EPFL

Thèse n° 9476

Uncovering the neural substrates of communication between the brain and ventral nerve cord in *Drosophila melanogaster*

Présentée le 23 septembre 2022

Faculté des sciences de la vie
Unité du Prof. Ramdya
Programme doctoral en neurosciences

pour l'obtention du grade de Docteur ès Sciences

par

Florian AYMANNS

Acceptée sur proposition du jury

Prof. B. D. McCabe, président du jury
Prof. P. P. Ramdya, directeur de thèse
Prof. J. Simpson, rapporteuse
Dr J. Ache, rapporteur
Prof. A. Ijspeert, rapporteur

École polytechnique fédérale de Lausanne

2022

Abstract

Despite a long history of research in motor control, the exact mechanism of how the brain communicates with the invertebrate ventral nerve cord (VNC) and the vertebrate spinal cord remains largely elusive at a single neuron resolution level. *Drosophila melanogaster* is an ideal model organism to address the role of individual neurons within a population, thanks to its stereotyped nervous system and genetic tools. Descending neurons (DNs) are interneurons in the brain that project to the invertebrate ventral nerve cord or vertebrate spinal cord. They have been shown to elicit specific behaviors upon optogenetic activation in *Drosophila*. These results suggest that behavior is controlled by sparse sets of command neurons, each eliciting a particular behavior. However, it remains to be seen how many DNs are involved in the control of naturalistic behaviors and what other information they might encode. In order to avoid commands leading to physically impossible or destabilizing actions, the brain has to be aware of the current behavior state. Ascending neurons (ANs) are interneurons in the invertebrate ventral nerve cord or vertebrate spinal cord projecting to the brain. They are likely to convey behavior state information to the brain. To which degree of fidelity and in what way ANs encode behavior state remains unclear.

To address these questions, we developed a dissection approach that allows functional imaging of ANs and DNs in the cervical connective. This dissection gives us optical access to the cervical connective and parts of the VNC in behaving adult *Drosophila*. Through the optical access a two-photon fluorescence microscope can be used to image neural activity via calcium indicators.

We began by imaging a new library of sparse split GAL4 driver lines targeting ANs. Our recordings from 247 genetically identifiable ANs revealed neurons encoding walking, resting, turning, eye grooming, foreleg movements, proboscis extension, and CO_2 puffs. Anatomical characterization of these ANs showed that they predominantly project to two brain regions: the anterior ventrolateral protocerebrum (AVLP) and the gnathal ganglion (GNG). Our results suggest that ANs encoding the motion of the animal with respect to the surrounding environment (self-motion) predominantly project to the AVLP, an integrative sensory hub. ANs projecting to the GNG mostly encode discrete actions.

We then proceeded to image populations of DNs. The majority of DNs we recorded encoded walking behavior, with smaller fractions encoding resting and head grooming. Some of these

neurons are likely to drive walking, but we also identified neurons encoding walking speed and turning. This suggests that a core set of DNs drives a behavior whose features can be modulated by additional DNs. Besides encoding behaviors, some DNs encode sensory information, including the presence of odors and deflection of the antennae.

Key words: descending neurons, ascending neurons, ventral nerve cord, *Drosophila melanogaster*, calcium imaging, two-photon fluorescence microscopy, motor system

Zusammenfassung

Trotz einer langen Geschichte der Forschung auf dem Gebiet der Bewegungskontrolle ist der genaue Mechanismus, wie das Gehirn mit dem Strickleiternervensystem von wirbellosen Tieren und dem Rückenmark von Wirbeltieren kommuniziert, auf der Ebene einzelner Neuronen noch weitgehend ungeklärt. Drosophila melanogaster ist ein idealer Modellorganismus, um die Rolle einzelner Neuronen als Teil einer Population zu untersuchen, da Neuronen zwischen Individuen identifiziert werden können. Absteigende Neuronen sind Interneuronen im Gehirn, die in das Strickleiternervensystem projizieren. Die optogenetische Aktivierung einzelner absteigender Neuronen kann spezifische Verhaltensweisen in Drosophila auslösen. Dies deutet darauf hin, dass das Verhalten durch kleine Gruppen von Befehlsneuronen gesteuert wird, die jeweils ein bestimmtes Verhalten auslösen. Es bleibt jedoch abzuwarten, wie viele absteigende Neuronen an der Steuerung natürlicher Verhaltensweisen beteiligt sind und welche anderen Informationen sie kodieren. Um zu vermeiden, dass Befehle zu physisch unmöglichem oder destabilisierendem Verhalten führen, muss sich das Gehirn des aktuellen Verhaltenszustands bewusst sein. Aufsteigende Neuronen sind Interneuronen im Strickleiternervensystem von wirbellosen Tieren oder im Rückenmark von Wirbeltieren die zum Gehirn projizieren. Sie übermitteln vermutlich Informationen über den aktuellen Verhaltenszustands an das Gehirn. Mit welcher Genauigkeit und auf welche Weise aufsteigende Neuronen Verhaltenszustände kodieren, ist noch unklar.
Um diese Fragen zu klären, haben wir eine Sezierungsmethode entwickelt, die die funktionelle Bildgebung von aufsteigenden und abssteigenden Neuronen im Nackenbereich ermöglicht. Die Sezierung gibt uns einen optischen Zugang zum Nackenbereich des Nervensystems und zu Teilen des Strickleiterenervensystems in erwachsenen sich bewegenden Drosophila. Durch den optischen Zugang kann ein Zwei-Photonen Fluoreszenzmikroskop benutzt werden um die neuronale Aktivität mit Hilfe eines Kalziumindikators zu messen.
Wir begannen mit einer neuen Kollektion von split GAL4-Treiberlinien, die jeweils wenige aufsteigende Neuronen als Zielgruppe haben. Unsere Aufnahmen von 247 genetisch identifizierbaren aufsteigenden Neuronen zeigen Neuronen, die Laufen, Ruhen, Drehen, Augenputzen, Bewegungen der Vorderbeine, Ausstreckung des Rüssels und CO_2 Stösse kodieren. Die anatomische Charakterisierung der Neuronen zeigt, dass sie vorwiegend in zwei Hirnregionen projizieren, das anterior ventrolaterale Protocerebrum (AVLP) und das Gnathalganglion (GNG). Unsere Ergebnisse deuten darauf hin, dass aufsteigende Neuronen, die die Bewegung

des Tieres in Bezug auf die Umgebung (Eigenbewegung) kodieren, vornehmlich in das AVLP, ein integratives sensorischen Zentrum, projizieren. Aufsteigende Neuronen, die in das GNG projizieren, kodieren meist diskrete Aktionen.
Anschließend haben wir Populationen von absteigenden Neuronen aufgenommen. Die Mehrheit der absteigenden Neuronen kodiert das Laufverhalten. Kleinere Anteile kodieren das Ruhen und das Putzen des Kopfes. Einige dieser Neuronen sind wahrscheinlich kausal für das Laufen verantwortlich. Andere kodieren die Laufgeschwindigkeit und das Drehen. Dies deutet darauf hin, dass ein Kern von absteigenden Neuronen ein Verhalten steuert, dessen Merkmale durch zusätzliche absteigende Neuronen moduliert werden können. Neben der Kodierung von Verhaltensweisen kodieren einige absteigende Neuronen auch sensorische Informationen, einschließlich des Vorhandenseins von Gerüchen und der Auslenkung der Antennen.

Stichwörter: Absteigende Neuronen, Aufsteigende Neuronen, Strickleiternervensystem, *Drosophila melanogaster*, Kalzium Bildgebung, Zwei-Photonen Fluoreszenzmikroskopie, Bewegungssystem

Résumé

Malgré une longue histoire de recherche sur le contrôle moteur, le mécanisme exact de communication du cerveau avec la moelle épinière des vertébrés et le cordon nerveux ventral des invertébrés reste largement insaisissable au niveau de la résolution d'un seul neurone. *Drosophila melanogaster* est un organisme modèle idéal pour aborder le rôle des neurones individuels au sein d'une population, grâce à son système nerveux stéréotypé. Il a été démontré que les neurones descendants (ND), des interneurones du cerveau qui se projettent dans le cordon nerveux ventral des invertébrés ou dans la moelle épinière des vertébrés, provoquent des comportements spécifiques après l'activation optogénétique chez les Drosophiles. Ces résultats suggèrent que le comportement est contrôlé par des ensembles épars de neurones de commande, chacun provoquant un comportement particulier. Cependant, il reste à savoir combien de ND sont impliqués dans le contrôle des comportements naturels et quelles autres informations ils pourraient encoder. Afin d'éviter que les commandes ne conduisent à des actions physiquement impossibles ou déstabilisantes, le cerveau doit être conscient de l'état actuel du comportement. Les neurones ascendants (NA), des interneurones du cordon nerveux ventral des invertébrés ou de la moelle épinière des vertébrés se projetant vers le cerveau, transmettent probablement ces informations au cerveau. Le degré de fidélité et la manière dont les NA codent l'état du comportement ne sont pas encore bien compris.
Pour répondre à ces questions, nous avons développé une approche de dissection qui permet l'imagerie fonctionnelle des NA et des ND dans la nuque. Cette dissection nous donne un accès optique à la partie cervicale du système nerveux et à certaines parties du cordon nerveux ventral de la Drosophile adulte en comportement, de sorte qu'un microscope à fluorescence à deux photons peut être utilisé pour imager l'activité neuronale via des indicateurs calciques. Nous avons commencé par imager une nouvelle bibliothèque de lignées de pilotes split GAL4 éparse ciblant les NA. Nos enregistrements de 247 NA génétiquement identifiables ont révélé des neurones ascendants codant pour la marche, le repos, la rotation, le nettooyage des yeux et les mouvements des pattes avant. La caractérisation anatomique de ces NA a montré qu'ils se projettent principalement dans deux régions du cerveau, le protocérébron ventrolatéral antérieur (PVLA) et le ganglion gnathal (GGN). Nos résultats suggèrent que les NA codant le mouvement de l'animal par rapport à l'environnement (auto-mouvement) se projettent principalement vers l'PVLA, un centre sensoriel intégratif. Les NA qui se projettent vers le GGN encodent principalement des actions discrètes.

Nous avons ensuite procédé à l'imagerie des populations de ND. La majorité des ND que nous avons enregistrés codaient le comportement de marche, avec des fractions plus petites codant le repos et le toilettage de la tête. Certains de ces neurones sont susceptibles de diriger la marche, mais nous avons également identifié des neurones codant pour la vitesse de marche et la rotation. Cela suggère qu'un ensemble central de ND dirige un comportement dont les caractéristiques peuvent être modulées par des ND supplémentaires. En plus de coder des comportements, certains ND codent des informations sensorielles, notamment la présence d'odeurs et la déviation des antennes.

Mots clefs: neurones descendants, neurones ascendants, cordon nerveux ventral, *Drosophila melanogaster*, imagerie calcique, microscopie à fluorescence à deux photons, système moteur

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motor control

It can be argued, that the primary purpose of brains is to produce intelligent behavior [1]. Therefore, understanding how the brain controls behavior is essential if we strive to understand the brain as a whole. Motor behaviors are of particular importance because they are the way humans and animals can interact with the world around us. The control of motor behaviors can be studied on several different levels, from motor planning and motor learning down to the recruitment of individual muscle fibers. The highest level, motor planning, concerns the spatiotemporal planning of a movement accomplishing an intention. In order to successfully complete the movement, the plan has to be adjusted based on the neural representation of the world around us. For instance, grabbing an object requires knowledge of where the object is located in space. While reaching for the object, the body has to counteract the motion such that it may not lose balance and fall over. More generally, the current behavior state of the body has to be taken into account when a movement is planned. The complex task of motor planning requires communication between several mammalian brain areas, including the premotor cortex, basal ganglia, cerebellum, and the thalamus [2]. Once the movement is planned, the appropriate down stream circuitry has to be activated in order to realize it. At the lowest level, this means muscles have to contract and relax with precise timing and accurate levels of force. This involves the activation of individual motor units [3] and central pattern generators [4] within the vertebrate spinal cord or invertebrate ventral nerve cord (VNC). At the same time, locally implemented reflexes have to be inhibited not to interfere with the planned movement [5]. How the control centers in the brain conduct the VNC circuitry to elicit behaviors and what information about behavioral state they receive from the VNC is the topic of this work.

1.2 *Drosophila melanogaster* as a model organism

1.2.1 *Drosophila melanogaster* has a stereotyped nervous system

One of the major challenges of neuroscience is the fact that no two brains are alike. This is particularly true for mammalian brains whose general structure is conserved across individuals but individual neurons are not identifiable across individuals. The nervous systems of some other organisms, including *Drosophila*, are stereotyped, making it possible to identify the same neuron in multiple animals. Despite this stereotypy, connectomes of individuals show variability in connectivity [6]. One of the factors that leads to such variability in genetically identical individuals, is the temperature during development [7]. Beyond variability in connectivity, there is also variability in behavioral traits in individuals with identical genomes, e.g. in the form of preferring one turning direction over the other in a Y-maze [8]. The ability to study inter-individual variability while being able to identify individual neurons makes *Drosophila* an excellent model organism for neuroscience. Its more complex behavior and the numerous genetic tools available further set *Drosophila* apart from other model organisms with stereotyped nervous systems, such as *Caenorhabditis elegans*. The next section discusses these genetic tools.

1.2.2 Genetic tools allow targeting of precise subsets of neurons

There is a long tradition of *Drosophila* as a model organism. This is strongly related to the genetic tools that are available. One of the first genetic tools discovered and perhaps the most important tool, are balancer chromosomes. These chromosomes have large multiply inverted segments of genome which prevent homologous recombination and are recessive lethal [9]. These two properties and clear phenotypes associated with the balancer chromosomes allow tracking and preserving transgenes with no phenotype across generations. This ability is paramount to the success of *Drosophila* as a model organism [10]. Much more recently developed, but equally ubiquitous in *Drosophila* research, are binary expression systems, the topic of the next section.

Binary expression systems

One of the main principles that makes genetic tools in *Drosophila* so versatile are binary expression systems. As the name suggests, binary systems consist of two components. The first component targets a specific subset of cell, such as all neurons, all sensory neurons, or even a single pair of neurons. The second component is used to express a particular transgene. Using combinations of the two components, the gene expression landscape of a targeted subset of cells can be modified in a specific way. The components are usually referred to as the driver line and the reporter line. The transgene of the driver line is placed under the control of a native enhancer restricting its expression to a subset of cells and a time where and when the enhancer is active.

The GAL4/UAS system [11, 12] is the most widely used binary system in *Drosophila*. Its two components are the the GAL4 gene and the upstream activation sequence (UAS). The GAL4 gene encodes a transcription activator from yeast which can bind to UAS, a regulatory sequence increasing expression of a neighboring gene upon binding of GAL4. On its own, GAL4 is invisible and has little to no effect on the cell. Its presence becomes visible and/or affects the cell through the use of reporter lines. A wide range of reporter lines is available to drive expression of different proteins. Some of the most relevant reporter lines for this work include genetically encoded calcium indicators and fluorescent proteins for labeling purposes. Combined with an ever growing number of GAL4 driver lines targeting subsets of neurons, this system makes *Drosophila* an ideal model organism for neuroscience research.

Due to the dependence on the spatiotemporal specificity of enhancers, it is challenging to identify GAL4 driver lines that target extremely sparse subsets of neurons. To overcome this hurdle, intersectional approaches have been developed. The expression of GAL80 inhibits activation of UAS by GAL4 [13]. If GAL4 and GAL80 are expressed in overlapping subset of cells, the reporter downstream of UAS will only be expressed in cells that exclusively express GAL4. This way, existing driver lines can be made more sparse by 'removing' neurons. Alternatively, driver lines can be made more sparse through the intersection of two lines. This intersection can be achieved with split-GAL4 lines [14]. The GAL4 gene consists of parts, the DNA-binding domain and the activation domain, that can be expressed separately. If the two parts are expressed in the same cell, they form a complete GAL4 protein which can trigger expression of the reporter.

1.3 Calcium imaging — A surrogate for neural activity

Calcium is an important second messenger in biology. It is involved in multiple mechanisms, including neurotransmitter release and muscle contraction. In order to study live calcium (Ca^{2+}) concentrations, calcium indicators were developed. Calcium indicators are molecules that react with Ca^{2+} ions resulting in a detectable change to the molecule. In its earliest form this was the production of purple showers of a calcium salt formed upon the injection of a high concentration solution of the dye alizarin [15]. Nowadays, the reaction is usually associated with an increase of fluorescence upon the binding of calcium. In the popular calcium indicator GCaMP, this happens through a conformational change [16]. The first calcium indicators were chemically synthesized and had to be added to the cells making it difficult to localize the measurement to a precise subset of cells. For a comprehensive review of these early calcium indicators, see [17]. Later, genetically encoded calcium indicators became available and made it possible to measure changes in calcium in specific subsets of cells, leaving the remaining cells unaffected [18]. Since then, new genetically encoded calcium indicators, with ever increasing signal-to-noise ratios and ever faster temporal dynamics, are released on a regular basis [16, 19, 20]. In this work GCaMP6 [19], a calcium indicator based on a combination of green fluorescent protein and calmodulin, is used to measure neural activity. Using calcium imaging to measure neural activity has several advantages over more direct measures of neural

activity. In flies, it is non-invasive with respect to the neural tissue, although gaining optical access to the nervous system is usually invasive. Furthermore, calcium imaging allow us to image multiple neurons at once, while also capturing their relative position and size. Relative position is especially useful when it comes to stereotyped nervous systems such as the one of *Drosophila*, where it can be used to identify and compare neural activity across animals. Perhaps most importantly, calcium imaging can be performed in vivo during behavior.

Next, I will briefly describe the roll of calcium in neurons focusing on why intracellular Ca^{2+} concentration is an indicator for neural activity. Then, I will discuss the main shortcomings of calcium indicators in relation to neural activity, namely buffering and temporal resolution.

1.3.1 The role of calcium in neurons

Neurons conduct signals through changes in potential difference between the intra- and extracellular medium. These changes occur through influx and efflux of ions via ion channels, pumps and exchangers. One of the these ions is Ca^{2+}. The most significant roll of Ca^{2+} in neurons is the conversion of electrical signals into output. At the synapse, it functions as an intracellular messenger triggering neurotransmitter release. When an action potential reaches the active zone of a synapse, voltage gated Ca^{2+} channels open in response to the depolarization [21]. The in-flowing calcium ions then bind to synaptotagmin, which results in vesicle fusion via the protein complex SNARE releasing neurotransmitter into the synaptic cleft [22]. Besides neurotransmitter release, calcium is involved in the modulation of ion channels, synaptic plasticity, and activation of transcription [23].

It should be mentioned that the electrochemical gradient for Ca^{2+} is particularly large allowing for fast influx through ion channels and a rapid multiple-fold change in terms of the number of Ca^{2+} ions in the cytoplasm [2]. These large changes make it a great candidate for monitoring neural activity via a calcium indicators.

1.3.2 Buffering and temporal resolution

Not all of the calcium that enters the neuron upon opening of ion channels is expelled again once the neuron repolarizes. Some of the calcium is stored inside the endoplasmic reticulum or bound by calcium binding proteins and can be released again at a later point [24]. This process is called buffering. Just as the naturally occurring calcium buffers, calcium indicators delay the efflux of calcium. Free Ca^{2+} is bound by the calcium indicator on the order of tens to hundreds of milliseconds after the firing of an action potential [19]. The release, however, takes on the order of hundreds of milliseconds to seconds [19]. The resulting dynamics in the fluorescence can be modeled by a first order auto-regressive processes [25] or by a convolution kernel of two competing exponential functions [26], one modeling the rise, and the other modeling the decay. It is important to take these dynamic into account when analyzing events whose duration is of the same order of magnitude as the fluorescence decay.

Deconvolution

The uptake and release dynamics of the calcium indicator are effectively a low-pass filter that can be mathematically described as the temporal convolution of the calcium concentration with the kernel described above. This kernel is sometimes called the calcium response [26] following the fMRI terminology of hemodynamic response. Although, convolution can mathematically be reversed by division in the Fourier space, measurement noise and inaccurate convolution kernels make this process difficult in practice. To combat the effects of noise, deconvolution algorithms make assumptions about the underlying signal, for instance, that sparse sets of action potentials generated the observed signal [25, 27]. In order to be able to infer individual action potentials, these algorithms require relatively high frequency neural data. Alternatively, the number of transitions between different activity levels can be regularized [26], relaxing the assumption of a spiking neuron and even allowing the deconvolution of data without single neuron resolution.

1.4 Choice of the experimental setup

1.4.1 Spherical treadmill

Naturalistic behavior is important when studying the nervous system. It allows us to observe the nervous system perform the tasks it evolved for [28]. Experimental setups that allow such naturalistic behavior while recording the activity of neurons include spherical treadmills [29], miniaturized microscopes [30], and fiber photometry [31]. *Drosophila* are about 3 mm in size [32] with an even small nervous system. With the exception of the spherical treadmill, the approaches above are prohibitively large and heavy or do not afford single neuron resolution. The spherical treadmill allows the fly to engage in all of the behaviors it most commonly performs, except for flight. Given that *Drosophila* spend very little time of their life flying [33] a spherical treadmill is a reasonable choice to study motor control in flies.

1.4.2 Two-photon microscopy

Multiphoton microscopy is a type of fluorescence microscopy which uses a much lager excitation wavelength than emission wavelength. One of the main advantages of it is that longer wavelengths are scattered less and can therefore penetrate deeper into the tissue [34]. In conventional fluorescence the molecule is lifted to an excited state through excitation by a photon whose energy matches the energy difference between the states. The molecule then releases some of its energy in non-radiative transitions and eventually drops back down to the ground state, releasing the energy in the form of light. If two photons reach the fluorophore with in a very short time, their energies can sum up to lift the fluorophore to an excited state. In order to excite the same fluorophore with two photons, their energy has to be half the energy of a single photon, i.e. the wavelength doubles. Multiphoton excitation requires a very high flux of photons, because in low flux conditions, photons are unlikely to excite the same molecule

within the required time interval. In practice this means that multiphoton excitation only takes place around the focal plane. This makes the use of pin holes obsolete, and fluorophores outside the focal plane are not excited, i.e. do not bleach [34]. Recent advances have shown that three-photon microscopy in principle allows imaging neural activity through the intact cuticle of the fly's head [35], making calcium imaging less invasive.

1.4.3 Behavior tracking

When studying motor circuits and motor control, it is essential to accurately measure the subjects behavior. Traditionally, summary variables, such as speed and turning, were recorded using optical flow sensors measuring the rotations of a spherical treadmill. For insects the treadmills are generally very small, making calibration of optical flow sensors difficult. Moore et al. [36] developed an approach that overcomes this issue by using a single camera to build a global map of the patterns on the treadmill. Not only does this approach not require calibration, it also reduces the effects of accumulating errors when estimating the path of the animal.
Recent advances in computer vision [37–41] have furthermore enabled us to study and quantify behavior in greater detail by tracking individual body parts. The position information can then be used to classify the behavior [42] and to analyze the control of individual joints by the nervous system. In *Drosophila*, we can even go one step further and replay the joint kinematic in a bio-realistic simulation [32] to estimate otherwise unknown parameters, such as joint torques and ground reaction forces.

1.5 The information bottleneck between the parts of the nervous systems

The majority of animals have central nervous systems that can be split into two parts: the brain and the invertebrate ventral nerve cord or the vertebrate spinal cord. Even the nervous system of the nematode *Caenorhabditis elegans* can be partitioned in to the head ganglia (brain) and the ventral and dorsal nerve cords [43] despite the fact that its nervous system only consists of 302 neurons [44]. The ventral nerve cord (VNC) or spinal cord is the part of the central nervous system interacting with the body. It drives motor actions and receives sensory information from the peripheral nervous system. The importance and sophistication of the circuitry in the VNC and the spinal cord is strikingly demonstrated by decapitation experiments. Decapitated flies can still perform complex motor actions such a walking and grooming [45, 46] and they can learn to avoid electric shocks [47]. In mammals, experiments with decerebrated cats [48] have a long history showing the importance of the circuitry besides the motor cortex. To perform conscious voluntary movements or to adapt movements to the specifics of the situation, the brain has to send signals to the motor circuitry. This happens through so called descending neurons. To make an informed decision and avoid unstable or physically impossible actions, the brain needs to know about the current behavioral state.

This information is relayed to the brain via ascending neurons.

1.5.1 Ascending neurons

Ascending neurons (ANs) are neurons that have their cell body in the invertebrate VNC or the vertebrate spinal cord, and project axons to the brain. They relay information about the body to the brain. How important this feedback is for motor control is strikingly demonstrated by the case of Ian Waterman, who lost most his proprioception and touch sense from the neck down [49, 50]. This subsection discusses the role of ANs in insects.

In *Drosophila*, sensory information about body state encoded by ANs can arise from six different sensors [51]: (i) gustatory sensilla on the legs and wings used for taste, (ii) external sensilla sensing deflection of bristles, (iii) stretch receptors sensing stretch and tension between leg segments, (iv) chordotonal organs sensing tension and vibrations, (v) campaniform sensilla sensing deformations of the cuticle, and (vi) multidendritic neurons encoding pain and temperature. The sensory neurons originating from theses sensors can either ascend directly to the brain as primary sensory ascending neurons or form connections in the VNC. Besides primary sensory ascending neurons, there are also secondary ascending interneurons [51], which likely integrate information from multiple sensory neurons [52]. However, what information they encode exactly and how they influence behavior is only known for few specific examples. Some of these examples point to a role of ANs in action selection. A first example that supports this notion is a pair of ascending neurons involved in backward walking [53]. Although the activation of a single pair of DNs is sufficient to elicit backward walking, the behavior is only sustained when the pair of ANs is not silenced [53]. When activated, the same pair of ANs interferes with forward walking, suggesting that it might inhibit forward walking in favor of backward walking. A more direct involvement of an individual pair of ANs in action selection was shown by Mann et al. [54]. This specific pair of ANs inhibits feeding when activated, and promotes feeding and inhibits locomotion when inhibited. Besides roles in action selection, ANs have also been shown to modulate visual neurons, presumably contributing to a mechanism for course correction during goal directed walking [55]. This sort of influence of motor regions on sensory regions is called corollary discharge. The influence on sensory circuits can range from reflex inhibition and sensory filtering to sensory analysis/stability and sensorimotor learning/planning [56]. Naturally, some of the neurons implementing corollary discharge might be ascending neurons connecting the motor circuits in the VNC to sensory regions in the brain. An example of an ascending neuron involved in sensory filtration was found in crickets [57]. This neuron allows crickets to retain auditory sensitivity during singing through in phase inhibition of the central auditory pathway [58].

