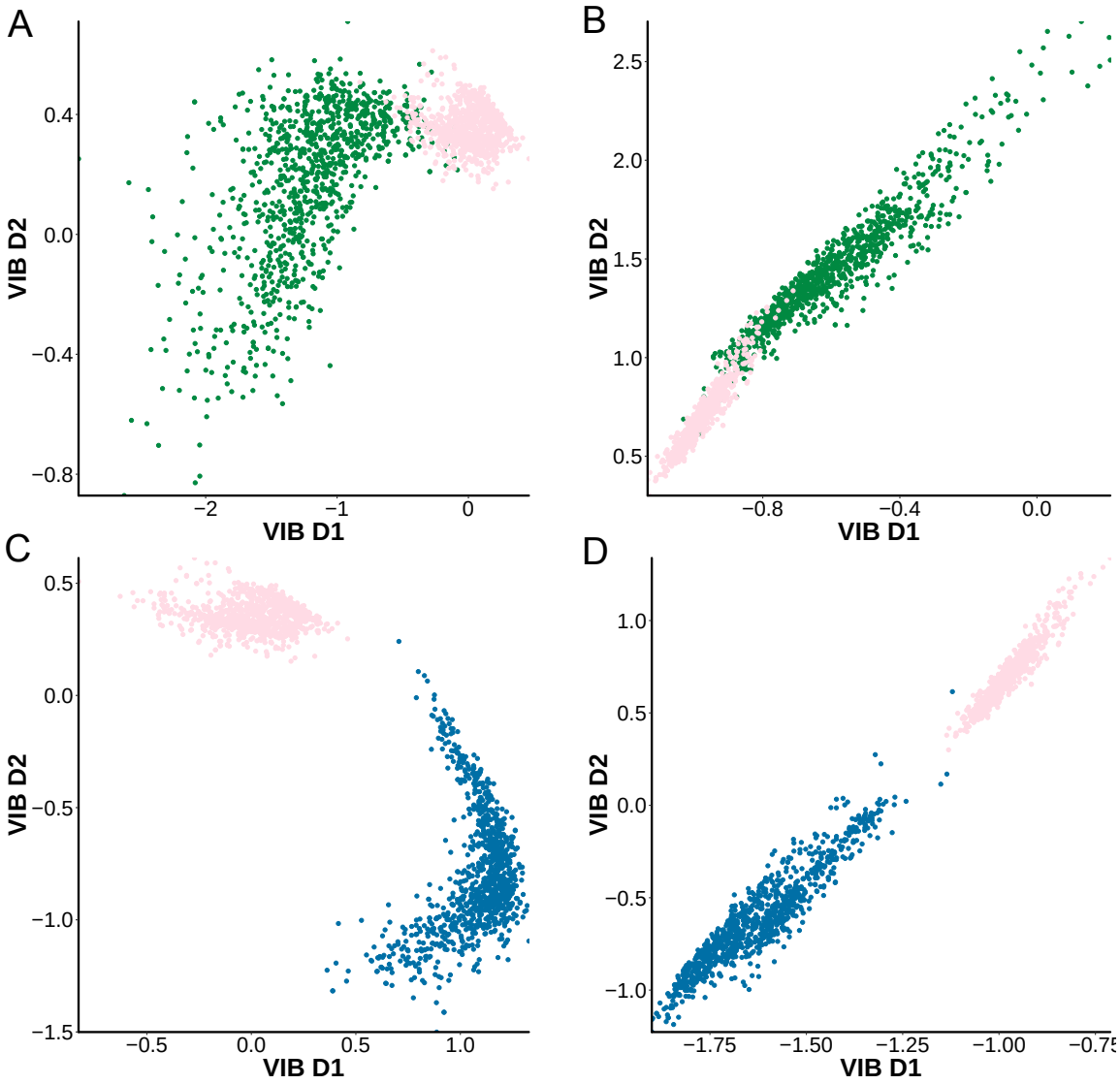

Fig. S8. The predictive information encoded by the VS network preferentially discriminates nearby stimuli. A) The predictive representation of stimuli at 19° and 37° obtained by mapping the respective axonal voltages of the entire VS network to the latent feature space generated by the VIB. B) Similar 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.