1.5.2 Descending neurons

Descending neurons (DNs) are neurons that have their cell bodies in the brain and project axons to the VNC or spinal cord respectively. It is generally assumed that they receive inputs

through dendrites in the brain and output in the VNC. However, in some instances they also have presynaptic terminals in the brain [59]. This subsection discusses the general organization and role of descending neurons in insects and mammals.

Command neurons

The concept of commmand neurons was first described by Kupfermann et al. [60]. They defined a command neuron as a neuron that is necessary and sufficient for a given behavior. The interpretation of necessary and sufficient as equivalence in terms of formal logic results in practical challenges [61]. Instead Kupfermann et al. [60] propose a set of three experiments to identify a neuron as a command neurons: (1) an experiment showing that a neuron is active during a stimulus elicited behavior, (2) an experiment showing that removing the neuron breaks the link between stimulus and elicited behavior, and (3) an experiment showing that artificial activation of the neuron elicits the behavior. In practice, neurons are often called command-like when their activation elicits the behavior in question and silencing results in a reduction of that behavior. The concept of command or command-like neurons is particularly useful for DNs as it describes a causal link between DN activity and behavior.

Insects

Descending neurons have been studied in numerous insects, including locusts [62, 63], crickets [64–67], moths [68, 69], stick insects [70], cockroaches [71–73], and flies [53, 59, 74–80]. Anatomical characterization of DNs has revealed that the general organization seems consistent across species [79]. Based on the location of a DN's cell body, they can be broadly classified into one of two categories, (i) supraesophageal ganglia descending neurons, and (ii) subesophageal ganglia descending neurons [80]. In some insects, the subesophageal ganglia are separated from the supraesophageal ganglia, which are sometimes referred to as the proper brain [80]. In *Drosophila*, they are not separated, and the subesophageal ganglia are usually referred to as the subesophageal zone (SEZ). Across species, subesophageal DNs tend to appear as clusters of neurons with similar morphologies and neurites running in parallel [59, 80]. These clusters are referred to as parallel DNs [80] or population DNs [59]. The remaining DNs form bilateral pairs with unique projection patterns [59]. The exact number of descending neurons is likely to vary across species due to different numbers of neurons in the cervical connective [81]. For *Drosophila*, the most recent estimate of the number of DNs is ~ 700 [59]. Previous estimates have suggested a number of DNs closer to ~ 1100 [79]. The relative number of supraesophageal DNs in *Drosophila* ranges from ~ 50% [79] to ~ 75% [59]. Beyond the broad classification into sub- and supraesophageal, DN somas appear in clusters in the supraesophageal ganglia [59, 64]. Since the cell bodies of neurons sit on the outside of the neuropil in insect brains, the cell body's location only provides limited information on the neurons function. A possibly better way to characterize DNs is to look at the innervation patterns of dendrites and axons. Namiki et al. [59] also describe these innervation patterns in the brain and the VNC. Notably, they find a large number of presynaptic neurites in the

ganthal ganglia (GNG), a part of the SEZ, putatively serving as an action selection hub [54, 82, 83]. In terms of dendritic innervation, the GNG is only surpassed by the inferior posterior slope (IPS) and the superior posterior slope (SPS) [59]. With connectomes of the brain [84] and VNC [85] becoming available, it will be possible to identify the pre- and postsynaptic partners of descending neurons to further refine the anatomical characterization. Once all synaptic locations and partners have been identified, it still remains to be seen what neurotransmitters are used by each of the DNs. Previous results suggest that most DNs are cholinergic or GABAergic at about 40% of the total population each [79]. This puts the number of inhibitory DNs at a much higher ratio than in mammals, whose DNs seem to be almost exclusively excitatory [86].

Besides anatomical characterization, the function of individual descending neurons has been studied in many organisms, including locusts [62, 63], crickets [65–67], stick insects [70], and *Drosophila* [53, 74–76, 78, 87–90]. Many of the neurons studied implement direct sensory motor transformations. One of the oldest examples implementing such a transformation is the locust descending motion detector [91], a neuron that receives input from the optical lobes, representing optical flow and drives escape responses [92]. A more recent example is a study in *Drosophila* showing that the giant fibers are activated by optical lobe neurons encoding the angular size and speed of looming stimuli [76]. The two factors determine the type of escape response based on spike timing in the giant fibers [90]. These visuomotor transformations often involve few synapses in the brain, allowing a fast response to the stimulus, but leave little room for context dependent gating that is necessary to avoid inappropriate behavioral responses. In recent work, Ache et al. [77] have identified two such gating mechanisms for the responses to looming stimuli based on whether the fly is currently flying or not. Besides the eyes, the antennae are another set of important sensory organs located on the insect's head. Sensory motor transformation of touches to the antennae have been shown to trigger grooming in flies [93] and turning in stick insects [94].

It is unlikely that all DNs implement such direct sensory motor transformation. For less discrete behaviors such as walking or flying the sensory input would have to be maintained over a prolonged period of time to explain long bouts of goal-directed behavior. DNs encoding walking [63, 65, 75] and flight [65, 87, 95] have been identified in several insects. In some cases, they encode specific behavioral features such as speed [96] and turning [75, 87, 89, 97]. Notably, they also show weak responses to odor presentation, suggesting that they might have connections to sensory neurons [65, 89, 97]. This combined with the fact that multiple neurons seem to encode the same behavior has lead to speculations that they might drive the same behavior in different sensory contexts [97]. This hypothesis is corroborated by the fact that optogenetic activation of DNs via sparse split GAL4 driver lines leads to similar behaviors for different neurons [74]. The results shows that many descending neurons are sufficient to elicit one behavior. It is not clear, however, whether they do so through direct activation of motor circuitry or if their activation leads to the recruitment of additional DNs either via axons in the brain or connections to ascending neurons. For them to be classified as command neurons, their activity also has to be necessary for the behavior [60]. In the formal sense of

the logic term 'necessary', this would require stimulating every neuron to show that no other neuron can elicit the behavior after the the neuron in question is silenced or killed [61]. In practice, the neuron is usually silenced using a neurotoxin [53] or over expression of inwardly rectifying potassium channels [77], and the effects on behavior are quantified.

Alternatively to sensory context, DNs might encode similar behaviors with differences that can go unnoticed by the human observer. An example for similar behaviors driven by different DNs are the anterior grooming neurons identified by Guo et al. [78], with individual neurons eliciting front leg rubbing, antennal grooming, or both. A similar case seems to exist for walking, with the DNp09 neurons driving walking with ipsilateral turning and a set of brain inter neurons driving fast walking without turning through a different descending pathway that is yet to be identified [75]. This makes it difficult to determine how many neurons are involved during natural behaviors. They could either be uniquely active to drive one specific sub-type of the behavior that is appropriate for the sensory input or be simultaneously active swaying the behavior more or less toward either sub-type.

As already mentioned above, individual DNs are sometimes capable of driving more than one behavior [78, 98]. In the case of courtship the progression in the cumulative sequence of behaviors is controlled by spike counts [98]. This suggests that not only behavioral features, such as speed and turning, but also entire behaviors might be driven by activity levels rather than binary active-versus-not-active thresholds.

Mammals

The anatomical organization of mammalian DNs is relatively well understood. Based on their origin and location in the spinal cord, they can be classified into different tracts [99]. Most of the evidence for the functional role of individual tracts has come from lesioning studies [99], but how individual DNs or sub-groups of DNs drive behavior remained largely unclear until recently. DNs can be coarsely classified into neurons originating from cortex or the brain stem [100].

Studies of subset of brain stem DNs have revealed populations of DNs involved in locomotion [101], turning [102], forelimb movements [103], bladder control [104], and resting [105]. These subsets have been termed command lines [106]. Notably, previous work on the function of specific regions within the brain step was unable to elicit such behavior. Only the ability to target subsets based on neurotransmitter content, gene expression, and connectivity revealed sub populations of DNs involved in specific behavior and showed that the neurons involved in different behaviors are often intermingled within the brain stem [107]. Although command lines consist of multiple neurons, the organization seems to resemble the concept of command neurons in insects. This suggests that some of the control principles found in insects might be conserved across species.

The role of corticospinal neurons is less clear as they have many collateral axons projecting to

brain areas along their way to the spinal cord. It has been hypothesized that they broadcast context dependent and cognitive information to the areas involved in motor control [106]. Furthermore, it has been shown that they can have many different effects in the spinal cord [99], including afferent input control [108], spinal reflex control [5], excitation and inhibition of motor neurons [109], autonomic control [110], long-term plasticity [111], synapse development and maintenance [112], and putatively recruitment of motor units [3].

1.6 Aim of this work

The aim of this work is to better understand how the *Drosophila* brain and VNC communicate with each other, on a single neuron level, to generate behavior. We aim to tackle this question from two different directions: (i) the encoding of body state information by ascending interneurons and (ii) the activity of populations of DNs during behavior.

For this undertaking, it will be crucial to record AN and DN activity during behavior. Current approaches only allow such recordings in the brain of behaving flies. Imaging neural activity of ANs in the brain is difficult as their axons might only project to optically inaccessible regions such as the GNG. For DNs, imaging in the brain is possible in principle, but it is challenging to record a large number of them simultaneous, as their cell bodies are spread across the brain's surface. Recording a volume that is large enough to include a large number of DNs would therefore lead to very slow frame rates. This is further complicated by the lack of a genetic driver line targeting DNs uniquely. We therefore aim to develop an approach that will allow us to image neural activity in the cervical connective and VNC.

So far, most of the work on ascending neurons has either focused on primary ascending neurons [51] or individual secondary ascending neurons with very specific functions [53–55]. In this study, we aim to take a more general look at what information is encoded by secondary ascending neurons.

Our current knowledge of descending control of behavior in insects suggests that individual command like DNs drive specific behaviors [53, 74, 75, 77, 78, 88]. However, this kind of control seems rather inflexible and it is not clear how complex movements are accurately controlled with only ~ 700 DNs in *Drosophila* [59]. Here, we aim to address the question of how many and in what way neurons are involved in spontaneous and odor-evoked behavior by imaging large numbers of DNs simultaneously in behaving *Drosophila*. This will help us determine the extent to which behaviors are controlled by many DNs simultaneously, a sparse sets of command-like DNs, or something in between. A control approach in between the two extremes could use a core set of DNs to drive the behavior and additional DNs to modulate it and adapt it to the specific situation. Different behaviors might employ control strategies from different ends of the spectrum depending on their need for adaptation. Furthermore, we aim to investigate whether different subsets of neurons drive the same behavior in different sensory contexts.

2 Imaging neural activity in the ventral nerve cord of behaving adult *Drosophila*

Chin-Lin Chen[1], Laura Hermans[1], Meera C. Viswanathan, Denis Fortun, **Florian Aymanns**, Michael Unser, Anthony Cammarato, Michael H. Dickinson, Pavan Ramdya. "Imaging neural activity in the ventral nerve cord of behaving adult *Drosophila*", *Nature Communications*, vol.9, pp.4390, 2018. The article can be found here: https://doi.org/10.1038/s41467-018-06857-z

[1]These authors contribute equally.

My contribution:
I was involved in the data analysis part of the paper. The motion correction approach showed some insufficient regularization causing swirling artifacts. To combat these artifacts, I developed a metric based on vector field divergence that is able to detect the artifacts. I then ran a grid search to determine optimal combinations of the regularization and the feature matching components of the registration (**Figure S2.4 & subsection 2.5.11**). In order to correlate neural activity to behavior we had to classify time points into behavior categories. For this purpose I developed a semi-automated behavior classifier that can detect walking and grooming events (**Figure 2.2 d**). Using the behavior information I generated heat maps based on pixel-wise regression with respect to behavior (**Figure 2.2b, c**).

2.1 Abstract

To understand neural circuits that control limbs, one must measure their activity during behavior. Until now this goal has been challenging, because limb premotor and motor circuits have been largely inaccessible for large-scale recordings in intact, moving animals – a constraint that is true for both vertebrate and invertebrate models. Here, we introduce a method for 2-photon functional imaging from the ventral nerve cord (VNC) of behaving adult *Drosophila melanogaster*. We use this method to reveal patterns of activity across nerve cord populations during grooming and walking and to uncover the functional encoding of moonwalker ascending neurons (MANs), moonwalker descending neurons (MDNs), and a novel class of locomotion-associated A1 descending neurons. Finally, we develop a genetic reagent to destroy the indirect flight muscles and to facilitate experimental access to the VNC. Taken together, these new approaches enable the direct investigation of circuits associated with complex limb movements.

2.2 Introduction

Limbs allow animals to rapidly navigate complex terrain, groom, manipulate objects, and communicate. In vertebrates, neural circuits in the spinal cord coordinate the actions of each arm or leg. Thoracic circuits perform comparable tasks in insects [113]. The thoracic segments of the fruit fly, *Drosophila melanogaster*, house the ventral nerve cord (VNC) which is a fusion of three thoracic and eight abdominal ganglia. The VNC contains six spherical neuromeres, each controlling one leg, a flat dorsal neuropil associated with the neck, wing, and halteres, and a set of intermediate neuropils including the tectulum that may coordinate the action of the legs and wings [114]. Also within the thoracic VNC are descending [59] and ascending [51] axons that connect the VNC and the brain. These tracts run through the neck or cervical connective, which – like the VNC – is inaccessible in most preparations.

In recent years, the VNC of adult *Drosophila* has gained attention as the site where some higher-order decisions are transformed into actions. Adult flies engage in complex limbed behaviors including walking [115, 116], reaching [117], escape jumping [118], courtship tapping [119], aggressive boxing [120] , and grooming [121]. Our current understanding of how thoracic circuits coordinate these actions is entirely based on behavioral genetics or recordings from a few neurons in tissue explants [54], immobilized animals [52, 122, 123], or sharp electrode studies in larger insects [124, 125].

To fully understand how thoracic circuits orchestrate limb movements, it is necessary to record the activity of individual cells and populations of neurons during behavior. To date, these experiments have not been performed in *Drosophila* due to the difficulty of accessing the VNC in intact, behaving animals. Here we describe a preparation that overcomes this obstacle and makes it possible to record the dynamic activity of populations and sparse sets of individual neurons within adult thoracic circuits during walking, grooming, and other actions involving

limb movement.

2.3 Results

2.3.1 A dissection for accessing the ventral nerve cord

The VNC lies on the thoracic sternum – a cuticular structure that anchors the leg muscles and the proximal leg segments to the thorax (**Figure 2.1**). Consequently, it is difficult to access the VNC by removing ventral thoracic cuticle without destroying musculoskeletal elements required for limb movement. We chose instead to access the VNC dorsally at the expense of flight-related behaviors [126]. This approach requires removing the prescutum and scutum of the dorsal thoracic cuticle, the indirect flight muscles (IFMs), and transecting the proventriculus, crop, and salivary glands of the gut (**Figure 2.1, Figure S2.1**, see Methods).

Using this technique, it is possible to uncover the VNC for functional imaging in flies that are still capable of exhibiting robust behavior, such as walking and grooming, for up to at least 4 h. In one round of studies (n = 46 flies) by a newly trained experimenter 46% of animals produced behaviors, 26% had limb movement deficiencies, and 28% were incapacitated. When comparing walking behaviors between flies with or without a thoracic dissection, we found that dissected flies generate locomotor bouts with likelihoods and velocities within the range of those observed in non-dissected animals (**Figure S2.2**; n = 20 dissected and 20 non-dissected flies). We note, however, that there are more examples of highly active non-dissected animals. We also found that, on average, dissected animals generate longer bouts **(Figure S2.2b**; $P<0.05$ Mann-Whitney U-test). This may be due to the fact that we only recorded from dissected animals that produced limb movements in response to touch or puffs of air. Therefore, they may also have been among those in a higher state of arousal.

Next, to illustrate the extent of optical access, we drove expression of the genetically encoded calcium indicator, GCaMP6s [19], together with tdTomato [127] — a fluorophore that serves as an anatomical fiduciary — throughout the entire nervous system (*GMR57C10>GCaMP6s; tdTomato*) [128], (**Figure 2.1b, d-g** and **Video S2.1**). To perform 2-photon microscopy in semi-intact, behaving animals, we constructed a customized fly holder and spherical treadmill (**Figure S2.3**) that, in contrast to previous methods used to record neural activity in the brain [29, 126, 129], permits optical access to the VNC along with unobstructed videography of limb movements.

2.3.2 Imaging the activity of populations of neurons in the VNC

By scanning horizontal x-y image planes in animals expressing GCaMP6s and tdTomato pan-neuronally (*GMR57C10>GCaMP6s; tdTomato*), we could record the detailed time course of neural activity in the prothoracic neuromere during walking and grooming (**Figure 2.1c, e** and **Video S2.2**). Alternatively, we could use a piezo-driven objective to scan coronal x-z image

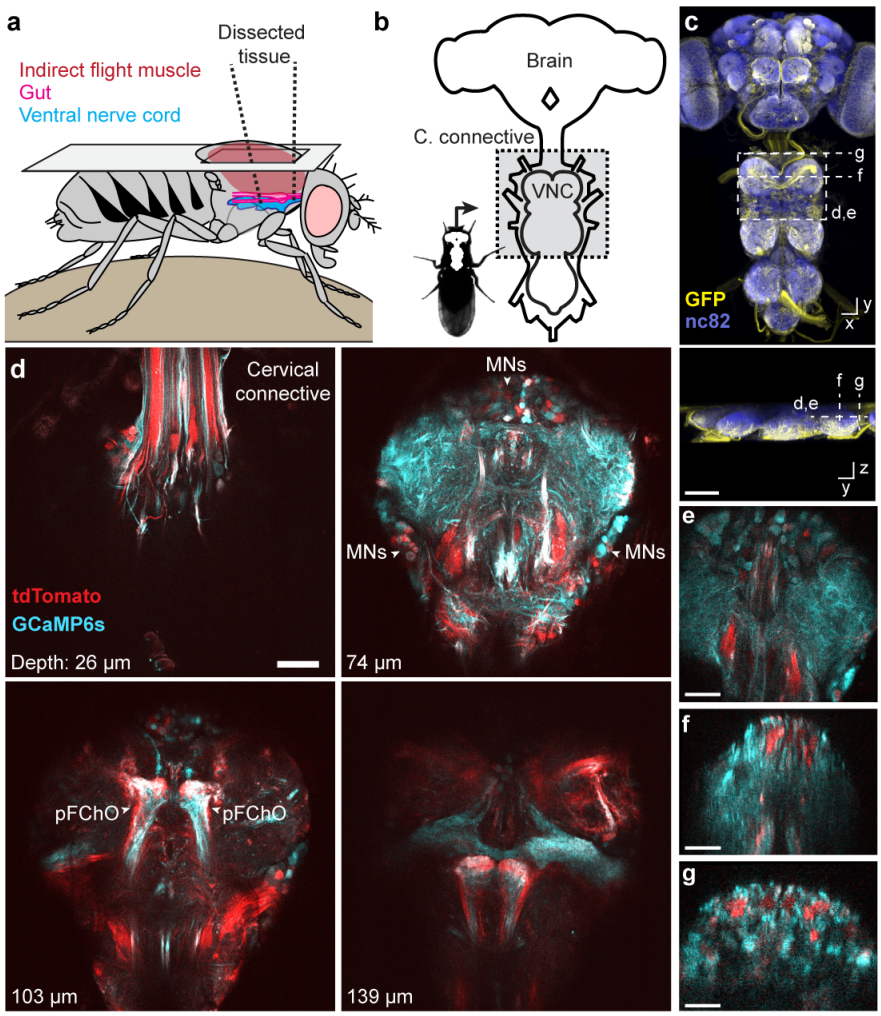

Figure 2.1: **Dissection for imaging the adult *Drosophila* ventral nerve cord (VNC). (a)** Schematic of the dorsal thoracic dissection. **(b)** Overview of newly accessible nervous tissue following thoracic dissection. **(c)** Confocal image of pan-neuronal driver line expression in the brain and VNC. Scale bar is 90 µm. GFP (yellow) and neuropil (nc82, blue) are labelled. Dashed lines highlight the horizontal and coronal imaging modalities used in this study. **(d)** Horizontal sections of the VNC imaged at different depths in an animal expressing GCaMP6s (cyan) and tdTomato (red) throughout the nervous system (*GMR57C10>GCaMP6s; tdTomato*). Motor neuron (MNs) cell bodies and prothoracic femoral chordotonal organ (pFChO) axon terminals are indicated (white arrowheads). Scale bar is 30 µm. **(e)** Horizontal section imaging of the VNC. Scale bar is 35 µm. **(f)** Coronal section imaging of the prothoracic neuromere. Scale bar is 50 µm. **(g)** Coronal section imaging of the cervical connective. Scale bar is 35 µm. Images in **e-g** were taken from flies expressing GCaMP6s and tdTomato throughout the nervous system (*GMR57C10>GCaMP6s; tdTomato*).

planes. These coronal sections allowed us to simultaneously record neural activity across different depths of the VNC corresponding to distinct layers housing sensory neuron axons [51],

interneurons [52], and motor neuron dendrites [130] (**Figure 2.1c, f** and [or to monitor activity patterns across populations of descending [53, 59] and ascending fibers [51, 53, 54] within the thoracic cervical connective (**Figure 2.1c, g** and **Video S2.4**). Thus, we confirmed that our new preparation provides optical access to previously inaccessible thoracic neural populations in behaving adult flies.

During behavior, the VNC moves and deforms. To overcome these image analysis obstacles, we used a non-parametric, variational image registration approach, designed to model arbitrarily complex deformations (**Figure S2.4** and **Video S2.5**, see Methods). After successful image registration, we used a semi-automated approach to annotate walking and grooming behaviors (**Video S2.6**, see Methods) and regressed these two datasets to identify VNC regions whose activities correlated with walking and grooming (**Figure 2.2**). We anticipate that further improvements in image registration will make it possible to build similar behavior-function maps from dense neural population imaging data in which the activity patterns of individual neurons can be identified.

2.3.3 Imaging the activity of sparse sets of neurons in the VNC

Using *Drosophila*, it is possible to repeatedly and systematically investigate the functional properties of sparse sets of genetically-identifiable neurons. In a recent study, a thermogenetic activation screen was used to identify a pair of descending neurons – Moonwalker Descending Neurons (MDNs) – that cause flies to walk backwards [53]. Additionally, concurrent thermogenetic activation of ascending neurons that project from the VNC to the brain – Moonwalker Ascending Neurons (MANs) – resulted in even more sustained backwards walking, perhaps by arresting forward walking [53]. Although these activation experiments show that MDNs and MANs play an important role in the control of backwards walking, their native activity patterns and the means by which they regulate or encode limb movements remain unknown.

Because MAN and MDN axons terminate in the gnathal ganglia (GNG) and the VNC – both relatively inaccessible regions of the nervous system – it is difficult to record the activity of these cells during behavior. We used our functional imaging approach to overcome this challenge and recorded the activity of ascending and descending neurons within the VNC. To overcome vertical movement artifacts associated with walking, we performed coronal section imaging of their axons within the cervical connective (e.g., **Figure 2.3a**).

2.3.4 Activity patterns of moonwalker ascending neurons

Using this approach, MAN axons are visible as small ellipses (**Figure 2.3b**). The MAN split-GAL4 line we used drives expression of GCaMP6s and tdTomato (*MAN>GCaMP6s; tdTomato*) in a pair of dorsal and a pair of ventral neurons. We focused our analysis on the dorsal pair of neurons – hereafter referred to as dMANs – because they showed conspicuous changes in activity (**Figure 2.3c**). The activity of left and right dMANs were strongly correlated (**Figure S2.5a**;

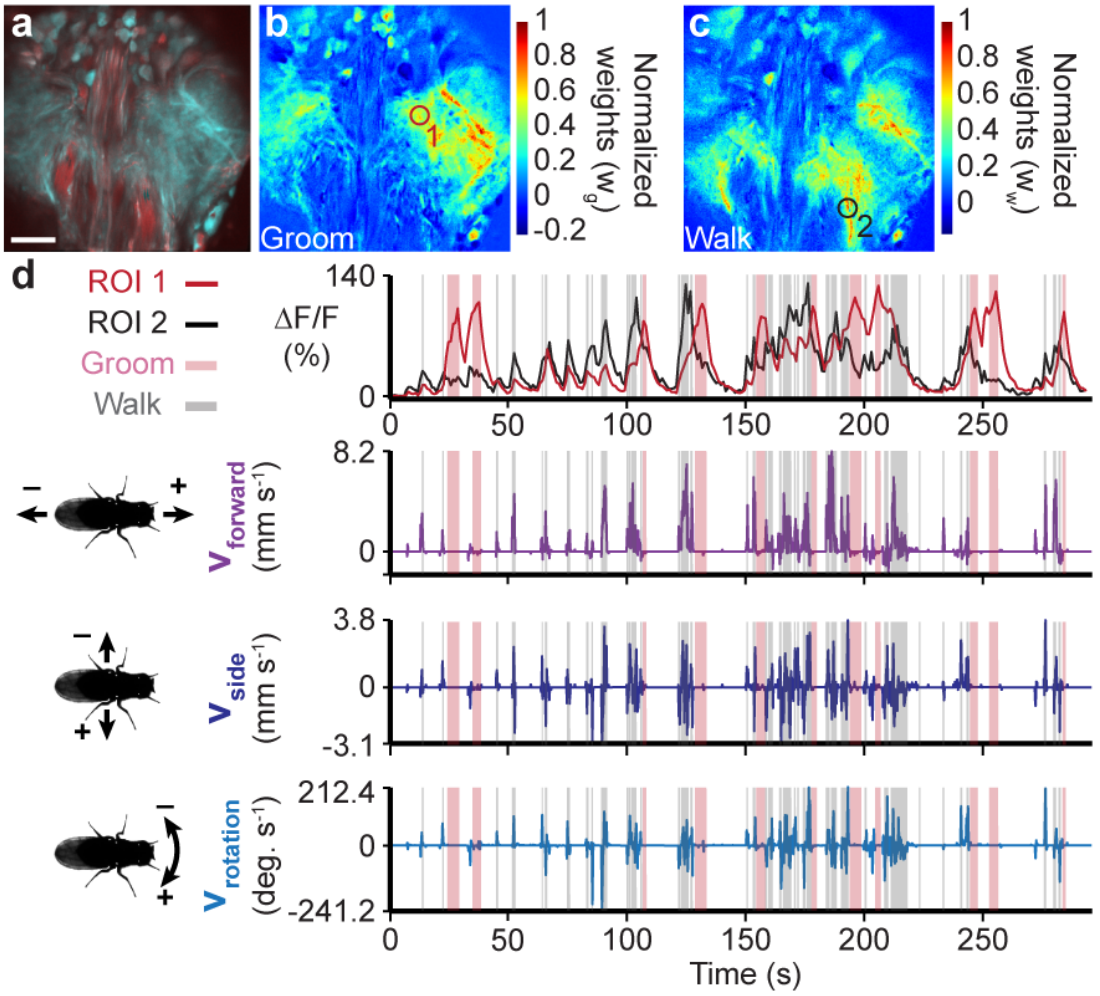

Figure 2.2: **Recording populations of neurons in the VNC during behavior.** (**a**) Standard deviation time projection for an experiment performing horizontal section imaging of the VNC. Scale bar is 35 µm. (**b-c**) Heat maps of linear regression weights w_g and w_w showing the pixel-wise relationships between fluorescence traces and (**b**) grooming or (**c**) walking, respectively. Weights are normalized to the maximum for each image. Data are from the experiment shown in panel **a**. (**d**) ROI-associated fluorescence traces (red from panel b, black from panel **c**) (**top**). Shaded regions indicate semi-automatically detected bouts of grooming (pink) or walking (gray). Corresponding forward, sideways, and rotational velocities of the fly (**bottom**).

Pearson's r = 0.96 ± 0.01, n = 5 flies), allowing us to study their collective response properties. Specifically, we automatically identified the occurrence of transient increases in dMAN fluorescence – referred to as 'events' – and examined corresponding behaviors reflected by rotations of the spherical treadmill (see Methods). Our analysis revealed that dMAN events were associated with rapid bimodal anterior-posterior rotations of the spherical treadmill (**Figure 2.3d**, n = 746 left and 748 right dMAN events from 9773 s of data from 5 flies). Through close inspection of the video data, we observed that these rotations occur when flies extend

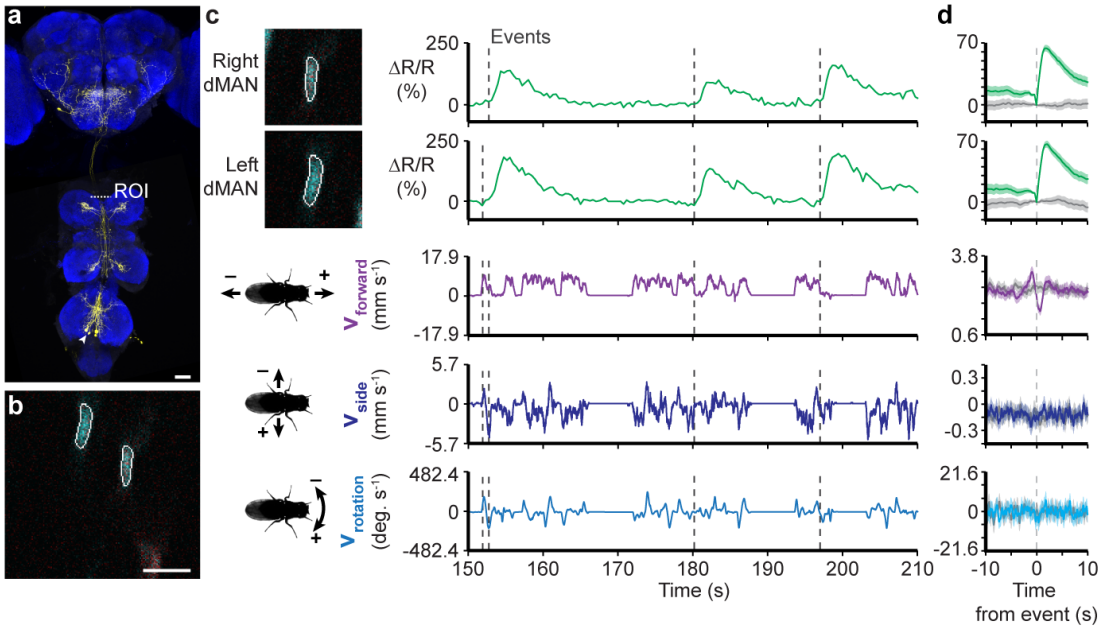

Figure 2.3: **Recording the activity of dorsal Moonwalker Ascending Neurons (dMANs) in the thoracic cervical connective during behavior.** (**a**) Confocal image of *MAN-GAL4* driver line expression in the brain and VNC. Scale bar is 40 µm. Neuronal GFP (yellow) and neuropil (nc82, blue) are labelled. A dashed white line highlights the thoracic x-z plane imaged. (**b**) Coronal section of the thoracic cervical connective in an animal expressing GCaMP6s (cyan) and tdTomato (red) in MANs (*MAN>GCaMP6s; tdTomato*). Scale bar is 3.5 µm. (**c**) Separated ROIs (**top-left**) and associated fluorescence signals from right and left dMANs (**top-right**). Corresponding forward, sideways, and rotational velocities of the fly (**bottom-right**). Events are indicated as dashed gray lines. (**d**) Summary of dMAN activity and spherical treadmill rotations with respect to fluorescence events aligned to 0 s (dashed gray line). Control data in which events are time-shuffled are overlaid in gray. Shown are the means (solid line) and bootstrapped 95% confidence intervals (transparencies).

all six legs to push down on the spherical treadmill (**Video S2.7** and **Video S2.8**).

2.3.5 Activity patterns of moonwalker descending neurons

Next, we asked to what extent MDNs are active during periods of backwards walking, a possibility suggested by behavioral responses to thermogenetic [53] and optogenetic [131] MDN stimulation. To address this question, we performed coronal section imaging of the thoracic cervical connective in flies expressing GCaMP6s and tdTomato in MDNs (*MDN-1>GCaMP6s; tdTomato*) (**Figure 2.4a-b**). As for dMANs, left and right MDN activity patterns were strongly correlated (**Figure 2.4c**), allowing us to study their collective response properties (**Figure S2.5b**; Pearson's r = 0.93 ± 0.001, n = 3 flies). As predicted, MDNs became active prior to anterior rotations of the spherical treadmill, corresponding to brief episodes of backward

walking (**Figure 2.4c-d**, n = 900 left and 900 right MDN events from 3 flies and 7790 s of data; **Video S2.9** and **Video S2.10**).

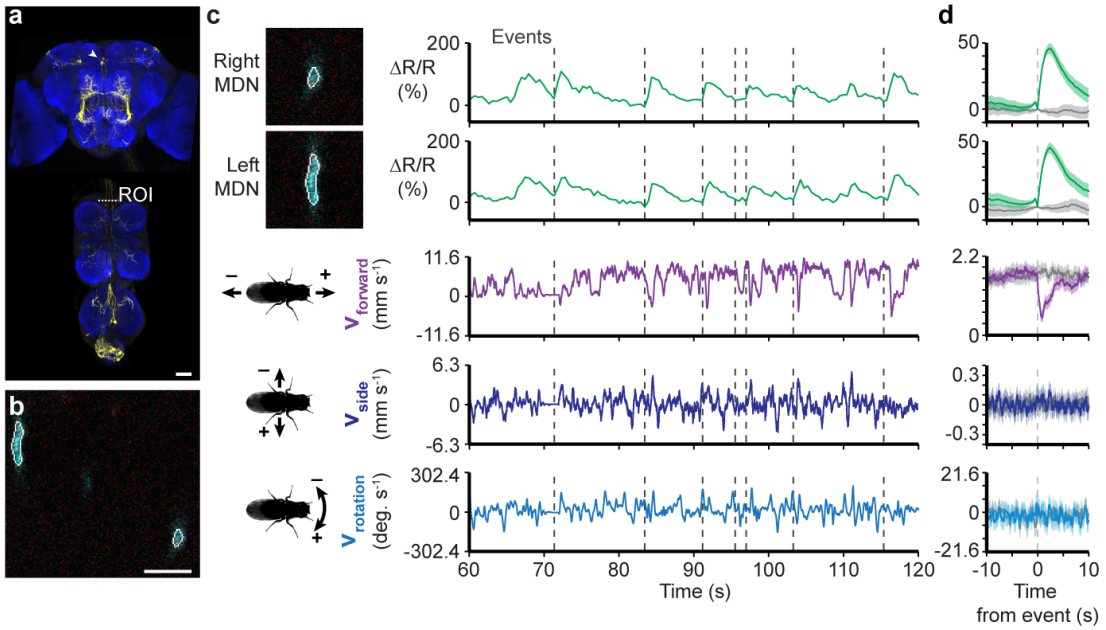

Figure 2.4: **Recording the activity of Moonwalker Descending Neurons (MDNs) in the thoracic cervical connective during behavior.** (**a**) Confocal image of *MDN-1-GAL4* driver line expression in the brain and VNC. Scale bar is 40 µm. Neuronal GFP (yellow) and neuropil (nc82, blue) are labelled. A dashed white line highlights the thoracic x-z plane imaged. (**b**) Coronal section of the thoracic cervical connective in an animal expressing GCaMP6s (cyan) and tdTomato (red) in Moonwalker Descending Neurons (*MDN-1>GCaMP6s; tdTomato*). Scale bar is 6 µm. (**c**) Separated ROIs (**top-left**) and associated fluorescence signals from right and left MDNs (**top-right**). Corresponding forward, sideways, and rotational velocities of the fly (**bottom-right**). Events are indicated as dashed gray lines. (**d**) Summary of MDN activity and spherical treadmill rotations with respect to fluorescence events aligned to 0 s (dashed gray line). Control data in which events are time-shuffled are overlaid in gray. Shown are the means (solid line) and bootstrapped 95% confidence intervals (transparencies).

2.3.6 Activity patterns of novel descending neurons

In addition to resolving the functional properties of previously identified neurons, our method can facilitate the discovery of novel cell classes that are active during walking, grooming, and other behaviors involving the limbs or abdomen. As a proof-of-concept, we selected four split-GAL4 lines [14] that drive sparse expression in pairs of descending neurons [59] whose axons project to leg neuromeres in the VNC (classes DNa01, DNb06, DNg10, and DNg13). We did not observe fluorescence responses during grooming or locomotion in DNg10, or DNg13 cells. DNb06 activity appeared to be only partially correlated with locomotion. By contrast, we observed that DNa01 neurons – hereon referred to as A1 cells – had activity patterns that were

clearly linked to locomotor behaviors (*A1>GCaMP6s; tdTomato*) (**Figure 2.5a-b** and **Video S2.11**). The activity of left and right A1 neurons were not highly correlated (**Figure 2.5c** and **Figure S2.5c**; Pearson's r = 0.53 ± 0.17, n = 4 flies). Therefore, we investigated the response properties of the left and right cells separately. We found that although the activities of both cells are linked to forward walking, events associated only with left A1 activity were correlated with negative medial-lateral and yaw rotations, or leftward turning by the fly (**Figure 2.5d** and **Video S2.12**; n = 1644 events from 4 flies and 8784 s of data). As expected from bilateral symmetry, activity in the right ¬A1 neuron coincided with positive medial-lateral and yaw rotations, or rightward turning (**Figure 2.5e** and **Video S2.13**; n = 1651 events from 4 flies and 8784 s of data).

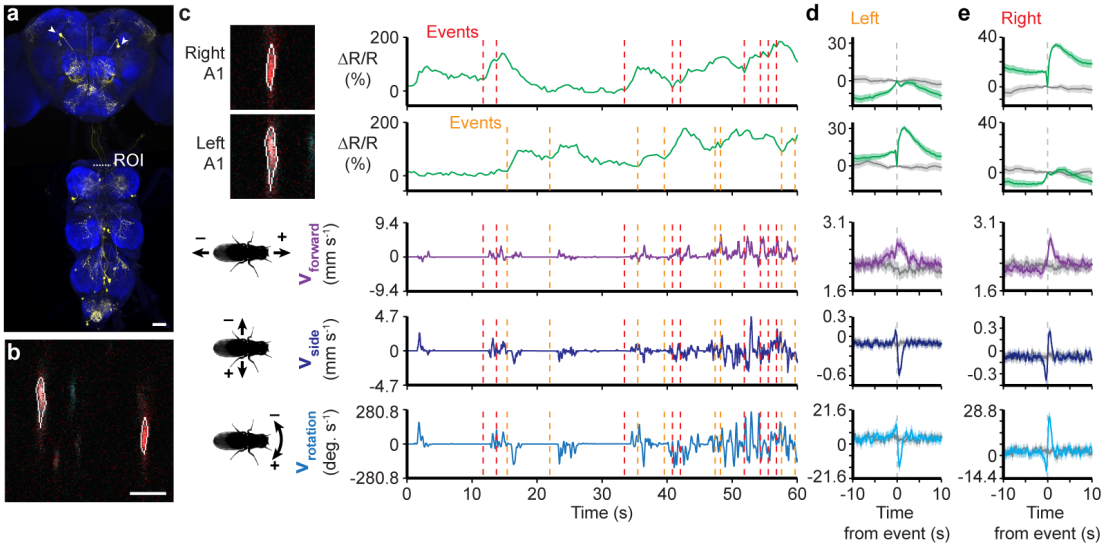

Figure 2.5: **Recording the activity of A1 neurons in the thoracic cervical connective during behavior.** (**a**) Confocal image of *DN0a1-GAL4* driver line expression in the brain and VNC. Scale bar is 40 µm. Neuronal GFP (yellow) and neuropil (nc82, blue) are labelled. A dashed white line highlights the thoracic x-z plane imaged. (**b**) Coronal section of the thoracic cervical connective in an animal expressing GCaMP6s (cyan) and tdTomato (red) in A1 neurons (*A1>GCaMP6s; tdTomato*). Scale bar is 5 µm. (**c**) Separated ROIs (**top-left**) and associated fluorescence signals from right and left A1 neurons (**top-right**). Corresponding forward, sideways, and rotational velocities of the fly (**bottom-right**). Events are indicated as dashed red and orange lines for right and left A1 neuron events, respectively. (**d-e**) Summary of A1 neural activity and spherical treadmill rotations with respect to (**d**) left or (**e**) right A1 neuron fluorescence events aligned to 0 s (dashed gray line). Control data in which events are time-shuffled are overlaid in gray. Shown are the means (solid line) and bootstrapped 95% confidence intervals (transparencies).

2.3.7 Facilitating access to the VNC by inducing cell death in IFMs

Our approach for recording neural activity in the VNC of behaving *Drosophila* opens up many new avenues for studying premotor and motor circuits. Nevertheless, we can envision further

improvements that will accelerate the study of the thoracic nervous system. For example, in our preparation we found it challenging and time-consuming to remove indirect flight muscles (IFMs) that fill most of the thorax. Although large, these muscles are quite fragile and tend to disintegrate over several hours after the cuticle of the notum is removed. However, to increase the speed and efficiency of our dissection, we devised a transgenic strategy to selectively ablate IFMs. We drove the expression of Reaper – a protein involved in apoptosis [132] – in IFMs using a 5' *Act88F* promotor sequence [133]. This loss results in highly elevated or slightly depressed wings – phenotypes seen in IFM developmental mutants [134]. Act88F:Rpr animals show a nearly complete loss of IFMs after 7 days post-eclosion (dpe) when raised at 25°C (**Figure 2.6a-b**). The heterozygous *Act88F:Rpr* transgenic background greatly accelerated the dorsal thoracic dissection. Immediately following eclosion (0 dpe), we observed prominent degradation of IFMs (**Figure 2.6c**). For imaging, *Act88F:Rpr* was most effective at up to 3 dpe: after this stage, the abdominal gut often entered the thoracic cavity. *Act88F:Rpr* also increased the success of dissections: in one round of studies (n = 15 flies) 73% of animals produced behaviors, only 13% had limb movement deficiencies, and only 13% were incapacitated.

We next assessed the degree to which *Act88F:Rpr* might negatively impact tissues beyond the IFMs, including neurons and muscles. Due to the difficulty of identifying regions with Rpr-driven cell death, we instead measured fluorescence in transgenic animals expressing GFP driven by the same promoter sequence (*Act88F:eGFP*) [133]. We observed green fluorescence very rarely and at very low levels outside – but not within – the central nervous system (**Figure S2.6**). We also did not observe any fluorescence in the leg muscles of *Act88F:eGFP* animals (**Figure S2.7**). These anatomical observations were further supported by behavioral responses to antennal infrared laser stimulation: Act88F:Rpr (*Act88F:Rpr; UAS-GCaMP6s-p2A-tdTomato; R57C10-GAL4*) and control (*+;UAS-GCaMP6s-p2A-tdTomato; R57C10-GAL4*) animals exhibited qualitatively indistinguishable walking behaviors at 7 dpe. However, we did observe very small, quantitative differences in walking speed near the end of the stimulation pulse (**Figure S2.8**), n = 15 *Act88F:Rpr* animals and n = 15 control animals; n = 10 responses per animal; $P<0.001$ Friedman test, then $P<0.05$ Mann-Whitney U-test with Holm-Bonferroni correction). Finally, we recorded neural activity in Act88F:Rpr flies and observed no qualitative differences between these animals (*Act88F:Rpr; elav-GAL4/+; UAS-GCaMP6s/+*) and their control counterparts (*elav-GAL4/+; UAS-GCaMP6s/+*) (**Video S2.14**).

2.4 Discussion

Several additional modifications might increase the power of our VNC imaging approach. First, we used coronal section imaging to record from sparse sets of descending and ascending neurons. This strategy was chosen to overcome movement issues observed during horizontal section imaging (**Video S2.15**). Technologies for reducing axial resolution to achieve video-rate 2-photon imaging could be used to overcome this problem [135]. Second, we currently resect the gut to gain access to the VNC. Although this intervention does not profoundly impact limb movements and we can consistently record behaviors for at least 40 minutes (**Video S2.16,**

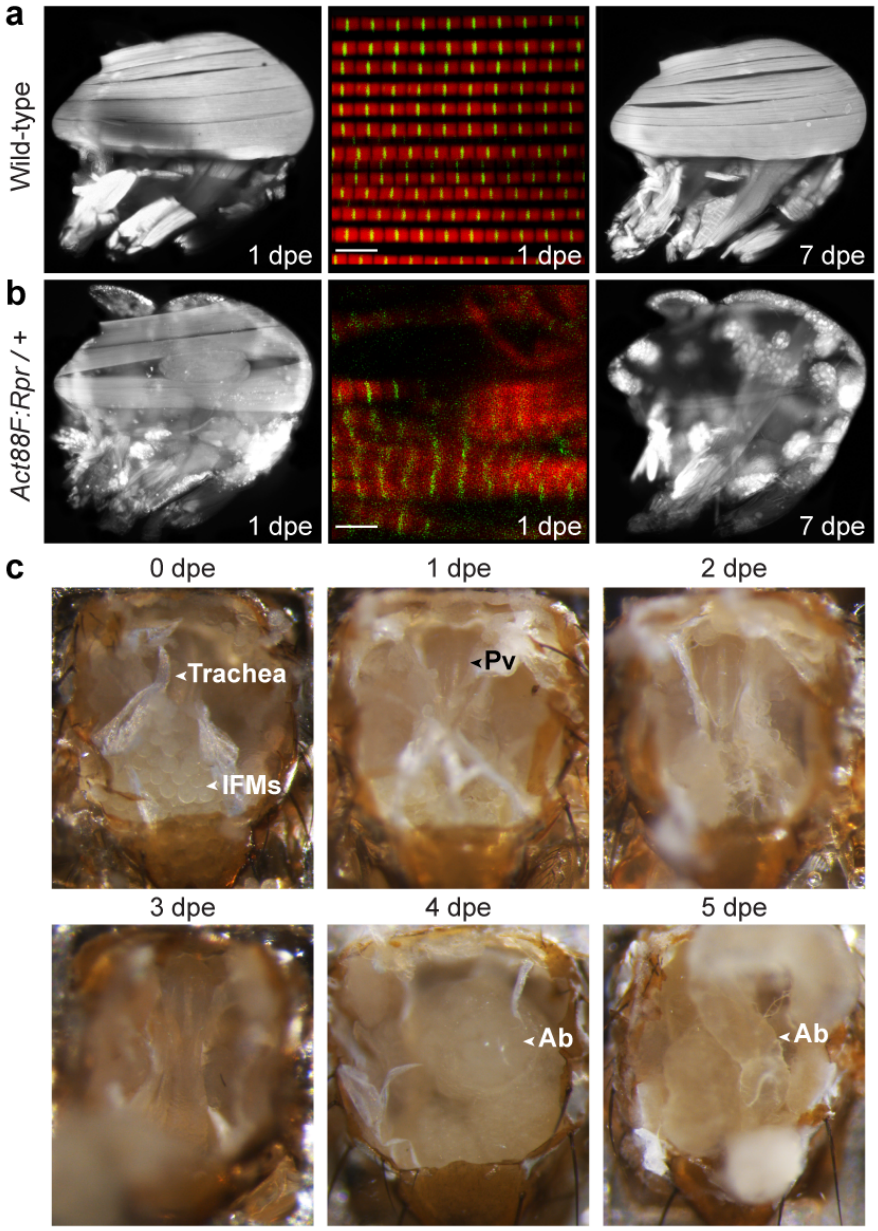

Figure 2.6: **Indirect flight muscle degradation in *Act88F:Rpr* animals.** Confocal images of dorsal longitudinal IFMs (DLMs) stained with TRITC-Phalloidin at 1 dpe (**left**), 7 dpe (**right**), or whole-mount confocal micrographs of myofibrillar structure (**middle**) for (**a**) wild-type, or (**b**) Act88F:Rpr heterozygous flies. Scale bars are 5 µm. (**c**) IFM degradation over time in heterozygous *Act88F:Rpr*animals (*Act88F:Rpr/elav-GAL4; UAS-GCaMP6s/+*). IFMs are already absent at 1 dpe, revealing the underlying proventriculus (Pv). At 4 dpe and later, abdominal gut (Ab) invades the thoracic cavity.

Video S2.17, Video S2.18), we predict that efforts to leave the gut intact will permit even longer recordings.

We have shown that we can record the activity of large VNC neural populations (**Figure 2.2**) as well as sparse cell classes with known (**Figure 2.3** and **Figure 2.4**), or unexplored functional properties (**Figure 2.5**). Our findings have been confirmatory – MDN activity correlates with backward walking – as well as unexpected – dMAN activity correlates with limb pushdown behaviors. With fluorescence decay transients ($t_{1/2}$) on the order of 1s [19], we cannot currently establish whether these signals precede these behaviors. However, based on previous observations [53], MDNs and perhaps A1 descending neurons likely drive locomotor behaviors. By contrast, MAN activity may report behavioral signals to higher-order decision-making centers in the brain. These first results suggest that our recording method, in conjunction with genetic behavioral screens [45, 53, 136], will become an indispensable tool for unraveling how signaling between the brain and VNC gives rise to complex motor actions.

2.5 Materials and methods

2.5.1 *Drosophila* lines

Several lines (*GMR57C10-GAL4*, *elav-GAL4*, *UAS-GCaMP6s*, *UAS-GCaMP6f*, *UAS-CD4:tdGFP*, and *UAS-tdTomato*) were obtained from the Bloomington Stock Center. *MAN-GAL4* (*VT50660-AD; VT14014-DBD*) and *MDN-1-GAL4* (*VT44845-DBD; VT50660-AD*) were provided by B. Dickson (Janelia Research Campus). *DNa01-GAL4* (SS00731: *GMR22C05-AD; GMR56G08-DBD*), *DNb06-GAL4* (SS02631: *GMR20C04-AD; BJD113E03-DBD*), *DNg13-GAL4* (SS02538: *BJD118A10-AD; BJD123E03-DBD*), and *DNg16-GAL4* (SS01543: *BJD104A07-AD; BJD107F12-DBD*) were provided by G. Rubin (Janelia Research Campus). *Act88F:GFP* was obtained from R. Benton (University of Lausanne). *Actin88F:Rpr* strains (*Act88F:Rpr* flies) were generated using the *Actin88F:eGFP* construct described previously [133] and injected (BestGene, Chino Hills, CA, USA) with the phiC31-based integration system using the attP18 (X chromosome) landing site [137]. For some experiments, this transgene was combined with *UAS-GCaMP6s-p2A-tdTomato* (generated in the laboratory of M.H.D.).

2.5.2 Generation of Act88F:Rpr construct and flies

Actin88F:Rpr strains (*Act88F:Rpr* flies) were generated using an *Actin88F:eGFP* construct29. The *Act88F:GFP* line, which houses an *eGFP* construct driven by a 2053 bp region of the *actin88F* promoter, was obtained from R. Benton (University of Lausanne). An *Act88F:Rpr* construct was generated by first using the following primer pair, to add a KpnI restriction site to the 5' end of a *rpr* cDNA clone (IP02530, Drosophila Genomics Resource Center, Bloomington, IN) and an XbaI site to the 3' end of the open reading frame, via a QuikChange Site-directed mutagenesis kit (Agilent Technologies):

Forward primer 5' AGACGGTACCATGGCAGTGGCATTC 3'

Reverse primer 5' GCCGCGTCTAGATCATTGCGATGGCTT 3'

The *Rpr* construct was then spliced into the *Act88F:eGFP* construct behind the *Act88F* promoter in place of the *eGFP* sequence. The *Act88F:Rpr* construct was injected into attp18 *Drosophila* embryos for PhiC31 integrase-mediated site-specific transgenesis35 (transgene landing site cytolocation 6C12) by BestGene Inc. (Chino Hills, CA). For some experiments, this transgene was combined with *UAS-GCaMP6s-p2A-tdTomato* (generated in the laboratory of M.H.D.).

2.5.3 Fluorescence imaging of indirect flight muscles

Fluorescent microscopy of hemi-thoraces was performed as described previously [138, 139]. Briefly, flies were anesthetized and their heads and abdomens were then removed. Thoraces were fixed overnight in 4% paraformaldehyde at 4°C and rinsed in 1x phosphate buffered saline (PBS) the following day. The specimens were arranged on a glass slide, snap frozen in liquid nitrogen and bisected down the midsagittal plane using a razor blade. IFMs were stained with Alexa-Fluor 568 Phalloidin (1:100 in PBS with 0.1% Triton-X (PBST)) overnight at 4°C, rinsed with PBS and visualized using EVOS® FL Cell Imaging System (Life Technologies) at 4x magnification. For whole mount imaging of IFM myofibrils, flies were prepared and thoraces bisected as described above. Hemi-thoraces were stained with Alexa-Fluor 568 Phalloidin (1:100 in PBST) overnight at 4°C. Samples were rinsed in PBS, mounted with Vectashield (Vector Laboratories) and visualized using a Leica TCS SPE RGBV confocal microscope (Leica Microsystems) at 100x magnification.

2.5.4 Immunofluorescence imaging of whole-mount brains and ventral nerve cords

Brains and VNCs were dissected out of 2-3 dpe female flies in PBS. Tissues were then fixed for 20 min in 4% paraformaldehyde in PBS at room temperature. After fixation, brains and VNCs were washed 2-3 times in PBS with 1% Triton-X-100 (PBST) for 10 min each and then incubated at 4°C overnight in PBST. Samples were then placed in PBST with 5% normal goat serum (PBSTS) for 20 min at room temperature. They were then incubated with primary antibodies (rabbit anti-GFP at 1:500, Thermofisher RRID: AB_2536526; mouse anti-Bruchpilot/nc82 at 1:20, Developmental Studies Hybridoma Bank RRID: AB_2314866) diluted in PBSTS for 48 h at 4°C. Brains and VNCs were rinsed 2-3 times in PBST for 10 min each before incubation with secondary antibodies (goat anti-rabbit secondary antibody conjugated with Alexa 488 at 1:500; Thermofisher; goat anti-mouse secondary antibody conjugated with Alexa 633 at 1:500; Thermofisher) diluted in PBSTS for 48 h at 4°C. Finally, brains and VNCs were rinsed 2-3 times for 10 min each in PBST and mounted onto slides with bridge coverslips in Slowfade mounting-media (Thermofisher).

Samples were imaged using a Carl Zeiss LSM 700 Laser Scanning Confocal Microscope with the following settings: 20x magnification, 8-bit dynamic range, 2x image averaging, 0.52 × 0.52 µm pixel size, 0.57 µm z-step interval. Standard deviation z-projections of imaging volumes were made using Fiji [140]. To compare GFP expression in the central nervous system, laser intensity and PMT gains for the green channel were kept constant across wild-type, *A1>GFP*, and *Act88F:GFP* samples.

2.5.5 Imaging GFP expression in leg muscles

Legs were manually dissected at the body-coxa joint and mounted onto glass slides using double-sided tape. Slowfade mounting-media (Thermofisher) was then added to the space between the cover slip and the slide. We then recorded GFP fluorescence in the green channel using an LSM 700 Laser Scanning Confocal Microscope (Zeiss). Laser intensity, PMT gains, and scanning parameters were kept constant across wild-type, *MHC>GFP* and *Act88F:GFP* animals: 20x magnification, 8-bit dynamic range, 8x image averaging, 0.63 × 0.63 µm pixel size, and 10 µm z-step interval. Cuticular auto-fluorescence was also recorded in the red channel.

2.5.6 Thoracic dissection for VNC imaging

Custom holders used to mount flies during imaging were fabricated as described previously [141]. For VNC imaging, these stages were modified to have (i) flat rather than folded steel shims, and (ii) chamfered vertices to make the spherical treadmill visible to optic flow sensors (Shapeways, https://github.com/NeLy-EPFL/Imaging-Drosophila-VNC-and-CC/blob/master/VNC-CC-Imaging-stage/vnc-cc-imaging-stage.STL). Steel shims were fabricated from 0.001" Stainless Steel, type 316 soft annealed (McMaster-Carr, part #2317K11). Shims were etched (Etchit, Buffalo, MN) to generate rectangular holes as described in http://ptweir.github.io/flyHolder/. The shim design file can be found here: https://github.com/NeLy-EPFL/Imaging-Drosophila-VNC-and-CC/blob/master/VNC-CC-Imaging-stage/shims.DXF.

All experiments were performed on 1-3 dpe female flies raised at 25ºC on standard cornmeal food on a 12 h light:12 h dark cycle. Flies were anesthetized at 4ºC. A female fly was selected and, in some cases, its wings were clipped to simplify the mounting process. The fly's dorsal thorax was then pushed through a hole in the steel shim of the imaging stage. The stage was then flipped over, UV-curing glue (Bondic, Aurora, ON Canada) was carefully applied around the perimeter of the thorax and cured by UV illumination (LED-200, Electro-Lite Co. Bethel, CT USA). UV glue was then used to fix the head and abdomen to the underside of the stage. The stage was then filled with extracellular saline as described previously [126]. Under a high-magnification dissection microscope (Leica M165C), a hypodermic needle (30G, BD PrecisionGlide, Franklin Lakes, NJ USA) was used to slice and lift the cuticle off the dorsal thorax [142], being careful not to sever the neck connective. Subsequently, in non-*Act88F:Rpr* animals, a pair of dull forceps was used to remove IFMs, predominantly from the anterior-medial region of the thorax overlying the gut (this step is unnecessary in

Act88F:Rpr animals). This process exposes the dorsal surface of the proventriculus – a large bulbous gut structure. With great care, a pair of super-fine forceps was then used to grasp and lift the proventriculus to displace much of the gut (including the crop and salivary glands) from the more ventrally located nervous tissue. With the gut thus elevated, ultra-fine clipper scissors (Fine Science Tools, Foster City, CA USA) were used to transect it at its anterior-most section. The proventriculus was then peeled back and a posterior incision was made to completely remove these portions of the gut, revealing the underlying nervous tissue. Notably, this dissection also removes the aorta, restricting hemolymph flow from the abdominal dorsal vessel. Nevertheless, we found that flies were viable and behaved for up to 4 h. In some cases, we observed that gut or muscle tissue would begin to obscure the VNC during imaging. Therefore, loose tissue should be removed at this stage while taking great care not to sever the VNC. After each dissection, we examined the extent to which the animal moved its legs in response to a puff of air or grabbed an object with each of its legs. This proved to be an accurate predictor of the success of the preparation. To evaluate the quality of a dissection, we examined the movements of each leg on the spherical treadmill. If a fly could walk in a coordinated manner, the dissection was considered successful. Otherwise, the animal was categorized as having a limb movement deficiency. Animals with multiple dysfunctional legs were categorized as incapacitated.

2.5.7 2-photon microscopy during behavior

Experiments were performed in the evening Zeitgeber time (Z.T.) and animals were typically imaged 30-60 min following dissection. Fly holders were secured to a raised platform over the spherical treadmill (**Figure S2.3**a). The VNC was then located using microscope oculars and positioned in the center of the field-of-view by 2-photon imaging.

The spherical treadmill is an aluminum rod with a ball-shaped hole milled at one end [29]. We fabricated 10 mm diameter foam balls (Last-A-Foam FR-7106, General Plastics, Burlington Way, WA USA) and manually spotted them using a Rapidograph pen (Koh-I-Noor, Leeds, MA USA) to provide high-contrast features for optic flow measurements. A 500-600 mL min-1 stream of filtered and humidified air was passed through the holder using a digital flow controller (Sierra Instruments, Monterey, CA USA). Movements of the ball were measured using two optical flow sensors (ADNS3080) outfitted with zoom lenses (Computar MLM3X-MP, Cary, NC USA). The ball and fly were illuminated using a pair of IR LEDs (850-nm peak wavelength) coupled to optic fibers and collimator lenses (ThorLabs, Newton, NJ USA). Optic flow measurements were passed to a microcontroller board (Arduino Mega2560) to be recorded using custom Python code. Simultaneously, video recordings of animals behaving on the ball were made using an IR-sensitive firewire camera (Basler, Ahrensburg, Germany) at approximately 30 fps.

We performed 2-photon microscopy using a Bergamo II microscope (ThorLabs) outfitted with two GaAsP PMT detectors for GCaMP6 and tdTomato imaging and coupled to a Ti:Sapphire laser (MaiTai DeepSee, Newport Spectra-Physics, Santa Clara, CA USA) tuned to 930 nm. We

used an Olympus 20x water-immersion objective lens with 1.0 NA (Olympus, Center Valley, PA USA). The microscope was controlled using ThorImage software (ThorLabs). Coronal section imaging experiments were performed in Galvo-Galvo imaging mode at 6 - 9 Hz. This framerate varied with image size which ranged between 26.58 µm x 26.58 µm and 53.15 µm x 53.15 µm. Laser power ranged between 3 mW and 5.7 mW. Volumetric imaging is also possible with appropriate hardware (e.g., Galvo-Resonance scanner and Piezo-driven objective collar).

Occasionally, a puff of air was used to elicit walking behaviors. These puffs were digitally encoded (Honeywell AWM 3300V, Morris Plains, NJ USA). Custom ROS software interfaced through an analog output device (Phidgets, Calgary, Canada) to ThorSync software (ThorLabs) was used to synchronize optic flow measurements, behavior videography, air puff measurements, and 2-photon image acquisition. For coronal section imaging, a Piezo collar (Physik Instrumente, Karlsruhe, Germany) was used to control rapid z-axis movements of the microscope objective lens.

To compare neural activity between control and *Act88F:Rpr* animals, we acquired 512 x 512 pixel images at 1.7 fps using a constant laser intensity and PMT gain. Selected imaging regions were empirically chosen as horizontal sections consisting of landmarks observed at 61-65 µm depth in **Video S2.1**.

2.5.8 Comparing walking behaviors in dissected or non-dissected animals

To evaluate the effects of dissection on locomotion, wild-type flies were subjected to the following procedure. Flies were mounted onto imaging stages and saline was added to each stage. A random subset of animals was dissected. All mounted flies were then placed onto the spherical treadmill for 30 min. Optic flow was recorded as described. To increase the likelihood of locomotion, a 500 ms pulse of 100% CO_2 was directed at the fly's antennae with a one min inter-pulse interval (0.05 l_n min^{-1} using a mass flow controller; Vögtlin Instruments, Switzerland).

2.5.9 Infrared laser antennal stimulation

To compare walking behaviors between *Act88F:Rpr* and control animals, we stimulated their antennae with an 830nm near infrared laser (Schäfter+Kirchhoff, Germany). We first anesthetized 7-8 dpe female animals at 4°C and mounted them on imaging stages. Flies were then acclimated for 10 min. For each experiment, an animal received ten 2 s laser stimulation pulses (18.1 mW) to its right antenna at a 60 s inter-pulse interval. Control and *Act88F:Rpr* animals were tested in alternation to minimize the effects of circadian time on behavioral comparisons.

2.5.10 Statistics

Sample sizes for animal experiments were chosen as follows: we performed at least three experiments to illustrate population and sparse neural recordings and performed more than ten experiments per group when performing statistical comparisons. A pre-established criteria of low signal-to-noise fluorescence signals resulted in the removal of two MDN experiments from our dataset. No randomization or blinding was used. For antennal laser stimulation, data were not normally distributed, thus Friedman and Mann-Whitney U-tests were performed. Estimates of variation are presented as mean and bootstrapped 95% confidence intervals.

2.5.11 Data analysis

We analyzed all data using custom Python scripts. Because the data acquisition frequency differed for optic flow, behavior videography, and 2-photon imaging we interpolated signals to match those of the highest frequency. Subsequently, optic flow data were smoothed using a running average (window = 200 ms) and then translated into rotations s^{-1} for anterior-posterior, medial-lateral, and yaw axes as described in [29]. To make these measurements more intuitive, rotations s^{-1} were then converted into mm s^{-1} (1 rot s^{-1} = 31.42 mm s^{-1} for anterior-posterior ($v_{forward}$) and medial-lateral (v_{side}) movements and into degrees s^{-1} (1 rot s^{-1} = 360° s^{-1} for yaw ($v_{rotation}$) movements [29].

Evaluation of locomotion in dissected animals (related to Figure S2.2)

The analysis of locomotion in dissected animals (**Figure S2.2**) was performed as follows. $V_{forward}$ optic flow data for 20 dissected and 20 non-dissected flies were downsampled to 1500 points s^{-1} and smoothed using a running average of duration 0.2 s. To compute the percentage of time walking forward/backward, or walking sideways two thresholds, -0.31 mm s^{-1} and +0.31 mm s^{-1} , were empirically defined to differentiate between standing still and forward (rightward) or backward (leftward) walking, respectively. Values above 0.31 mm s^{-1} were considered moments of forward (rightward) walking and values below $-0.31\,\mathrm{mm\,s^{-1}}$ were considered moments of backward (leftward) walking. Optic flow values between these thresholds were considered moments of standing still. The percentage of time walking was calculated as the proportion of data points in which an animal was not considered standing still. Similarly, thresholds of 10.8 and -10.8 degree s^{-1} were used to defined moments of turning. A bout was defined as a continuous period of walking or turning.

Pan-neuronal image registration, ROI identification, and fluorescence processing (related to Figure 2.2)

Large tissue deformations could occur during behavior. Therefore, we performed post-hoc registration of pan-neuronal imaging data. To do this, we registered all frames of an imaging experiment to one reference image. Because the complexity of deformations could not be

captured using simple parametric motion models (e.g., affine transformations), we used a non-parametric, variational approach, designed to model arbitrarily complex deformations. We computed the motion field **w** between the reference image, denoted I_r, and the image at time t, denoted I_t, by solving the minimization problem

$$\hat{\mathbf{w}} = \underset{w}{\operatorname{argmin}} D(\mathbf{w}) + \lambda \sum_{x \in \Omega} ||\nabla \mathbf{w}(\mathbf{x})||_2^2, \tag{1}$$

where $D(\mathbf{w})$ is a data fitting term, the second term is a regularization promoting smoothness of **w** by penalizing its gradient $\nabla \mathbf{w}$ [143], Ω is the discrete image domain, and the parameter λ balances the contributions of the two terms.

GCaMP6s images present a challenge for motion estimation because neural activity produces large local intensity changes. Therefore, we used an additional activity independent fluorophore, tdTomato, and defined a data term of the form

$$D(w) = \rho(w, I_r, I_t) + \gamma \phi(w, I_r, I_t). \tag{2}$$

The first term models the standard assumption of conservation of intensity along the trajectory of each pixel. It is defined by

$$\rho(w, I_r, I_t) = \sum_{x \in \Omega} |I_t(x + w(x)) - I_r(x)|, \tag{3}$$

where we use an l_1 norm to gain partial robustness to intensity changes [144]. The second term in (2) is a feature matching constraint inspired by Revaud and co-workers [145], written as

$$\phi(w, I_r, I_t) = \sum_{x \in \Omega} ||w(x) - m(x, I_r, I_t)||_1. \tag{4}$$

In equations (2) to (4), I_r and I_t are from the tdTomato channel. Minimizing the function ϕ favors motion vectors $\mathbf{w}(\mathbf{x})$ to be close to feature correspondences $\mathbf{m}(\mathbf{x}, I_r, I_t)$, computed on a sparse set of relevant keypoints. We obtain **m** with the feature matching algorithm proposed by Revaud and co-workers [145], which is specifically designed to handle large image deformations. We compute **m** using the tdTomato imaging channel, such that the correspondences are also insensitive to the intensity changes between I_r and I_t. As a result, the estimation is guided by reliable feature matches. The parameter γ balances the two terms in (2).

For each experiment, we optimized the values for λ and γ using a grid search to register horizontal section images of the VNC (**Figure S2.4**). As an objective function for optimization, we used the gradient of the temporal mean image [146]. Small values of λ (i.e. λ <1000), occasionally led to artifacts in the registered images. These artifacts were associated with strong convergence in the vector field $\mathbf{w}(\mathbf{x})$ (**Figure S2.4c**). Therefore, we empirically defined artifacts as clusters of pixels with $div\ \mathbf{w(x)} < -1.2$ and cardinality > 20 (we obtained similar

results with cardinality > 5). Finally, we selected λ and γ values as those with no artifacts and the highest gradient of the mean image. Sample unregistered images, transformation vector fields, and registered images of the three optimized examples are shown in **Video S2.5**.

We solved the optimization problem (1) with an alternated direction method of multiplier (ADMM) algorithm [147]. We introduced two splitting variables, associated with the regularization and the feature matching terms, respectively. Each sub-problem of the algorithm was solved analytically. We used parts of the inverse problems library described in [148]. A post processing based on weighted median filtering was applied using the method from [149].

In **Figure 2.2**, behaviors were semi-automatically annotated, using a custom Python module. This module allows the user to select two ROIs on the video's first frame. The first ROI is used to detect walking and must be positioned over the metathoracic and mesothoracic legs. The second ROI is responsible for detecting prothoracic leg grooming and must be positioned in front of the fly. To detect motion in those regions, consecutive frames are subtracted. Resulting differential images are then median blurred (radius = 5 pixels), to reduce noise. Based on this blurred image, a threshold on the number of non-zeros pixels in each of the two ROIs is applied to extract binary sequences of grooming and walking bouts. Note that prothoracic leg movements observed during walking are ignored (i.e., grooming classification is subservient to walking classification). A hysteresis filter was then applied to low-pass filter binary behavioral sequences and to remove transitions that occur over too few frames to be biologically plausible. Example ROIs and behavioral annotations are illustrated in **Video S2.6**. This behavior data was used in **Figure 2.2** as shown in **Video S2.2**. It was annotated using the following parameters: threshold for walking = 400, threshold for grooming = 5, hysteresis length for walking = 8, hysteresis length for grooming = 10.

For **Figure 2.2b-c**, we used linear regression to find regions in the VNC associated with either walking or grooming. Regressors X_w and X_g (for walking and grooming, respectively) were constructed from the two behavioral sequences, S_w and S_g, using equation (5) by convolution with an exponentially decaying Calcium signal Impulse Response (CIR) derived from the time constant measured for GCaMP6s ($t_{1/2}$=1.1448s) [19].

$$X_w = S_w \otimes CIR$$

$$X_g = S_g \otimes CIR, \tag{5}$$

Target functions were pixel-wise $\Delta F/F$ traces, where $\Delta F = F_t - F$. F_t is the fluorescence at time, t. F is a baseline fluorescence signal measured as the average pixel value for the first ten sequential GCaMP6s images in which no cellular activity was observed (i.e., minimal and unchanging GCaMP6s fluorescence). The regressor weights were calculated using equation (6).

$$w_w = (X_w^T X_w)^{-1} X_w^T y$$

$$w_w = (X_g^T X_g)^{-1} X_g^T y, \tag{6}$$

where y is the pixel-wise $\Delta F/F$ trace.

Figure 2.2b-c shows heat maps of the regressor weights, w_w for walking and w_g for grooming, normalized to their respective maxima. ROI 1 was chosen as a region of the heat map with a high weight for grooming but a low weight for walking. ROI 2 was chosen as the spatial location with the highest value of w_w. Each ROI encompasses a region with a 15 pixel radius.

Sparse neuron ROI identification, and fluorescence processing (related to Figure 2.3, Figure 2.4, Figure 2.5)

For single-neuron fluorescence data, ROIs were selected using custom Python scripts that depended on OpenCV and Numpy libraries. First, a reference frame was selected for which the software identified all potential ROIs. To do this, the GCaMP6s image was smoothed to reduce background noise and then an Otsu filter threshold was applied to the image. An erosion factor was then applied on all objects detected within the image. Contours of all detected objects were then presented to the user for manual selection. Once these reference ROIs were selected for left and right neurons, we used a cross-correlation-based image registration algorithm [150] to identify the most likely left and right ROIs for each image frame based on those manually selected on the reference frame. A second script was used to manually verify automatically selected ROIs and, if incorrect, to display all potential ROIs within the frame for manual selection. If erosion values yielded malformed ROIs, another script was used to manually position elliptical ROIs of arbitrary orientation on any given frame. Finally, binary ROI images were used as an image mask to extract mean fluorescence signals from the original GCaMP6s or tdTomato images. These signals were reported as $\%\Delta R/R$ as in [151] to reduce the effects of motion on our measurements. Due to the absence of stimuli, the baseline R was calculated as the minimum ratio of GCaMP6s/tdTomato within a 2.5 s bin.

To detect transient increases in activity, we developed an algorithm based partly on [152]. We first determined when the first derivative of the $\%\Delta R/R$ signal crossed a threshold, which was determined by examining all derivative values for a given neuron class (MDN, MAN, or A1). We reasoned that threshold values should be characteristic and potentially different for each type of neuron because fluorescence dynamics are related to intrinsic physiological properties that can differ across neuron classes but not across experiments for a single class. We set this threshold as the 97.5th percentile for MDNs and dMANs and the 90th percentile for A1 neurons. A lower threshold value was selected for A1 neurons because many more fluorescence transients were observed in A1 traces. These transients would have been overlooked using a 97.5th percentile threshold. To identify the onset of fluorescence increases we found the nearest preceding time point where the derivative crossed zero. This zero-crossing is considered the time-point of an 'event' associated with the identified fluorescence increase. Events detected close to one another with no intervening derivative zero-crossing were compressed into one event associated with the first time-point. There were 10 separate experiments per

animal. Events in the first and last 10 s of each experiment were not considered since the data presentation window encompassed 10 s before and 10 s after each event.

Because left and right MDN and dMAN activities strongly covaried (**Figure S2.5**), an additional step was performed for event detection: if events were detected in both left and right neurons within 2 s of one another, both events were retained; otherwise, an event identified for neuron A (e.g., left MDN) and not neuron B (e.g., right MDN) was also added to neuron B's event library. By contrast, left and right A1 activities did not strongly covary. Therefore, events were associated with one and not the other neuron. To accomplish this, if an event was detected for both left and right A1 neurons within a time window of 0.25 s, neither of the events were used for analysis.

$\%\Delta R/R$ and optic flow traces linked to each event were aligned by setting the event time points to 0 s. We then computed the mean and bootstrapped 95% confidence intervals for these aligned traces using the Python Seaborn library. Optic flow and $\%\Delta R/R$ measurements were downsampled to 500 values s-1 for this analysis. To increase clarity, $\%\Delta R/R$ traces were baseline-subtracted to make them zero at the time of the event in the summary panels (**Figure 2.3d, Figure 2.4d, Figure 2.55d-e**). Control, shuffled data (gray traces) were computed by instead assigning random time-points in place of real, identified events. These random time points were treated as real events and their mean and bootstrapped 95% confidence intervals were computed and plotted for comparison.

Covariance analysis (related to Figure S2.5)

Covariance analysis was performed using a custom Python script that depended on the Matplotlib and Numpy libraries. Scatter plots were computed to compare left and right neuron $\%\Delta R/R$ values from all experiments for each fly separately. Pearson's r values are reported as mean ± standard deviation.

Event-related behaviors (related to Video S2.8, Video S2.10, Video S2.12, and Video S2.13)

Events for behavioral summary movies were manually selected from automatically detected events as described above. For dMANs, events were selected from among those that maximized the difference in anterior-posterior ball rotations between 1 s before and 2 s after the event. For MDNs, events were selected from among those that minimized anterior-posterior ball rotations up to 2 s after the event. For A1 neurons, events were selected from among those that maximized the average yaw ball rotations (positive for left A1 neuron examples and negative for right A1 neuron examples) for up to 2 s after the events.

Laser-stimulated walking (related to Figure S2.8)

Responses to near infrared laser stimulation were averaged across 10 trials for each animal. Optic flow was downsampled to 500 values s^{-1}. Mean and 95% bootstrapped confidence intervals for optic flow traces were measured and plotted using the Python Seaborn library. The Python Scipy library was used to perform Friedman and Mann-Whitney *U*-tests.

2.6 Code availability

Code and sample datasets used for this study are available at: https://github.com/NeLy-EPFL/Imaging-Drosophila-VNC-and-CC

2.7 Data availability

The data used in this study is available from the authors on reasonable request.

2.8 Acknowledgments

We thank B.J. Dickson (Janelia Research Campus, VA) for MDN-1-GAL4 and MAN-GAL4 fly strains. We thank G. Rubin (Janelia Research Campus, VA) for DNa01-GAL4, DNb06-GAL4, DNg13-GAL4, and DNg16-GAL4 fly strains. AC acknowledges support from the National Institutes of Health (R01HL124091). MHD acknowledges support from the National Institute of Neurological Disorders and Stroke of the National Institutes of Health (U01NS090514). PR acknowledges support from the Swiss National Science Foundation (31003A_175667).

2.9 Author contributions

C.L.C. generated strains; performed experiments; analyzed data **L.H.** performed experiments; analyzed data **M.C.V.** generated strains; performed experiments; analyzed data **D.F.** wrote analysis code **F.A.** analyzed data **A.C.** designed and supervised the project **M.H.D.** designed and supervised the project **P.R.** conceived of, designed, and supervised the project; performed experiments; analyzed data All authors contributed to writing the paper

2.10 Supplementary information

2.10.1 Supplementary Figures

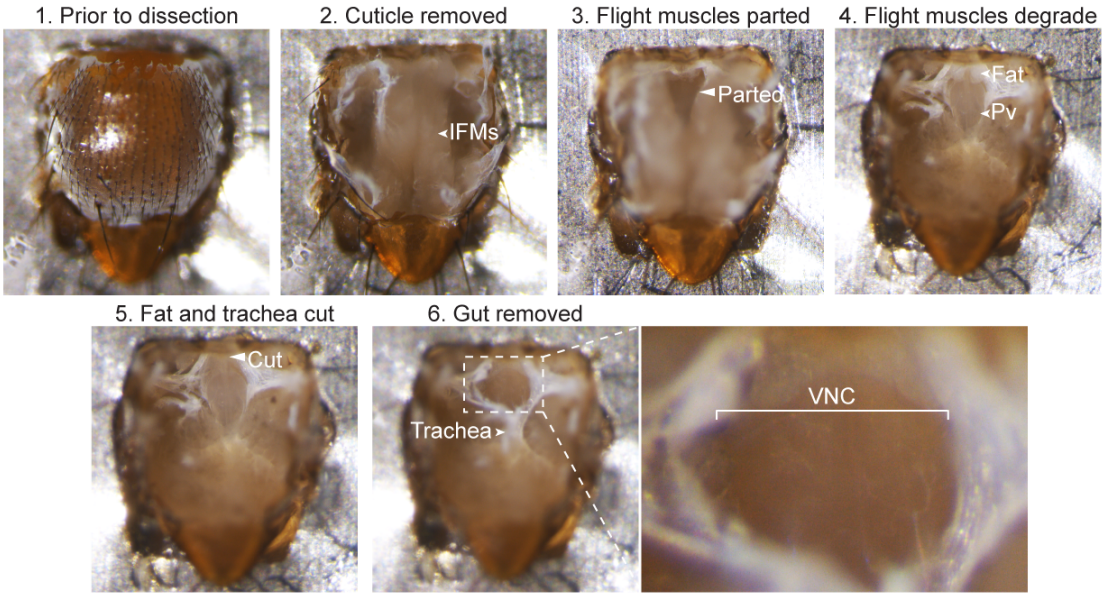

Figure S2.1: **Illustration of dorsal dissection to access the VNC.** First, the dorsal cuticle is removed, revealing indirect flight muscles (IFMs) (**steps 1-2**). The IFMs are then parted at the anterior midline (**step 3**). After >1hr, the IFMs degrade, exposing midline trachea, fat bodies (Fat), and the proventriculus (Pv) (**step 4**). After removing these structures (**steps 5-6**), the prothoracic and mesothoracic VNC (**inset**) become accessible for 2-photon microscopy.

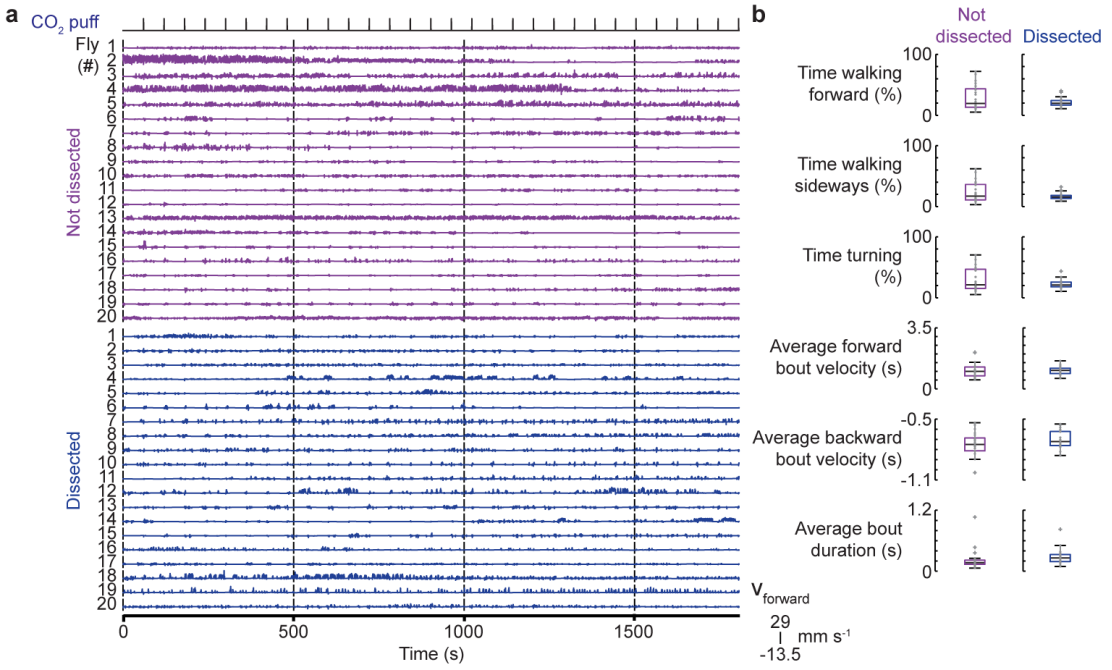

Figure S2.2: **Locomotion in flies with or without thoracic dissection.** (**a**) Raw spherical treadmill traces showing forward velocity for animals without **(top, purple)**, or following the thoracic dissection revealing the VNC (bottom, blue). Timing of CO_2 pulses delivered to the antennae are indicated. (**b**) Locomotor parameters from animals without **(left, purple)** or with **(right, blue)** the thoracic dissection. Parameters include percent of time walking forward, walking sideways, or turning, as well as the average forward bout velocity per animal, average backward bout velocity per animal, and average bout duration per animal. Box plots show the median, upper, and lower quartiles.

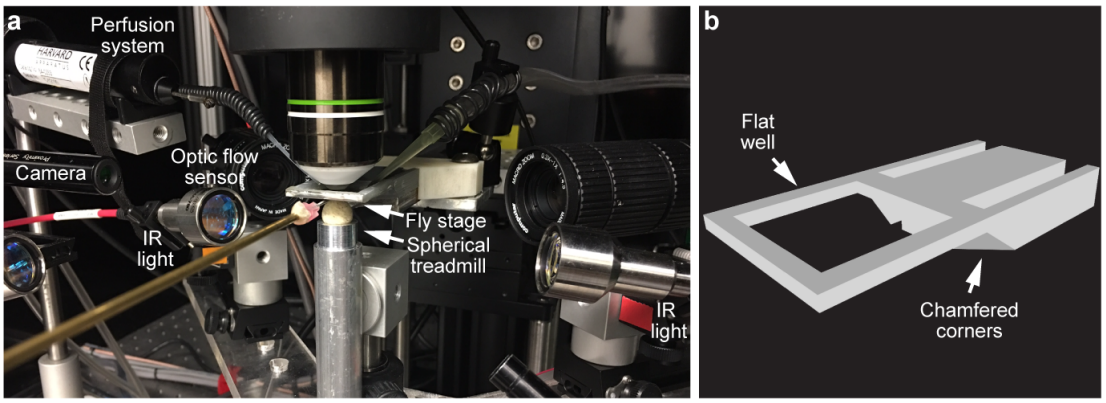

Figure S2.3: **System for VNC imaging.** (**a**) Photograph of the experimental system and (**b**) a CAD schematic of the custom fly stage used in this study.

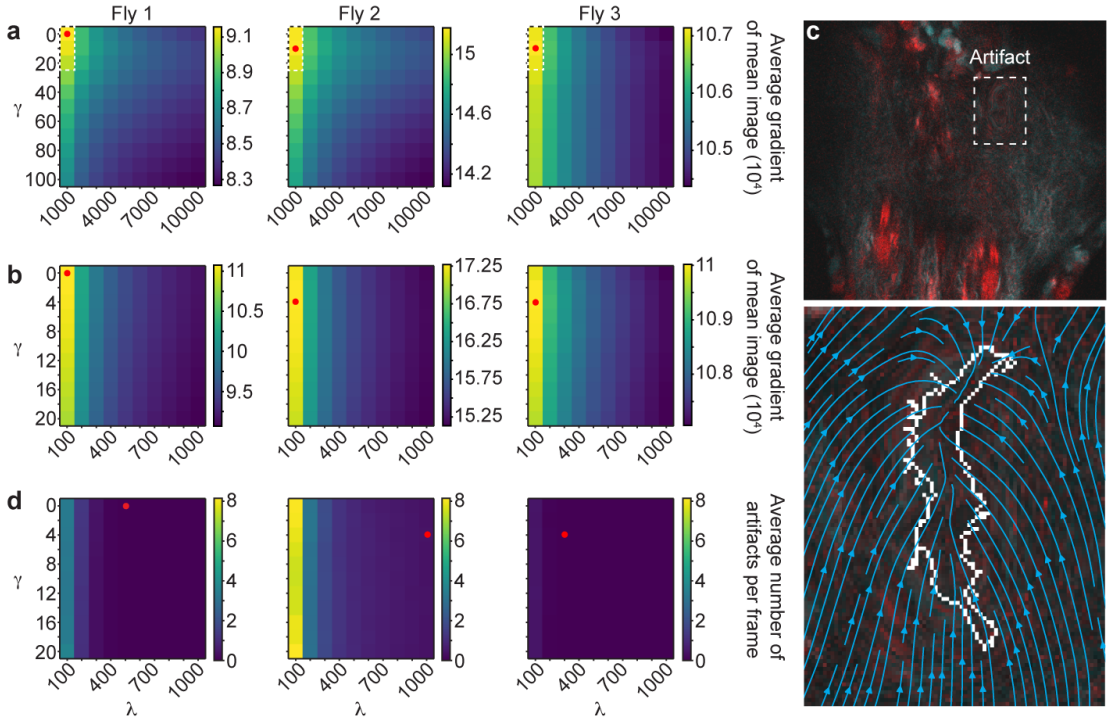

Figure S2.4: **Optimization of λ and γ values for image registration.** (**a-b**) The average gradient of the mean image for (**a**) a coarse grid search of λ and γ values for three horizontal section VNC imaging datasets (n = 3 flies; same data and ordering as in **Video S2.5**). Red dots indicate the maximum value. White dashed boxes indicate the regions explored using (**b**) a finer-scale grid search. (**c**) Illustration of an image registration artifact observed for data from fly 2 when using $\lambda = 100$ and $\gamma = 0$ (**top**). The region in the white inset magnified and overlaid with the transformation vector field (blue arrows) and the boundary (white outline) of the artifact as determined using a threshold on divergence (**bottom**). (**d**) Average number of artifacts per frame. Red dots indicate the final λ and γ values used to register the data shown in **Video S2.5**. These values result in images with no artifacts but a maximal average gradient of the mean image.

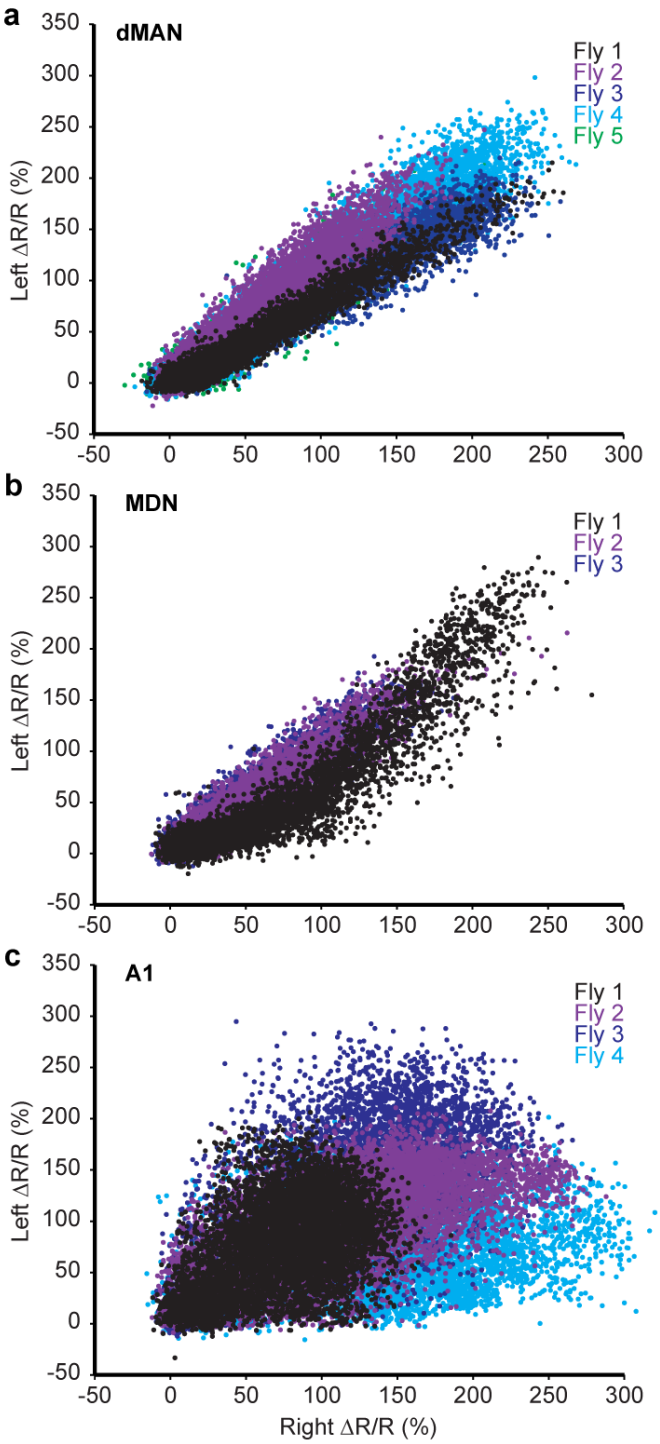

Figure S2.5: **Covariance in fluorescence signals between bilateral pairs of neurons.** Scatter plots comparing $\%\Delta R/R$ signals recorded from right and left (**a**) dMAN, (**b**) MDN, or (**c**) A1 neuron pairs.

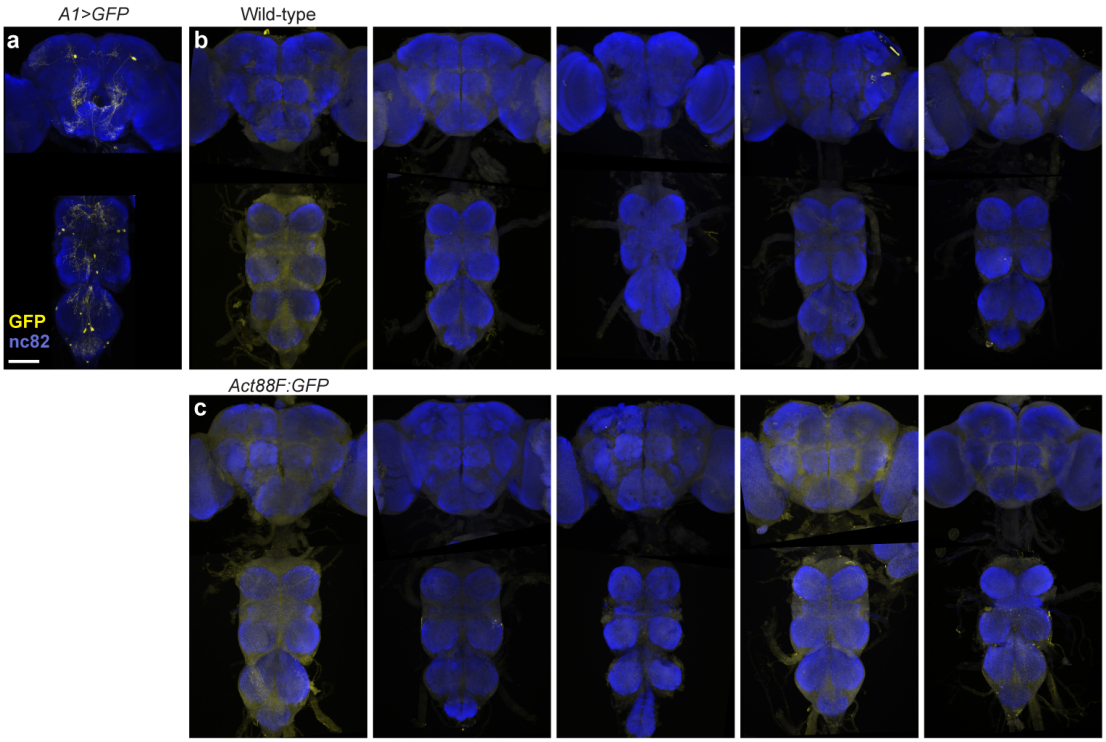

Figure S2.6: **GFP expression in the central nervous system.** Representative confocal images of brains and VNCs from (**a**) one *A1>GFP* fly: a positive control with sparse GFP expression, (**b**) five wild-type flies: negative controls without GFP expression, or (**c**) five *Act88F:GFP* flies. Immunostaining is against GFP (yellow) and nc82 (blue). Scale bar is 180 µm.

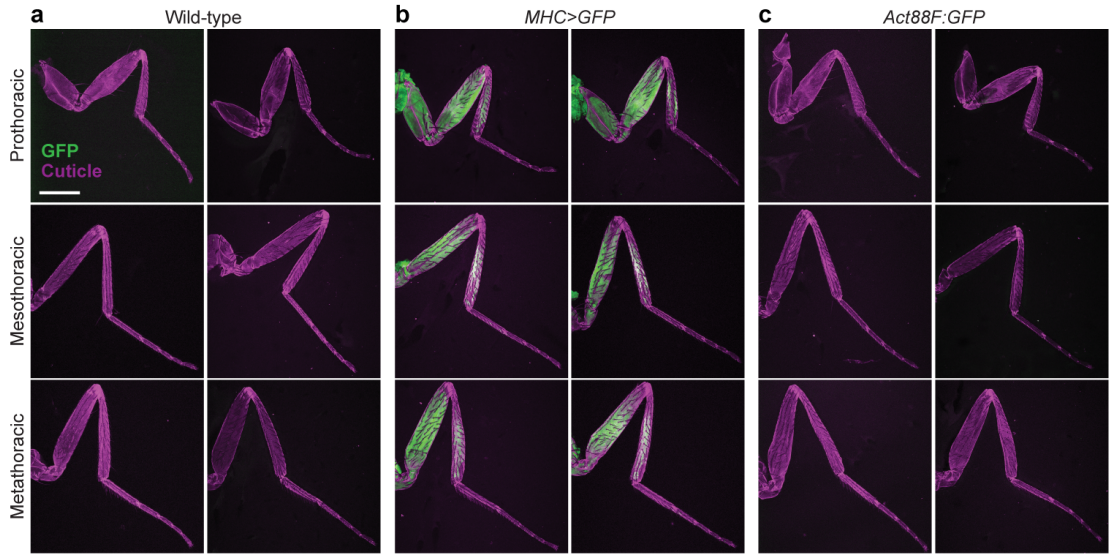

Figure S2.7: **GFP expression in leg muscles.** Representative confocal images of prothoracic, mesothoracic, or metathoracic legs from (**a**) two wild-type flies: negative controls without GFP expression, (**b**) two *MHC>GFP* flies: positive controls with leg muscle GFP expression, or (**c**) two *Act88F:GFP* flies. Shown are endogenous GFP fluorescence (green) and cuticular autofluorescence (magenta). Scale bar is 300 µm.

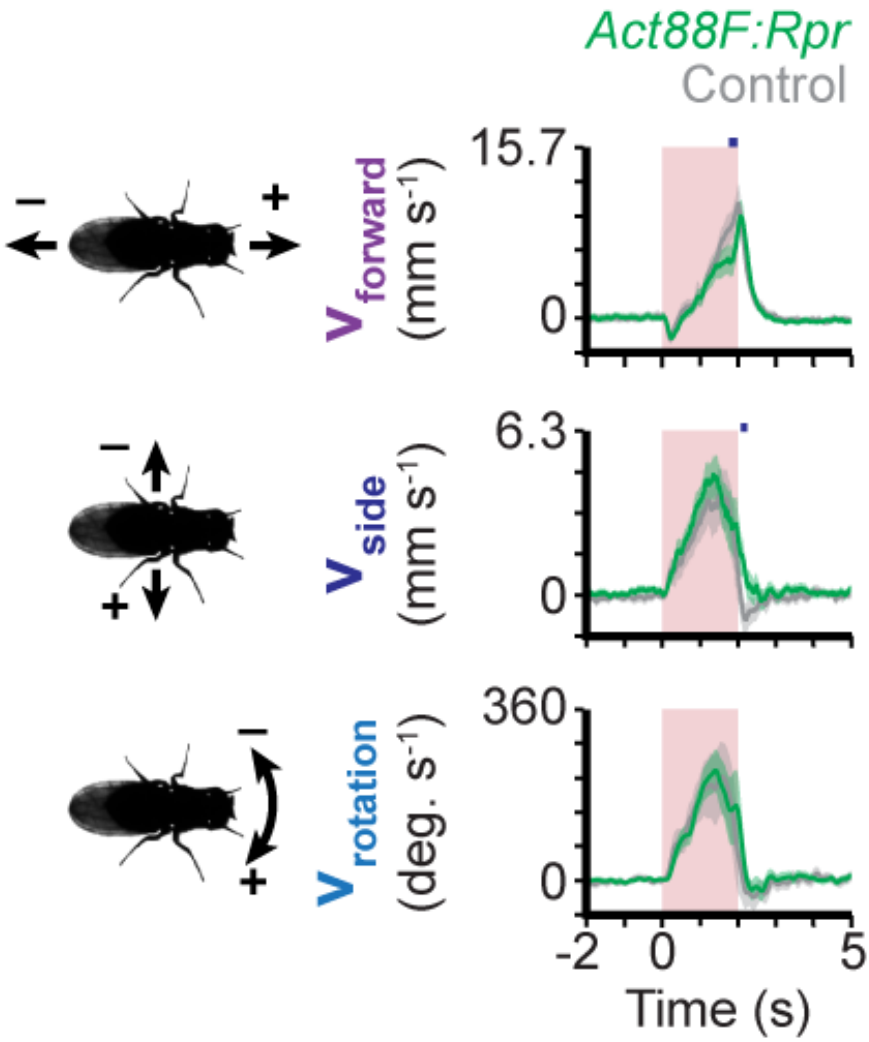

Figure S2.8: **Locomotor behaviors in control and *Act88F:Rpr* animals.** Forward, sideways, and rotational locomotor velocities for *Act88F:Rpr* (green: *Act88F:Rpr; R57C10-GAL4/UAS-GCaMP6s-p2A-tdTomato*), or control (grey: *+; R57C10-GAL4/UAS-GCaMP6s-p2A-tdTomato*) animals. Shown are the means (solid lines) and bootstrapped 95% confidence intervals (transparencies). Red transparency indicates period with near-infrared laser stimulation of the right antenna. Blue bars show time points with significant differences between Act88F:Rpr and control animals (P<0.05; Mann-Whitney U-test with a Holm-Bonferroni correction).

2.10.2 Supplementary Videos

Video S2.1: **VNC imaging volume.** 2-photon z-stack illustrating horizontal sections across the dorsal-ventral extent of the VNC. GCaMP6s (cyan) and tdTomato (red) are expressed throughout the nervous system (*GMR57C10>GCaMP6s; tdTomato*). Imaging depth indicated on the top-left.
Download Video S2.1

Video S2.2: **Horizontal VNC imaging.** 2-photon imaging of a single horizontal section of the VNC in a fly that walks and grooms. GCaMP6s (cyan) and tdTomato (red) are expressed throughout the nervous system (*GMR57C10>GCaMP6s; tdTomato*). Shown are synchronized raw fluorescence images **(top-left)**, $\%\Delta F/F$ images **(top-right)**, behavior video images (**bottom-left**), and forward, sideways, and rotational velocities of the animal (bottom-right). Experimenter-administered air puffs are indicated by the appearance of a red box above the behavior video images. Registration was performed using $\gamma = 0$ and $\lambda = 500$. Movie is 4x faster than real-time.
Download Video S2.2

Video S2.3: **Coronal VNC imaging.** 2-photon imaging of a single coronal section of the VNCin a fly that walks. GCaMP6s (cyan) and tdTomato (red) are expressed throughout the nervoussystem (*GMR57C10>GCaMP6s; tdTomato*). Shown are synchronized raw fluorescence images(**top-left**), %ΔF/F images (**top-right**), behavior video images (**bottom-left**), and forward, side-ways, and rotational velocities of the animal (**bottom-right**). Registration was performed using $\gamma = 50$ and $\lambda = 10{,}000$. Movie is 4x faster than real-time.
Download Video S2.3

Video S2.4: **Coronal cervical connective imaging.** 2-photon imaging of a single coronal section of the cervical connective in a fly that walks. GCaMP6s (cyan) and tdTomato (red) are expressed throughout the nervous system (*GMR57C10>GCaMP6s; tdTomato*). Shown are synchronized raw fluorescence images (**top-left**), $\%\Delta F/F$ images (**top-right**), behavior video images (bottom-left), and forward, sideways, and rotational velocities of the animal (bottom-right). Registration was performed using $\gamma = 50$ and $\lambda = 5{,}000$. Movie is 4x faster than real-time.
Download Video S2.4

Video S2.5: **Motion and deformation correction.** 2-photon imaging of a single horizontal section of the VNC in three different animals during behavior. Data are from **Figure 2.4** (Fly 1: λ = 500 and γ = 0; Fly 2: λ = 1000 and γ = 4; Fly 3:λ = 300 and γ = 4). Raw, unregistered, fluorescence images **(top)**. Spatially down sampled transformation vector field used to register raw images **(middle)**. Final, registered fluorescence images **(bottom)**. Images for flies 1 and 2 are 512 x 512 pixels. Images for fly 3 are 256 x 256 pixels.
Download Video S2.5

Video S2.6: **Semi-automated behavioral classification.** Illustration of semi-automated grooming and walking annotations used for **Figure 2**. Grooming and walking ROIs are shown (white boxes). The current behavioral descriptor is indicated on the top-left.
Download Video S2.6

Video S2.7: **Coronal cervical connective imaging of dorsal Moonwalker Ascending Neurons.** Moonwalker Ascending Neurons. 2-photon imaging of a single coronal section of the cervical connective in a behaving fly. GCaMP6s (cyan) and tdTomato (red) are expressed in MANs (*MAN>GCaMP6s; tdTomato*). Raw fluorescence images of the left and right dMANs are presented and outlined by ROIs **(top-left)**. These images are used to calculate $\%\Delta R/R$ traces for each neuron **(top-right)**. Corresponding behavior videography **(bottom-left)** and forward, sideways, and rotational velocities of the animal **(bottom-right)** are shown.
Download Video S2.7

Video S2.8: **Behavioral responses associated with dorsal Moonwalker Ascending Neuron activity events.** Behavioral responses associated with dorsal Moonwalker Ascending Neuron activity events. Three example behaviors (rows) produced at the onset of dMAN fluorescence events for three flies (columns). Red squares indicate the onset time of each fluorescence event (t = 0 s). Movie is 3x slower than real-time.
Download Video S2.8

Video S2.9: **Coronal cervical connective imaging of Moonwalker Descending Neurons.** 2-photon imaging of a single coronal section of the cervical connective in a behaving fly. GCaMP6s (cyan) and tdTomato (red) are expressed in MDNs (*MDN-1>GCaMP6s; tdTomato*). Raw fluorescence images of the left and right MDNs are presented and outlined by ROIs (top-left). These images are used to calculate $\%\Delta R/R$ traces for each neuron (top-right). Corresponding behavior videography (bottom-left) and forward, sideways, and rotational velocities of the animal (bottom-right) are shown.
Download Video S2.9

Video S2.10: **Behavioral responses associated with Moonwalker Descending Neuron activity events.** Descending Neuron activity events. Three example behaviors (rows) produced at the onset of MDN fluorescence events for three flies (columns). Red squares indicate the onset time of each fluorescence event (t = 0 s). Movie is 3x slower than real-time.
Download Video S2.10

Video S2.11: **Coronal cervical connective imaging of A1 neurons.** Descending Neuron activity events. 2-photon imaging of a single coronal section of the cervical connective in a behaving fly. GCaMP6s (cyan) and tdTomato (red) are expressed in A1 neurons (*A1>GCaMP6s; tdTomato*). Raw fluorescence images of the left and right A1 neurons are presented and outlined by ROIs **(top-left)**. These images are used to calculate $\%\Delta R/R$ traces for each neuron **(top-right)**. Corresponding behavior videography **(bottom-left)** and forward, sideways, and rotational velocities of the animal **(bottom-right)** are shown.
Download Video S2.11

Video S2.12: **Behavioral responses associated with left A1 neuron activity events.** Three example behaviors (rows) produced at the onset of A1 fluorescence events for three flies (columns). Red squares indicate the onset time of each fluorescence event (t = 0 s). Movie is 3x slower than real-time.
Download Video S2.12

Video S2.13: **Behavioral responses associated with right A1 neuron activity events.** Three example behaviors (rows) for each of three flies (columns) generated at the onset of right A1 neuron fluorescence events. Red squares indicate the time of each fluorescence event (t = 0s). Movie is 3x slower than real-time.
Download Video S2.13

Video S2.14: **Behaviors and neural activity in Act88F:Rpr and control animals.** Horizontal section imaging data from the prothoracic neuromere of Act88F:Rpr animals (**left**, *Act88F:Rpr/elav-GAL4; GCaMP6s/+;*), or control animals (**right**, *+/elav-GAL4; GCaMP6s/+;*). 2-photon imaging data are not registered.
Download Video S2.14

Video S2.15: **Coronal and horizontal section imaging of A1 neurons in the VNC.** A comparison of coronal **(left)**, or horizontal **(right)** section imaging of the cervical connective in the same animal. GCaMP6s (cyan) and tdTomato (red) are expressed in A1 neurons (*A1>GCaMP6s; tdTomato*). Raw fluorescence images **(top)** are presented alongside corresponding behavior video images **(bottom)**.
Download Video S2.15

Video S2.16: **Long-term behavior in animals used to image dMAN activity.** Behavior movies for three flies (*MAN>GCaMP6s; tdTomato*) during the first imaging experiment (0 min) as well as 20 min and 40 min afterwards.
Download Video S2.16

Video S2.17: **Long-term behavior in animals used to image MDN activity.** Behavior movies for three flies (*MDN>GCaMP6s; tdTomato*) during the first imaging experiment (0 min) as well as 20 min and 40 min afterwards.
Download Video S2.17

Video S2.18: **Long-term behavior in animals used to image A1 activity.** Behavior movies for three flies (*A1>GCaMP6s; tdTomato*) during the first imaging experiment (0 min) as well as 20 min and 40 min afterwards.
Download Video S2.18

3 Ascending neurons convey behavioral state to integrative sensory and action selection centers in the brain

Chin-Lin Chen, **Florian Aymanns**, Ryo Minegishi, Victor D. V. Matsuda, Nicolas Talabot, Semih Günel, Barry J. Dickson, Pavan Ramdya. "Ascending neurons convey behavioral state to integrative sensory and action selection centers in the brain", *bioRxiv*, 2022.02.09.479566, 2022. The article can be found here: https://doi.org/10.1101/2022.02.09.479566

My contribution:
My main contribution to this project is developing the analysis methodology. I developed a behavior classifier that classifies individual frames into different behavior categories based on joint angles and ball rotations (**Figure S3.1e, subsection 3.5.6**).
I then investigated how behavior and neural activity are linked using ridge regression models (**Figure 3.2, Figure 3.4a,g, Figure 3.5a,i,j, Figure 3.6a,g, Figure 3.7a,c,d, Figure S3.4a,b, subsection 3.5.8, subsection 3.5.7**).
For the confocal images, I created a template and registered all of the confocal images (**subsection 3.5.11**).
Additionally, I contributed to data curation, writing, and editing.

3.1 Abstract

Knowing one's own behavioral state has long been theorized as critical for contextualizing dynamic sensory cues and identifying appropriate future actions. Ascending neurons (ANs) in the motor system that project to the brain are well-positioned to provide such behavioral state signals. However, what ANs encode and where they convey these signals remains largely unknown. Here, through large-scale functional imaging in behaving flies, deep learning-based analyses, and morphological quantification we investigate the encoding, brain targeting, and motor patterning of 247 genetically-identifiable ANs in the adult fly, *Drosophila melanogaster*. We reveal that ANs encode high-level behaviors, specifically conveying self-motion to the AVLP, an integrative sensory hub, and discrete actions to the GNG, a locus for action selection. Additionally, AN projection patterns within the motor system are predictive of their encoding. Thus, ascending populations inform distinct brain hubs of self-motion and ongoing actions: crucial substrates for computations that give rise to adaptive behaviors.

3.2 Introduction

To generate adaptive behaviors, animals [56] and robots [153] must not only sense their environment but also be aware of their own behavioral state including low-level movements of their limbs and high-level behaviors such as walking and resting. This self-awareness has long been theorized to overcome at least two major challenges for robust, autonomous control. First, knowing if one is at rest or in motion permits the accurate interpretation of whether sensory cues, like visual motion during feature tracking or odor intensity fluctuations during plume following, result from exafference (the movement of objects in the world), or reafference (self-motion with respect to stationary objects) [56]. Second, being aware of one's current posture enables the selection of appropriate future actions that are not destabilizing, or physically impossible.

In line with these theoretical predictions, neural representations of behaviors have been observed widely across the brains of mice [154–156], and in the fly, *Drosophila melanogaster* [126, 157–159]. Furthermore, studies in *Drosophila* have supported roles for behavioral state signals in sensory contextualization (e.g., flight [126] and walking [157] modulate neurons in the visual system [158, 160]), and action selection (e.g., an animal's walking speed regulates its decision to run or freeze in response to a fear-inducing stimulus [161]).

Despite these advances, the cellular origins of behavioral state signals in the brain remain largely unknown. On one hand, they might arise from efference copies generated by descending neurons (DNs) in the brain that project to and drive downstream motor systems [56]. However, these efference copies would not be expected to provide the most precise readout of one's own behavioral state: the brain's descending commands are sculpted by musculoskeletal interactions with the environment. Instead, a more categorically and temporally precise readout of ongoing behaviors might be obtained from ascending neurons in the motor system

that process proprioceptive and tactile signals and then convey a holistic representation of behavioral states to the brain. Although these behavioral signals may come from a subset of primary mechanosensory neurons in the limbs [52], they are more likely to be computed and conveyed by second- and higher-order ascending neurons (ANs) residing in the spinal cord of vertebrates [162–165], or insect ventral nerve cord (VNC) [166, 167]. In *Drosophila*, ANs have been shown to process limb proprioceptive and tactile signals [52, 168, 169], perhaps to generate a complex readout of ongoing movements and behaviors.

To date only a few genetically-identifiable AN cell types have been studied in behaving animals—primarily in the fly, *Drosophila melanogaster*, which has a relatively small number of neurons that can also be genetically targeted for repeated investigation. These studies support the hypothesis that ANs are a prominent source of behavioral state signals in the brain. First, microscopy recordings of AN terminals in the brain have shown that Lco2N1 and Les2N1D ANs are active during walking [51], and that LAL-PS-ANs convey walking signals to the visual system [55]. Second, artificial activation of pairs of PER_{in} ANs [54], or Moonwalker ANs [53] regulate action selection and behavioral persistence, respectively.

These first insights urgently motivate a more comprehensive and quantitative analysis of large AN populations to investigate three fundamental questions. First, what information do ANs convey to the brain **(Figure 3.1A)**? They might encode low-level movements of the joints or limbs, or high-level behavioral states like whether an animal is walking, or grooming. Second, where do ANs convey this information to in the brain **(Figure 3.1B)**? They might project widely across brain regions, or narrowly target circuit hubs with specific functions. Third, what can an AN's patterning within the VNC tell us about how it derives its encoding **(Figure 3.1C, red)**? Answering these questions would open the door to a cellular-level understanding of how neurons encode behavioral states by integrating proprioceptive, tactile, and other sensory feedback signals. It would also enable the study of how behavioral state signals are used by brain circuits to contextualize multimodal cues and select appropriate future actions.

To address these questions, we developed and used a number of advanced experimental and analytical tools. First, we screened a library of split-GAL4 *Drosophila* driver lines (R.M. and B.J.D., unpublished). These, along with the published MAN-spGAL4 [53] and 12 sparsely expressing GAL4 lines [128], together allowed us to gain repeated genetic access to 247 ANs **(Figure 3.1D; Table S3.1)**. Using these driver lines and a multi-color flip-out (MCFO) approach [170], we then quantified the projections of ANs within the brain and VNC **(Figure 3.1E)**. Second, we screened the encoding of these ANs by two-photon microscopy-based functional recordings of neural activity within the VNC of tethered, behaving flies [171]. To overcome noise and movement-related deformations in imaging data, we developed and used 'AxoID', a deep learning-based software to semi-automatically identify and track axonal Regions-of-Interest (ROIs) (see Methods). Third, to precisely quantify joint angles and limb kinematics, we used a multicamera array to record behavior during two-photon imaging. We processed these videos using DeepFly3D, a deep learning-based 3D pose estimation software [37]. By combining these 3D joint positions with measured spherical treadmill rotations, a proxy for

locomotor velocities [29], we could then segment and classify behavioral time-series to study the relationship between behavioral states and ongoing neural activity using linear models.

These analyses uncovered a number of fundamental characteristics of ANs. First, as a population, ANs do not project broadly across the brain but principally target two hubs: (i) the anterior ventrolateral protocerebrum (AVLP), a site for higher-order multimodal convergence—vision [172], olfaction [173], audition [174–176], and taste [177]—, and (ii) the gnathal ganglion (GNG), a region important for action selection [54, 59, 82]. Second, ANs encode high-level behavioral states, primarily walking, rather than low-level joint or limb movements. Third, distinct behavioral states are systematically conveyed to different brain targets. The AVLP is informed of self-motion states like resting, walking, and the presence of gust-like stimuli, perhaps to contextualize sensory cues. By contrast, the GNG receives precise signals about actions—turning, eye grooming, and proboscis extension—likely to guide action selection.

To understand the relationship between AN behavioral state encoding and brain projection patterns, we then performed a more in-depth investigation of seven AN classes. We observed a correspondence between the morphology of ANs in the VNC and their behavioral state encoding: ANs with neurites targeting all three VNC neuromeres (T1-T3) encode global locomotor states (e.g., resting and walking) while those with projections only to the T1 prothoracic neuromere encoded foreleg-dependent behaviors (e.g., eye grooming). Notably, we also observed AN axons within the VNC. This suggests that ANs are not simply passive relays of behavioral state signals to the brain but that they may also help to orchestrate motor actions and/or compute state encodings. This latter possibility is illustrated by a class of 'PE-ANs' that seems to encode the number of proboscis extensions generated over tens of seconds, possibly through recurrent interconnectivity within the VNC. In summary, these data provide a first comprehensive view of ascending signals to the brain, opening the door for a cellular-level understanding of how behavioral states are computed, and how ascending motor signals enable the brain to contextualize sensory signals and select appropriate future actions.

3.3 Results

3.3.1 A large-scale screen of ascending neuron movement encoding, brain targeting, and motor system patterning

We performed a functional screen of 108 driver lines that each express fluorescent reporters in a small number of ANs **(Figure 3.1D)**. This allowed us to address to what ANs encode low-level joint and limb movements, or high-level behavioral states. To quantify limb movements, we recorded each fly using six synchronized cameras (a seventh camera was used to position the fly on the ball) **(Figure 3.1F)**. We processed these videos using DeepFly3D [37], a markerless 3D pose estimation software that outputs joint positions and angles **(Figure 3.1G)**. We also measured spherical treadmill rotations using two optic flow sensors [29] and converted these into three fly-centric velocities: forward (mm/s), sideways (mm/s), and yaw (degree/s) **(Fig-**

ure 3.1H) that correspond to forward/backward walking, side-slip, and turning, respectively. A separate DeepLabCut [39] deep neural network was used to track proboscis extensions (PEs) from one camera view **(Figure S3.1A-D)**. We studied spontaneously generated actions but also used a puff of CO_2 to elicit behaviors from sedentary animals.

Synchronized with movement quantification, we recorded the activity of ANs by performing two-photon imaging of the cervical connective within the thoracic ventral nerve cord (VNC) [171]. The VNC houses motor circuits that are functionally equivalent to those in the vertebrate spinal cord **(Figure 3.1I, left)**. Neural activity was read-out as changes in the fluorescence of a genetically-encoded calcium indicator, OpGCaMP6f, expressed in a small number of ANs. Simultaneously, we recorded tdTomato fluorescence as an anatomical fiduciary. Imaging coronal (x-z) sections of the cervical connective allowed us to keep AN axons within the imaging field-of-view despite behaviorally-induced motion artifacts that would disrupt conventional horizontal (x-y) section imaging [171]. Sparse spGAL4 and GAL4 fluorescent reporter expression facilitated axonal region-of-interest (ROI) detection. To semi-automatically segment and track AN ROIs across thousands of imaging frames, we developed and used AxoID, a deep network-based software **(Figure 3.1I, right; Figure S3.2)**. AxoID also helped perform ROI detection despite significant movement-related ROI translations and deformations as well as, for some driver lines, relatively low transgene expression levels and suboptimal imaging signal-to-noise ratios (SNR).

To relate AN neural activity with ongoing limb movements, we trained classifiers using 3D joint angles and spherical treadmill rotational velocities to accurately and automatically detect nine behaviors—forward and backward walking, spherical treadmill pushing, resting, eye and antennal grooming, foreleg and hindleg rubbing, and abdominal grooming **(Figure 3.1J)**. This classification was highly accurate **(Figure S3.1E)**. Additionally, we classified non-orthogonal, co-occurring behaviors like proboscis extensions (PEs) and recorded the timing of CO_2 puff stimuli **(Video S3-52)**.

Our final dataset encompassed 247 ANs targeted using 70 sparsely-labelled driver lines (more than 32 h of data). These data included (i) anatomical projection patterns, and temporally synchronized (ii) neural activity, (iii) joint angles, and (iv) spherical treadmill rotations. Here we focus on the results for 157 of the most active ANs taken from 50 driver lines (more than 23 h of data) **(Video S3.2)**. The remainder were excluded due to redundancy with other driver lines, a lack of neural activity, or a low SNR (as determined by smFP confocal imaging, or two-photon imaging of tdTomato and OpGCaMP6f). Representative data from each of these selected driver lines illustrate the richness of our dataset **(Video S3-52)**.

3.3.2 Ascending neurons encode high-level behaviors

With these data, we first asked to what extent AN activity encode low-level joint angles and leg movements, or high-level behaviors like walking, resting, and grooming **(Figure 3.1A)**. We expected that, unlike primary limb mechanosensory neurons, second- and higher-order ANs

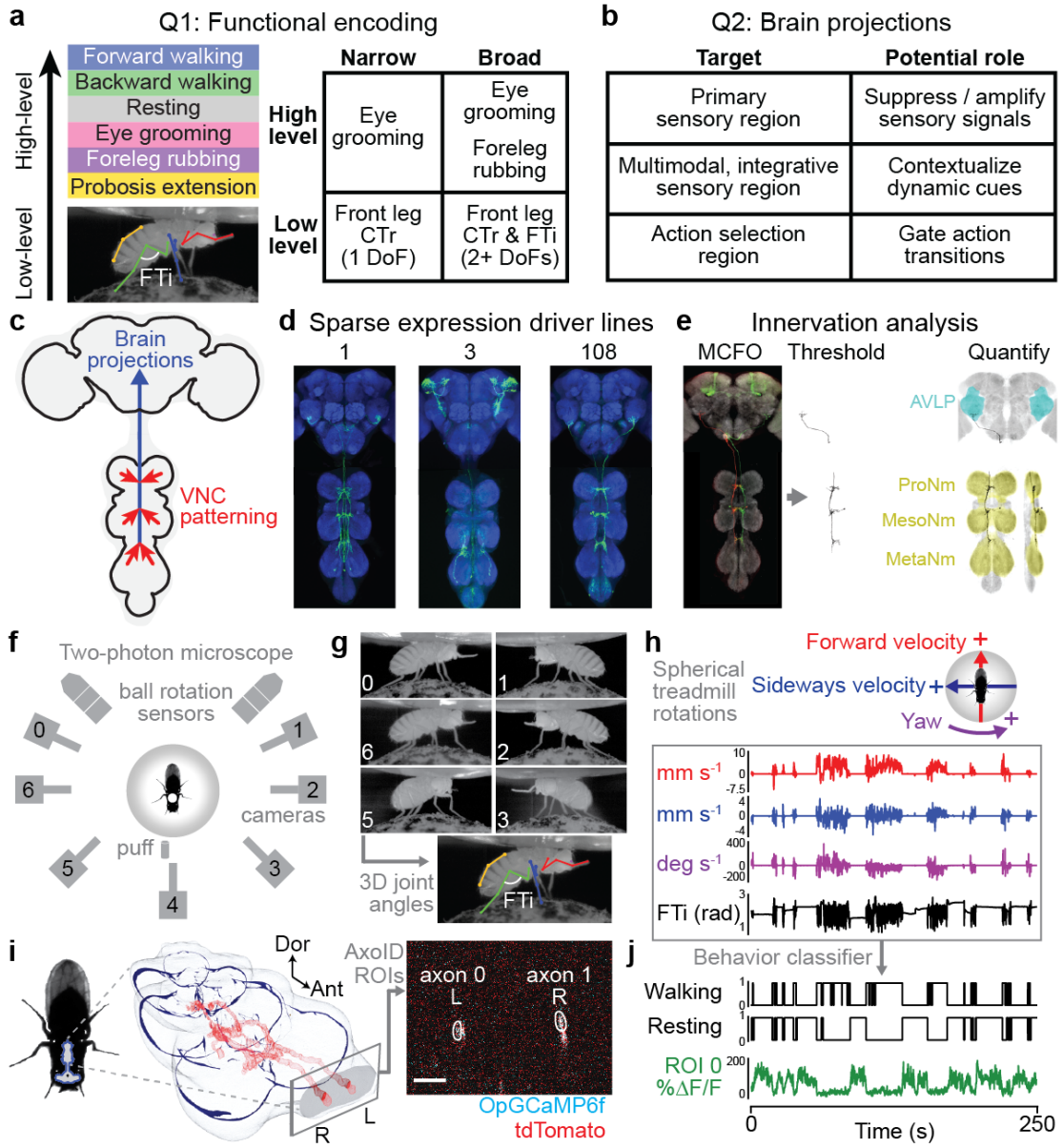

Figure 3.1: **Large-scale functional and morphological screen of ascending neuron movement encoding and nervous system targeting.** (**a-c**) Schematics and tables of the main questions addressed. (**a**) To what extent do ascending neurons (ANs) encode high-level behaviors, or low-level limb movements? This encoding may be either narrow (e.g., one behavior, or joint degree-of-freedom), or broad (e.g., several behaviors, or joint DoFs). (**b**) Where in the brain do ANs convey behavioral states? ANs might target the brain's (i) primary sensory regions (e.g., optic lobe, or antennal lobe) for sensory gain control, (ii) multimodal and integrative sensory regions (e.g., anterior ventrolateral protocerebrum, or mushroom body) to contextualize dynamic, time-varying sensory cues, and (iii) action selection centers (e.g., gnathal ganglion, or central complex) to gate action transitions. Individual ANs may project broadly to multiple

brain regions, or narrowly to one region. (**C**) To what extent is an AN's patterning within the VNC predictive of its brain targeting and encoding? (**d**) We screened 108 sparsely expressing driver lines. The projection patterns of the lines with active ANs and high signal-to-noise ratio (157 ANs) were examined in the brain and VNC. (**e**) These were quantified by tracing single-cell MCFO confocal images. **(f)** Overhead schematic of the behavior measurement system used during two-photon microscopy. A camera array captures six views of the animal. Two optic flow sensors measure ball rotations. A puff of CO_2 (or air) is used to elicit behavior from sedentary animals. (**g**) 2D poses are estimated for six camera views using DeepFly3D. These data are triangulated to quantify 3D poses and joint angles for six legs and the abdomen (color-coded). The Femur-Tibia (FTi) joint angle is indicated (white). (**h**) Two optic flow sensors measure rotations of the spherical treadmill as a proxy for forward (red), sideways (blue), and yaw (purple) walking velocities. Positive directions of rotation ('+') are indicated. (**i, left**) A volumetric representation of the ventral nerve cord (VNC) including a reconstruction of ANs targeted by the *SS27485-spGAL4* driver line (red). Indicated are the dorsal-ventral ('Dor') and anterior-posterior ('Ant') axes, as well as the fly's left (L) and right (R) sides. (**i, right**) Sample two-photon cross-section image of the thoracic neck connective showing ANs that express OpGCaMP6f (cyan) and tdTomato (red). AxoID is used to semi-automatically identify two axonal regions-of-interest (ROIs, white) on the left ('L') and right ('R') sides of the connective. (**j**) Spherical treadmill rotations and joint angles are used to classify behaviors. Binary classifications are then compared with simultaneously recorded neural activity for 250 s trials of spontaneous and puff-elicited behaviors. Shown is an activity trace from ROI 0 (green) in panel **i**.
See also **Figure S3.1**, and **Video S3.1**, **Video S3.2**, and **Video S3-52**.

---

would more likely integrate and process proprioceptive and tactile sensory signals to encode high-level behavioral states. This remained unknown because previous studies of AN encoding [51, 54, 55] did not quantify movements at high enough resolution, or study more than a few ANs in total. To address this gap, with the data from our large-scale functional screen, we performed a linear regression analysis to quantify the degree to which the movements of individual joints, legs, pairs of legs, or epochs of high-level behaviors could explain the time-course of AN activity. Specifically, we quantified the unique explained variance (UEV, or ΔR^2) for each movement, or behavioral regressor via cross-validation by subtracting a reduced model R^2 from a full regression model R^2. Specifically, in the reduced model, the regressor of interest was shuffled while keeping the other regressors intact (see Methods). To compensate for the temporal mismatch between fast leg movements and slower calcium signal decay dynamics, every joint angle and behavioral state regressor was convolved with a calcium indicator decay kernel chosen to maximize the explained variance in neural activity, with the aim of reducing the occurrence of false negatives.

First we examined to what extent individual joint angles could explain the activities of 157 ANs. We confirmed that the vast majority of joint angles do not covary with others—with the exception of the middle and hindleg CTr and FTi pitch angles, which were correlated to one another **(Figure S3.3)**. This is important because if two regressors are highly correlated, one

regressor can compensate when shuffling the other, resulting in a potential false negative. We did not find any evidence of joint angles explaining AN activity **(Figure 3.2A)**. Similarly, individual leg movements (tested by shuffling all of the joint angle regressors for a given leg) could not explain the variance of AN activity **(Figure 3.2B)**. Additionally, with the exception of ANs from SS25469 whose activities could be explained by movements of the front legs **(Figure 3.2C)**, AN activity largely could not be explained by the movements of pairs of legs. By contrast, the activity of ANs could be explained by high-level behavioral states **(Figure 3.2D)**. Most ANs encoded self-motion—forward walking and resting—but some also encoded specific actions like eye grooming, proboscis extensions, as well as responses to puff stimuli.

Our regression approach is inherently conservative: it avoids false positives. However, because it is thus prone to false negatives for infrequently occurring behaviors like abdominal grooming and hindleg rubbing, as an additional alternative approach, we measured the mean normalized $\Delta F/F$ of each AN for each high-level behavioral state. Using this complementary approach, we confirmed and extended our results **(Figure S3.4A)**. We considered results from both our linear regression and mean normalized $\Delta F/F$ analyses when selecting neurons for further in-depth analyses.

3.3.3 Ascending neurons target integrative sensory, or action selection brain regions as a function of their encoding

Having identified high-level behavioral state encoding for a large population of 157 ANs, we next wondered to what extent these distinct state signals are routed to specific and distinct brain targets **(Figure 3.1B)**. On one hand, individual ANs might project diffusely to multiple brain regions. Alternatively, they might target one, or only a few regions. For instance, locomotor signals carried by walking and resting encoding ANs might be conveyed to brain regions to contextualize time-varying visual and olfactory cues with respect to an animal's own self-motion. On the other hand, ANs that signal when an animal is grooming might target action selection brain regions to prohibit future actions that might result in unstable postures. To address these possibilities, we quantified the brain projections of all 157 ANs by staining and imaging the expression of spFP and MCFO reporters in these neurons **(Figure 3.1E)**.

Strikingly, we found that AN projections to the brain were largely restricted to two regions: the AVLP, a site known for multimodal, integrative sensory processing [172–177] and the GNG, a hub for action selection [54, 59, 82] **(Figure 3.3A)**. ANs encoding resting and puff-responses almost exclusively target the AVLP **(Figure S3.4B,C)** providing a means for interpreting whether sensory cues arise from self-motion, or the movement of objects in the external environment (i.e., while resting, an animal might still perceive visual motion due to moving objects, or odor fluctuations due to gust-like puffs of air). By contrast, the GNG is targeted by ANs encoding a wide variety of behavioral states including walking, eye grooming, and proboscis extensions **(Figure S3.4B,C)**. These signals may help to ensure that future actions are compatible with ongoing ones.

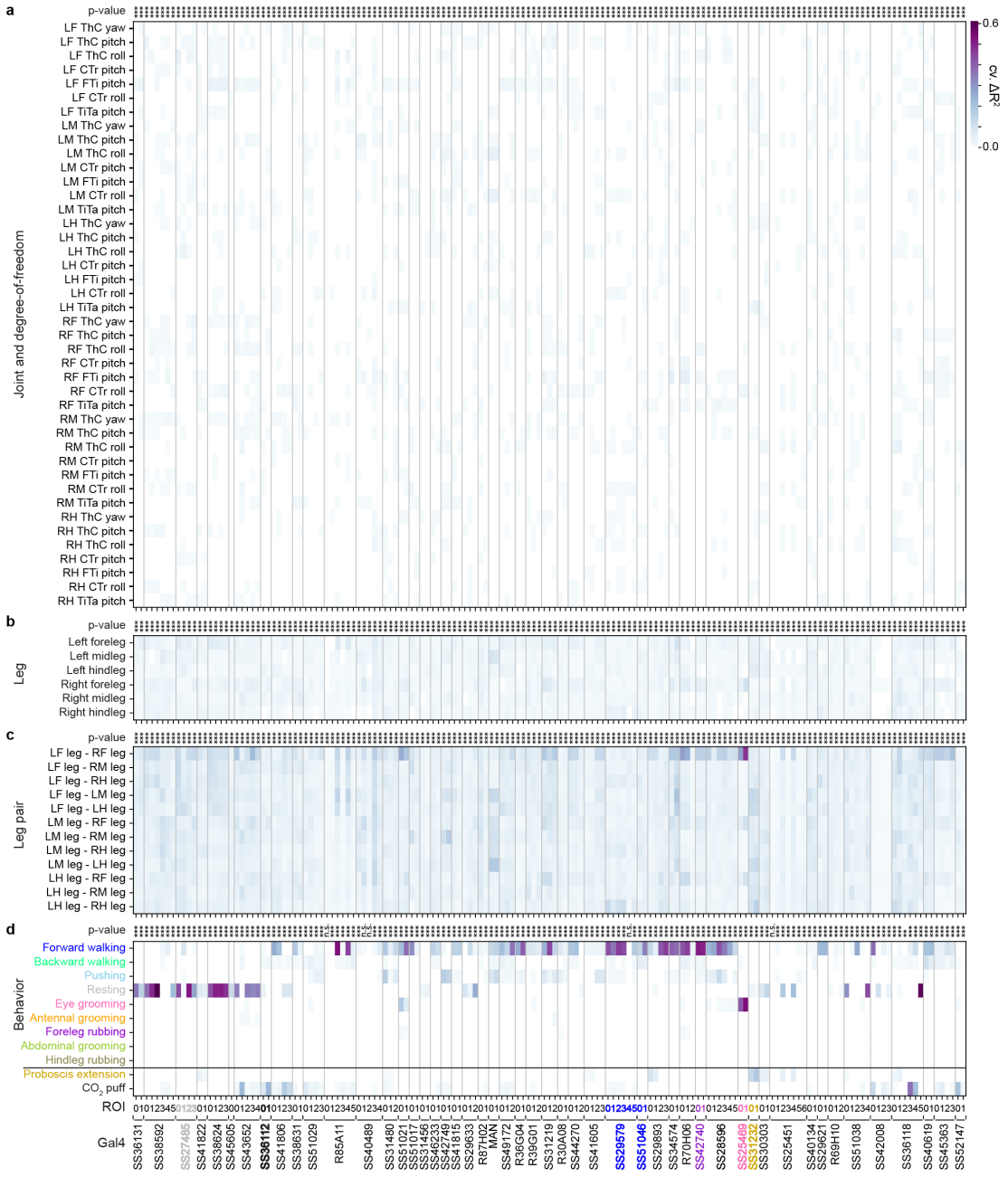

Figure 3.2: **Ascending neurons encode high-level behaviors.** Proportion of variance in AN activity that is uniquely explained by regressors (cross-validated ΔR^2) based on (**a**) joint movements, (**b**) the movements of individual legs, (**c**) the movements of pairs of legs, (**d**) high-level behaviors. Regression analyses were performed for 157 ANs recorded from 50 driver lines. Lines selected for more in-depth analysis are color-coded by the behavioral class best explaining their neural activity: SS27485 (resting), SS36112 (puff responses), SS29579 (walking), SS51046 (turning), SS42740 (foreleg movements), SS25469 (eye grooming), and SS31232 (proboscis extensions). Non-orthogonal regressors (PE and CO_2 puffs) are separated

from the others. p-values report the F-statistic of overall significance of the complete regression model with none of the regressors shuffled (*p<0.05, **p<0.01, and ***p<0.001).
See also **Figure S3.3** and **Figure S3.4**.

Because AN dendrites and axons within the VNC might be used to compute behavioral state encodings, we next asked to what extent their projection patterns within the VNC are predictive of an AN's encoding. For example, ANs encoding resting might require sampling each VNC leg neuromere (T1, T2, and T3) to confirm that every leg is inactive. By quantifying AN projections within the VNC **(Figure 3.3B)**, we found that, indeed, an AN's VNC projection pattern can be predictive of behavioral state encoding. As hypothesized, ANs encoding resting (e.g., SS27485) each project to all VNC leg neuromeres **(Figure S3.4B,D)**. By contrast, ANs encoding foreleg-dependent eye grooming (SS25469) only project within the T1 VNC neuromere housing motor circuits to control the front legs **(Figure S3.4B,D)**. To more deeply understand how the morphological features of ANs relate to behavioral state encoding, we next performed a detailed study of a diverse subset of ANs that encode resting, puff-responses, walking, turning, foreleg-dependent behaviors, eye grooming, and proboscis extensions.

3.3.4 Distinct rest- and puff-encoding by morphologically similar ANs

AN classes that encode resting and puff responses exhibit coarsely similar projection patterns: both almost exclusively target the brain's AVLP while also both sampling from all three VNC leg neuromeres (T1-T3) **(Figure S3.4)**. We therefore next investigated which more detailed morphological features might be predictive of their very distinct encoding.

We addressed this question by closely examining the functional and morphological properties of specific pairs of 'rest-ANs' (SS27485) and 'puff-ANs' (SS36112). Neural activity traces of rest-ANs and puff-ANs could be reliably predicted by regressors for resting **(Figure 3.4A)**, and puff-stimuli **(Figure 3.4G)**, respectively. This was statistically confirmed by comparing behavior-triggered averages of AN responses at the onset of resting **(Figure 3.4B)**, versus puff stimulation **(Figure 3.4H)**, respectively. Importantly, although CO_2 puffs frequently elicited brief periods of backward walking, close analysis revealed that puff-ANs primarily respond to gust-like puffs and do not encode backward walking **(Figure S3.5A-D)**. They also did not encode responses to CO_2 specifically: the same neurons responded equally well to puffs of air **(Figure S3.5E-M)**.

As mentioned, rest- and puff-ANs, despite their very distinct encoding, exhibit similar innervation patterns in the brain and VNC. However, MCFO-based single neuron analysis revealed a few subtle but potentially important differences. First, rest- and puff-AN cell bodies are located in the T2 **(Figure 3.4C)** and T3 **(Figure 3.4I)** neuromeres, respectively. Second, although both AN classes project medially into all three leg neuromeres (T1-T3), rest-ANs have a simpler morphology **(Figure 3.4D)** compared with the more complex arborization of puff-ANs in the

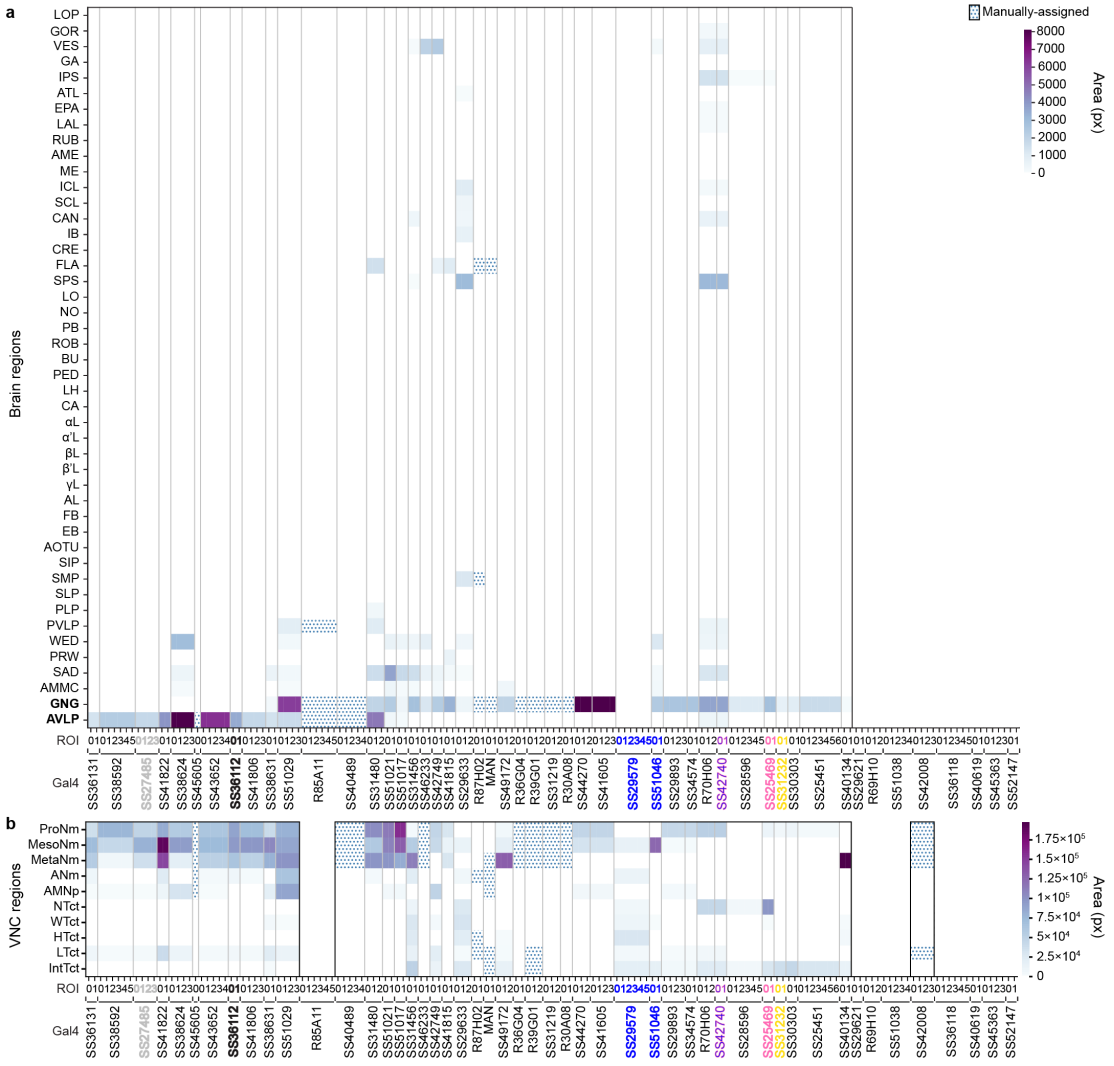

Figure 3.3: **Ascending neurons principally project to the brain's AVLP and GNG and the VNC's leg neuromeres.** Regional innervation of (**a**) the brain, or (**b**) the VNC. Data are for 157 ANs recorded from 50 driver lines and automatically quantified through pixel-based analyses of MCFO labeled confocal images. Other, manually quantified driver lines are indicated (dotted). Lines for which projections could not be unambiguously identified are left blank. Lines selected for more in-depth evaluation are color-coded by the behavioral state that best explains their neural activity: SS27485 (resting), SS36112 (puff responses), SS29579 (walking), SS51046 (turning), SS42740 (foreleg-dependent behaviors), SS25469 (eye grooming), and SS31232 (proboscis extensions).
See also **Figure S3.4**.

VNC **(Figure 3.4J)**. In the brain, both AN types project to nearly the same ventral region of the AVLP where they exhibit varicose terminals **(Figure 3.4E** and **Figure 3.4K)**. Using *syt:GFP*, a GFP tagged synaptotagmin (presynaptic) marker, we confirmed that these varicosities house

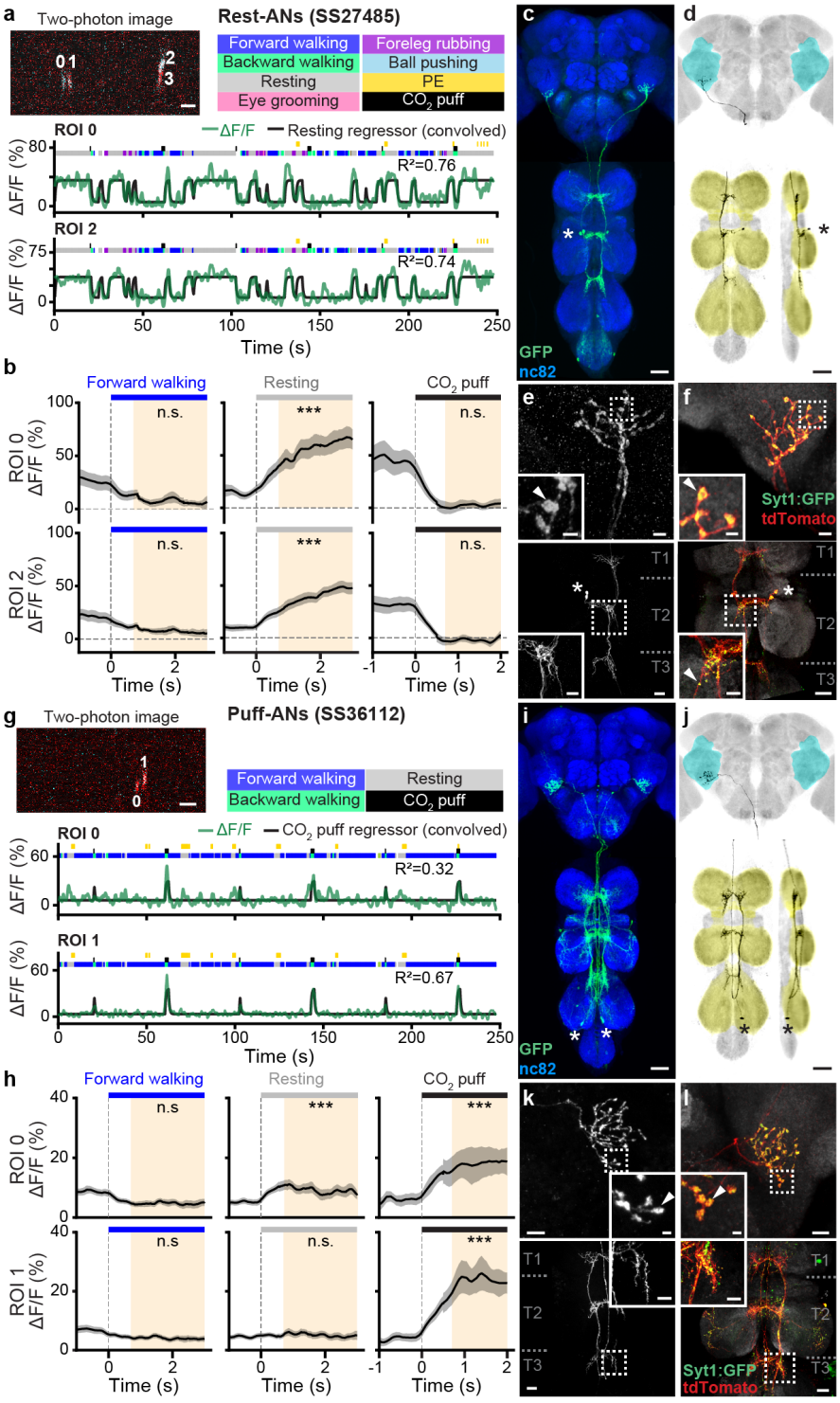

Figure 3.4: **Functional and anatomical properties of ascending neurons that encode resting, or responses to puffs.** (**a**,**g**) (**top-left**) Two-photon image of axons from an (**a**) *SS27485-GAL4*, or (**g**) *SS36112-GAL4* animal expressing OpGCaMP6f (cyan) and tdTomato (red). ROIs are

numbered. Scale bars are 5µm. **(bottom)** Behavioral epochs are color-coded. Representative $\Delta F/F$ time-series from two ROIs (green) overlaid with a prediction (black) obtained by convolving (**a**) resting epochs, or (**g**) puff stimuli with Ca^{2+} indicator response functions. Explained variances are indicated (R^2). (**b,h**) Mean (solid line) and 95% confidence interval (gray shading) of $\Delta F/F$ traces for (**b**) rest-ANs, or (**h**) puff-ANs during epochs of forward walking (left), resting (middle), or CO_2 puffs (right). 0 s indicates the start of each epoch. Here and for similar analyses in later figures, data more than 0.7s after onset (yellow region) are compared with an otsu thresholded baseline (ANOVA and Tukey posthoc comparison, $^{***}p<0.001$, $^{**}p<0.01$, $^{*}p<0.05$, n.s. not significant). (**c,i**) Standard deviation projection image of an (**c**) *SS27485-GAL4*, or **(i)** *SS36112-GAL4* nervous system expressing smFP and stained for GFP (green) and Nc82 (blue). Cell bodies are indicated (white asterisk). Scale bars are 40 µm. (**d,j**) Projection as in **c,i** but for one MCFO-expressing, traced neuron (black asterisk). The brain's AVLP (cyan) and VNC's leg neuromeres (yellow) are color-coded. Scale bars are 40 µm. (**e,f,k,l**) Higher magnification projections of (**top**) brains and (**bottom**) VNCs from (**e,f**) *SS27485-GAL4*, or (**k,l**) *SS36112-GAL4* animals expressing (**e,k**) the stochastic label MCFO, or (**f,l**) the synaptic marker, syt:GFP (green), and tdTomato (red). Insets magnify dashed boxes. Indicated are cell bodies (asterisks), bouton-like structures (white arrowheads), and VNC leg neuromeres ('T1, T2, T3'). Scale bars for brain images and insets are (**e**) 5 µm or (**k**) 10 µm, and 2 µm for insets. Scale bars for VNC images and insets are 20 µm and 10 µm, respectively.
See also **Figure S3.5**.

---

synaptic terminals **(Figure 3.4F, top** and **Figure 3.4L, top)**. Notably, in addition to smooth, likely dendritic arbors, both AN classes have axon terminals within the VNC **(Figure 3.4F, bottom** and **Figure 3.4L, bottom)**.

Taken together, these results demonstrate that even very subtle differences in VNC patterning can give rise to dramatically different AN tuning properties. In the case of rest- and puff-ANs, we speculate that this might be due to physically close, but distinct presynaptic partners—possibly leg proprioceptive afferents for rest-ANs, and leg tactile afferents for puff-ANs.

3.3.5 Walk- or turn- encoding correlates with the laterality of VNC projections

Among the ANs we analyzed, most encode walking **(Figure 3.2D)**. However, this broad category of locomotion includes more subtle dimensions including walking direction and turning. We reasoned that an AN's patterning within the VNC may be predictive of whether it encodes locomotion broadly (e.g., walking), or narrowly (e.g., turning).

Indeed, we observed that while the activity of one pair of ANs (SS29579, 'walk-ANs') was remarkably well explained by the timing and onset of walking epochs **(Figure 3.5A-C)**, for other ANs a broad walking regressor could account for much less of the variance in neural activity **(Figure 3.2D)**. We reasoned that these ANs might instead encode narrower locomotor dimensions like turning. For a bilateral pair of DNa01 descending neurons, their difference in

activity correlates with turning direction [89, 171]. To see if this relationship might also hold for some pairs of walk-encoding ANs, we quantified the degree to which the difference in pairwise activity can be explained by spherical treadmill yaw or roll velocity—a proxy for turning **(Figure 3.5H)**. Indeed, we found several pairs of ANs for which turning explained a relatively large amount of variance. For one pair of 'turn-ANs' (SS51046), although a combination of forward and backward walking regressors poorly predicted neural activity **(Figure 3.5I)**, a regressor based on spherical treadmill roll velocity strongly predicted the pairwise difference in neural activity **(Figure 3.5J)**. When an animal turned right, the right (ipsilateral) turn-AN was more active, and the left turn-AN was more active during left turns **(Figure 3.5K)**. During forward walking, both turn-ANs were active **(Figure 3.5L)**.

We next asked how VNC patterning might predict this distinction between broad (walk-ANs) versus narrow (turn-ANs) locomotor encoding. Both AN classes have cell bodies in the VNC's T2 neuromere **(Figure 3.5D,M)**. However, walk-ANs bilaterally innervate the T2 neuromere **(Figure 3.5E)**, whereas turn-ANs unilaterally innervate T1 and T2 **(Figure 3.5N, black)**. Their ipsilateral T2 projections are smooth and likely dendritic **(Figure 3.5O_1,P_1)**, while their contralateral T1 projections are varicose and exhibit syt:GFP puncta, suggesting that they harbor presynaptic terminals **(Figure 3.5O_2,P_2)**. Both walk-ANs **(Figure 3.5D,E)** and turn-ANs **(Figure 3.5M,N)** project to the brain's GNG. However, only turn-ANs project to the WED (**Figure 3.5N**). Notably, walk-AN terminals in the brain **(Figure 3.5F)** are not labelled by *syt:GFP* **(Figure 3.5G)**, suggesting that they may be neuromodulatory in nature, relying on another class of synaptic machinery [178].

These data support the notion that broad versus narrow AN behavioral state encoding may depend on the laterality of VNC patterning. Additionally, we observed that pairs of broadly-tuned walk-ANs that bilaterally innervate the VNC are synchronously active. By contrast, pairs of narrowly-tuned turn-ANs are asynchronously active. This correlation between the bilaterality of an AN pair's VNC projections and their synchrony appears to be a general principle **(Figure S3.6)**.

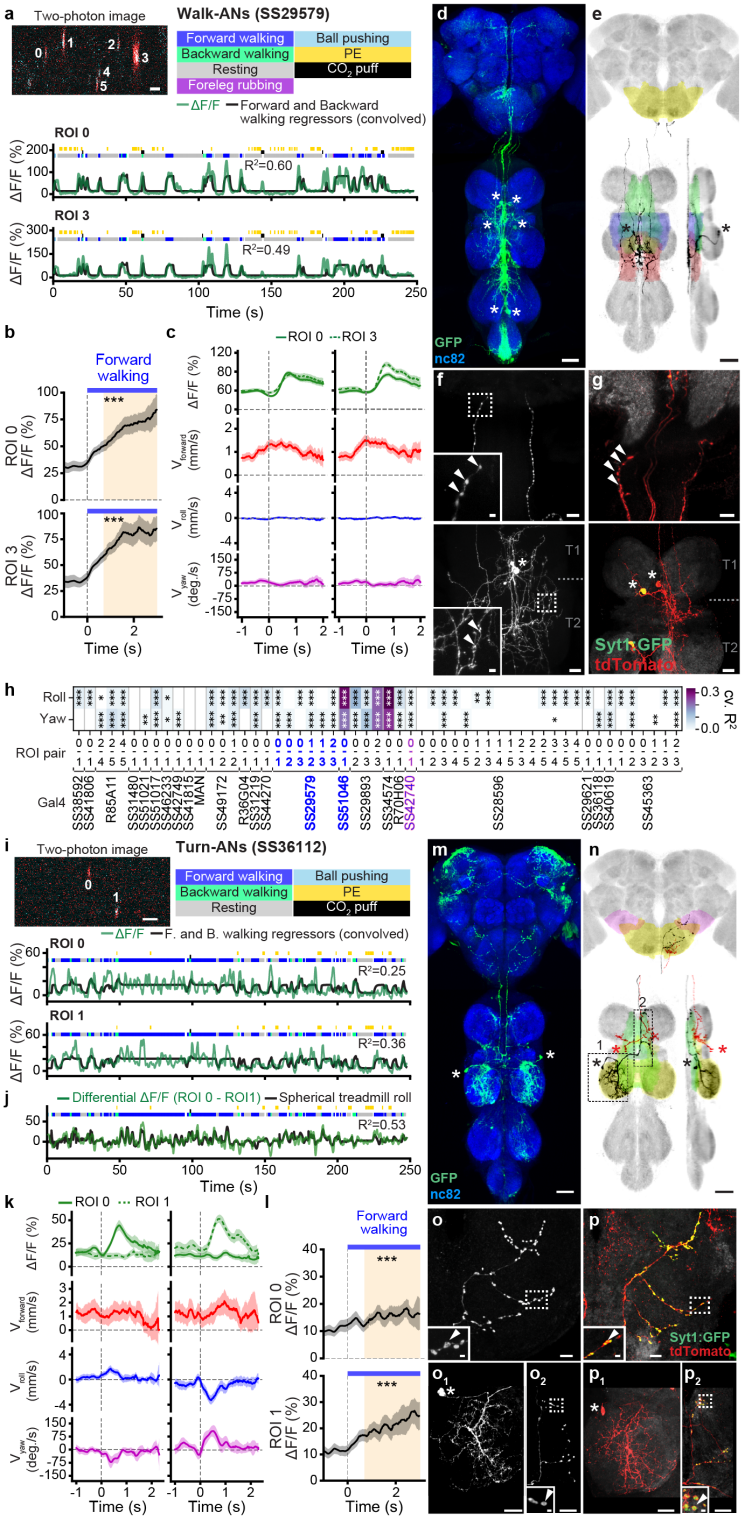

Figure 3.5: **Functional and anatomical properties of ascending neurons that encode walking, or turning.** (**a, i**) (**top-left**) Two-photon image of axons from an (**a**) *SS29579-GAL4*, or (**i**) *SS51046-GAL4* animal expressing OpGCaMP6f (cyan) and tdTomato (red). ROIs are numbered.

Scale bars are 5 μm. **(bottom)** Behavioral epochs are color-coded. Representative $\Delta F/F$ time-series from two ROIs (green) overlaid with a prediction (black) obtained by convolving forward and backward walking epochs with Ca^{2+} indicator response functions. Explained variance is indicated (R^2). (**b,l**) Mean (solid line) and 95% confidence interval (gray shading) of $\Delta F/F$ traces during epochs of forward walking. 0 s indicates the start of each epoch. (**c,k**) Fluorescence (OpGCaMP6f) event-triggered average ball rotations for **(left)** ROI 0, or **(right)** ROI 3 of (**c**) an *SS29579-GAL4* animal, or (**left**) ROI 0, or (**right**) ROI 1 of (**k**) an *SS51046-GAL4* animal. Fluorescence events are time-locked to 0 s (green). Shown are mean and 95% confidence intervals for forward (red), roll (blue), and yaw (purple) ball rotational velocities. (**d,m**) Standard deviation projection image for a (**d**) *SS29579-GAL4*, or (**m**) SS51046 nervous system expressing smFP and stained for GFP (green) and Nc82 (blue). Cell bodies are indicated (white asterisks). Scale bar is 40 μm. (**e,n**) Projection as in **d, m** but for one MCFO-expressing, traced neuron (black asterisks). The brain's GNG (yellow) and WED(pink), and VNC's intermediate (green), wing (blue), haltere (red) tectulum and mesothoracic leg neuromere (yellow) are color-coded. Scale bar is 40 μm. (**f,g,o,p**) Higher magnification projections of (**top**) brains and (**bottom**) VNCs of (**f,g**) *SS29579-GAL4* animals, or (**o,p**) *SS51046-GAL4* animals expressing (**f,o**) the stochastic label MCFO, or (**g,p**) the synaptic marker, *syt:GFP* (green), and tdTomato (red). Insets magnify dashed boxes. Indicated are cell bodies (asterisks), bouton-like structures (white arrowheads), and VNC leg neuromeres ('T1, T2'). $\mathbf{o_1}$ and $\mathbf{p_1}$, or $\mathbf{o_2}$ and $\mathbf{p_2}$ correspond to the locations '1' and '2' in **(n)**. Scale bars for brain images and insets are 10 μm and 2 μm, respectively. Scale bars for VNC images and insets are 20 μm and 4 μm, respectively.
See also **Figure S3.6**.

3.3.6 Foreleg-dependent actions are encoded by ANs in the anterior VNC

In addition to locomotion, flies use their forelegs to perform complex movements including reaching, boxing, courtship tapping, and several kinds of grooming—eye grooming, antennal grooming, and foreleg rubbing. An ongoing awareness of these behavioral states is critical to select appropriate future actions that do not lead to unstable postures. For example, before deciding to groom its hindlegs, an animal must first confirm that its forelegs are stably on the ground and not also grooming.

We noted that some ANs project only to the VNC's anterior-most, T1 leg neuromere **(Figure S3.4D)**. This pattern implies a potential role in encoding actions that only depend on the forelegs. Indeed, close examination revealed two classes of ANs that encode foreleg-related behaviors. We found ANs (SS42740) that were broadly active during multiple foreleg-dependent behaviors including walking, pushing, and grooming ('foreleg-ANs'; overlaps with R70H06) **(Figure S3.4A)(Figure 3.6A,B)**. By contrast, another pair of ANs (SS25469) was narrowly tuned and sometimes asynchronously active only during eye grooming ('eye groom-ANs') **(Figure S3.4A,B) (Figure 3.6G,H)**. Similar to walking and turning, we hypothesized that this broad (foreleg) versus narrow (eye groom) behavioral encoding might be reflected by a difference in the promiscuity and laterality of AN innervations in the VNC.

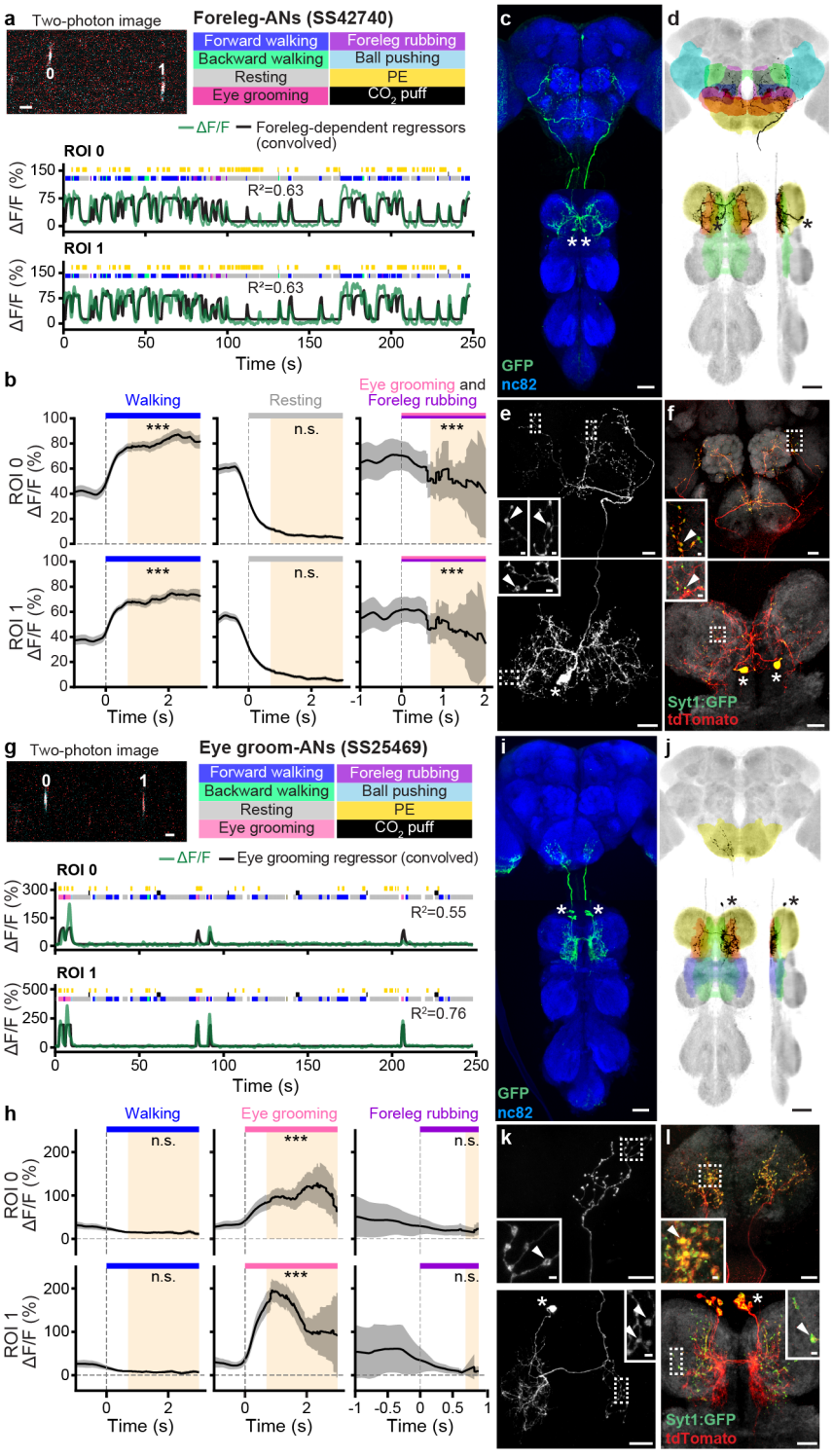

Figure 3.6: **Functional and anatomical properties of ascending neurons that encode multiple foreleg behaviors, or only eye grooming.** (**a**,**g**) (**top-left**) Two-photon image of axons from an (**a**) *SS42740-GAL4*, or (**g**) *SS25469-GAL4* animal expressing OpGCaMP6f (cyan) and tdTomato (red). ROIs are numbered. Scale bar is 5 µm. **(bottom)** Behavioral epochs are color-coded.

Representative $\Delta F/F$ time-series from two ROIs (green) overlaid with a prediction (black) obtained by convolving all foreleg-dependent behavioral epochs (forward and backward walking as well as eye, antennal, and foreleg grooming) for an (**a**) *SS42740-GAL4* animal, or eye grooming epochs for an (**g**) *SS25469-GAL4* animal with Ca^{2+} indicator response functions. Explained variance is indicated (R^2). (**b,h**) Mean (solid line) and 95% confidence interval (gray shading) of $\Delta F/F$ traces for (**b**) foreleg-ANs during epochs of forward walking (left), resting (middle), or eye grooming and foreleg rubbing (right), or (**h**) eye groom-ANs during forward walking, eye grooming, or foreleg rubbing epochs. 0 s indicates *the start of eac*h epoch. (**c,i**) Standard deviation projection image for an (**c**) *SS42740-GAL4*, or (**i**) *SS27485-GAL4* nervous system expressing smFP and stained for GFP (green) and Nc82 (blue). Cell bodies are indicated (white asterisks). Scale bars are 40 μm. (**d,j**) Projections as in **c,i** but for one MCFO-expressing, traced neuron (black asterisks). The brain's GNG (yellow), AVLP (cyan), SAD (green), VES (pink), IPS (blue), SPS (orange), and VNC's neck (orange), intermediate tectulum (green), wing tectulum (blue), and prothoracic leg neuromere (yellow) are color-coded. Scale bars are 40 μm. (**e,f,k,l**) Higher magnification projections of **(top)** brains and (**bottom**) VNCs from (**e,f**) *SS42740-GAL4*, or (**k,l**) *SS25469-GAL4* animals expressing (**e,k**) the stochastic label MCFO, or (**f,l**) the synaptic marker, syt:GFP (green), and tdTomato (red). Insets magnify dashed boxes. Indicated are cell bodies (asterisks), and bouton-like structures (white arrowheads). Scale bars for brain images and insets are 20 μm and 2 μm, respectively. Scale bars for VNC images and insets are 20 μm and 2 μm, respectively.
See also **Figure S3.4**.

To test this hypothesis, we compared the morphologies of foreleg- and eye groom-ANs. Both had cell bodies in the T1 neuromere, although foreleg-ANs were posterior **(Figure 3.6C)** and eye groom-ANs were anterior **(Figure 3.6I)**. Foreleg- and eye groom-ANs also both projected to the dorsal T1 neuromere with eye groom-AN neurites restricted to the tectulum **(Figure 3.6D** and **Figure 3.6J)**. Notably, foreleg-AN puncta **(Figure 3.6E, bottom)** and syt:GFP expression **(Figure 3.6F, bottom)** were bilateral and diffuse while eye groom-AN puncta **(Figure 3.6K, bottom)** and syt:GFP expression **(Figure 3.6L, bottom)** were largely restricted to the contralateral T1 neuromere. Projections to the brain paralleled this difference in VNC projection promiscuity: foreleg-ANs terminated across multiple brain areas—GNG, AVLP, SAD, VES, IPS, and SPS **(Figure 3.6E,F top)**— while eye groom-ANs narrowly targeted the GNG **(Figure 3.6K,L top)**.

These results further illustrate how an AN's encoding relates to its VNC patterning. Here, diffuse, bilateral projections are associated with encoding multiple behaviors that require foreleg movements whereas focal, unilateral projections are related to a narrow encoding of eye grooming.

3.3.7 Temporal integration of proboscis extensions by a cluster of ANs

Flies often generate spontaneous proboscis extensions (PEs) while resting **(Figure 3.7A, yellow ticks)**. We observed that 'PE-ANs' (SS31232, overlap with SS30303) **(Figure 3.2D)** become active

during PE trains—a sequence of PEs that occur within a short period of time **(Figure 3.7A)**. Close examination revealed that PE-AN activity slowly ramped up over the course of PE trains. This made them difficult to model using a simple PE regressor: their activity levels were lower than predicted early in PE trains, and higher than predicted late in PE trains. On average, across many PE trains, PE-AN activity reached a plateau by the seventh PE **(Figure 3.7B)**.

Thus, PE-AN activity seemed to convey the temporal integration (or counting) of discrete events [179, 180]. Therefore, we next asked if PE-AN activity might be better predicted using a regressor that integrates the number of PEs within a given time window. By testing a range of window sizes, we determined that the most accurate prediction of PE-AN dynamics could be obtained using an integration window of more than 10 s **(Figure 3.7C, red circles)**. This additional integration window made it possible to predict both the undershoot and overshoot of PE-AN activity at the start and end of PE trains, respectively **(Figure 3.7D)**.

Temporal integration can be implemented using a line attractor model [181, 182] based on recurrently connected circuits. To explore the degree to which PE-AN might support an integration of PE events via recurrent interconnectivity, we examined PE-AN morphologies more closely. PE-AN cell bodies were located in the anterior T1 neuromere **(Figure 3.7E)**. From there they projected dense neurites into the midline of the T1 neuromere **(Figure 3.7F)**. Among these neurites in the VNC, we observed puncta and syt:GFP expression consistent with presynaptic terminals **(Figure 3.7G,H, bottom)**. Their dense and highly overlapping arbors would be consistent with a mutual interconnectivity between PE-ANs. These putative recurrent connections could enable the integration of PE events over tens-of-seconds. This integration might be used to filter out sparse PE events associated with feeding and allow PE-ANs to convey long PE trains observed during deep rest-states [183] to the brain's GNG **(Figure 3.7G,H, top)**.

3.4 Discussion

Animals must be aware of their own behavioral states to accurately interpret sensory cues and select appropriate future actions. Here, we examined how this self-awareness might be conveyed to the brain by studying the activity and targeting of ascending neurons residing within the *Drosophila* motor system. Specifically, we addressed a number of fundamental questions **(Figure 3.1A-C)**. First, to what extent do ANs encode the low-level movements of joint and legs, versus high-level behavioral states like walking and grooming? Second, are individual AN encodings narrow (conveying one movement or behavior), or broad (conveying multiple movements or behaviors)? Third, to what extent do ANs target multiple or single brain regions? Fourth, do ANs that convey distinct signals also target distinct brain regions? Fifth, which characteristics of an AN's patterning in the VNC are predictive of their encoding? Sixth, are ANs a simple feedforward relay of signals to the brain, or might they also contribute to computations within the VNC? To address these questions, we performed a large-scale functional and anatomical screen that leveraged a library of *Drosophila* driver lines expressing

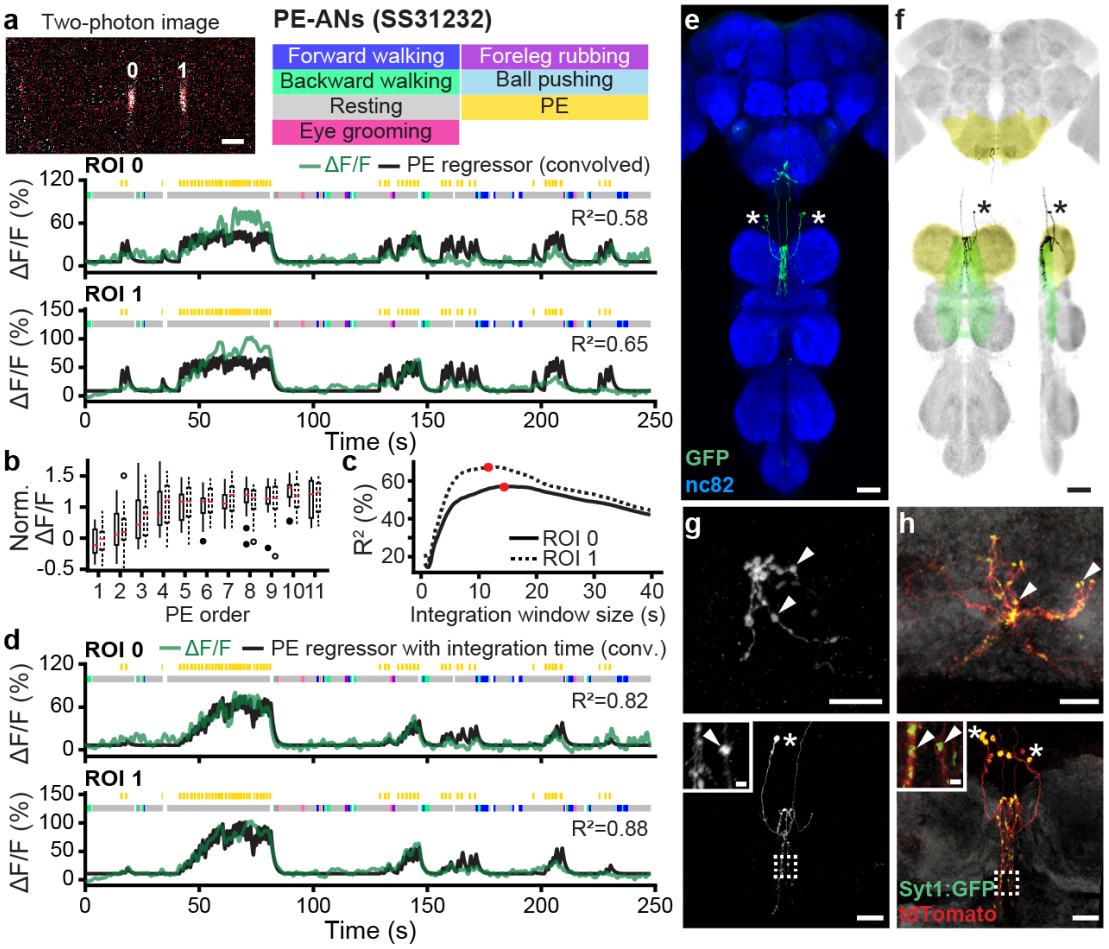

Figure 3.7: **Functional and anatomical properties of ascending neurons that integrate the number of proboscis extensions over time.** (**a**) (**top-left**) Two-photon image of axons from an *SS31232-GAL4* animal expressing OpGCaMP6f (cyan) and tdTomato (red). ROIs are numbered. Scale bar is 5 µm. **(bottom)** Behavioral epochs are color-coded. Representative $\Delta F/F$ time-series from two ROIs (green) overlaid with a prediction (black) obtained by convolving proboscis extension epochs with a Ca^{2+} indicator response function. Explained variance is indicated (R^2). (**b**) $\Delta F/F$, normalized with respect to the neuron's 90th percentile, as a function of proboscis extension (PE) number within a PE train for ROIs 0 (solid boxes, filled circles), or 1 (dashed boxes, open circles). (**c**) Explained variance (R^2) between $\Delta F/F$ time-series and a prediction obtained by convolving proboscis extension epochs with a Ca^{2+} indicator response function and a time-window. Time-windows that maximize the correlation for ROIs 0 (solid line) and 1 (dashed line) are indicated (red circles). (**d**) Behavioral epochs are color-coded. Representative $\Delta F/F$ time-series from two ROIs (green) are overlaid with a prediction (black) obtained by convolving proboscis extension epochs with a Ca^{2+} response function as well as the time windows indicated in panel **c** (red circles). Explained variance is indicated (R^2). (**e**) Standard deviation projection image of a *SS31232-GAL4* nervous system expressing smFP and stained for GFP (green) and Nc82 (blue). Cell bodies are indicated (white asterisks). Scale bar is 40 µm. (**f**) Projection as in **e** but for one MCFO-expressing, traced neuron (black asterisks).

The brain's GNG (yellow) and VNC's intermediate tectulum (green), and prothoracic leg neuromere (yellow) are color-coded. Scale bar is 40 μm. (**g,h**) Higher magnification projections of (**top**) brains and (**bottom**) VNCs for *SS31232-GAL4* animals expressing (**g**) the stochastic label MCFO, or (**h**) the synaptic marker, syt:GFP (green), and tdTomato (red). Insets magnify dashed boxes. Indicated are cell bodies (asterisks), and bouton-like structures (white arrowheads). Scale bars for brain images are 10 μm. Scale bars for VNC images and insets are 20 μm and 2 μm, respectively.

reporters in small sets of ANs, as well as new experimental and computational tools for recording and quantifying neural activity in behaving animals.

3.4.1 Encoding of high-level behavioral states

We discovered that ANs functionally encode high-level behavioral states **(Figure 3.8A)**, predominantly those related to self-motion like walking and resting. These encodings could be further distinguished as either broad (e.g., walk-ANs and foreleg-ANs), or narrow (e.g., turn-ANs and eye groom-ANs). Similarly, neurons in the vertebrate dorsal spinocerebellar tract have been shown to be more responsive to whole limb versus individual joint movements [184]. To compensate for the technical hurdle of relating relatively rapid joint movements to slow calcium indicator decay kinetics, we convolved joint angle time-series' with a kernel that would maximize the explanatory power of our regression analyses. Additionally, we confirmed that potential issues related to the non-orthogonality of joint angles and leg movements would not obscure our ability to explain the variance of AN neural activity **(Figure S3.3)**. Our observation that eye groom-AN activity could be explained by movements of the forelegs gave us further confidence that, in principle, leg movement encodings could be detected **(Figure 3.2C)**. To further confirm the absence of low-level joint and leg movement AN encoding, future work might directly manipulate the legs of animals while recording AN activity [185]. Finally, we sometimes observed that the activity of putative walk-encoding ANs was not fully explained by our walking regressor, nor our turn analysis, (e.g., SS44270, overlaps with SS41605). This suggests that some ANs may encode other features of locomotion.

3.4.2 Predominant projection to the brain's AVLP and GNG

We found that the majority of ANs do not project diffusely across the brain but rather specifically target either the AVLP and GNG **(Figure 3.8B)**. We hypothesize that this may reflect the contribution of AN behavioral state signals to two fundamental brain computations. First, the AVLP is a site known for multimodal, integrative sensory convergence [172–177]. Thus, the projection of ANs encoding resting, walking, and gust-like puffs to the AVLP **(Figure 3.8C)** may serve to contextualize time-varying visual and olfactory signals to indicate if they arise from self-motion, or from objects moving and odors fluctuating in the world. A

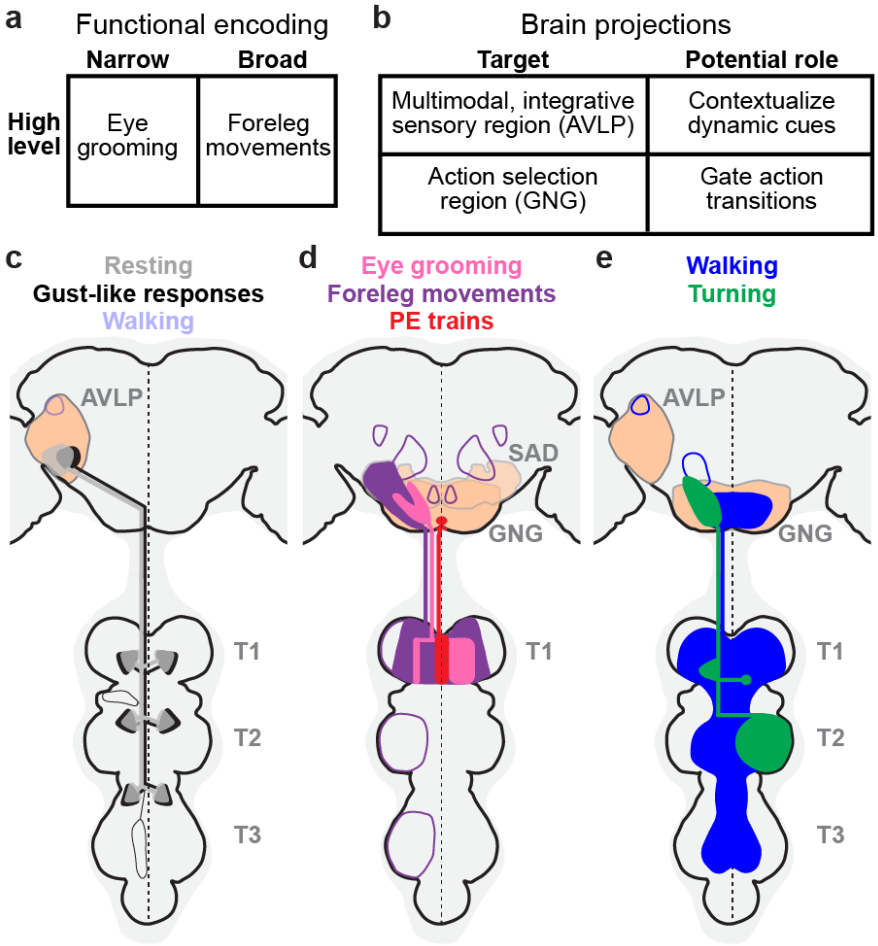

Figure 3.8: **Summary of ascending neuron functional encoding, brain targeting, and VNC patterning.** (**a**) ANs encode high-level behaviors in a narrow (e.g., eye grooming), or broad (e.g., foreleg movements) manner. (**b**) Corresponding anatomical analysis shows that ANs primarily target the AVLP, a multimodal, integrative brain region, and the GNG, a region associated with action selection. **(c,d)** By comparing functional encoding with brain targeting and VNC patterning, we find that (**c**) signals critical for contextualizing object motion—walking, resting, and gust-like stimuli—are sent to the AVLP, while (**d**) signals indicating diverse ongoing behaviors are sent to the GNG, potentially to influence future action selection. (**e**) Broad (e.g., walking), or narrow (e.g., turning) behavioral encoding is associated with diffuse and bilateral, or restricted and unilateral VNC innervations, respectively. (**c-e**) AN projections are color-coded by behavioral encoding. Axons and dendrites are not distinguished from one another. Brain and VNC regions are labelled. Frequently innervated brain regions—the GNG and AVLP—are highlighted (light orange). Less frequently innervated areas are outlined. The midline of the central nervous system is indicated (dashed line).

similar role—conveying self-motion—has been proposed for neurons in the vertebrate dorsal spinocerebellar tract [165]. Second, the GNG is thought to be an action selection center [54, 59,

82]. Thus, the projection of ANs encoding diverse behaviors—walking, turning, foreleg movements, eye-grooming, and proboscis extensions **(Figure 3.8D,E)**—to the GNG may contribute to the computation of whether potential future actions are compatible with ongoing behaviors. This role would be consistent with a hierarchical control approach used in robotics [153]. Notably, walk-ANs that project to the ventral GNG may be neuromodulatory in nature. Thus, they may be particularly well-poised to shift an animal's internal state and the relative values of potential future actions.

Notably, the GNG is also heavily innervated by descending neurons (DNs). Because ANs and DNs both contribute to action selection [53, 54, 59, 74], we speculate that they may connect within the GNG, forming a feedback loop between the brain and motor system. Specifically, ANs that encode specific actions might excite DNs that drive the same actions to generate behavioral persistence, while also suppressing DNs that drive conflicting actions. For example, turn-ANs may excite DNa01 and DNa02 which control turning [75, 89, 171], and foreleg-ANs may excite aDN1 and aDN2 that control grooming [93]. Of course the opposite might also be true: ANs might *inhibit* DNs that encode the same action to ensure that motor actions are terminated once they have been performed. These competing hypotheses may soon be tested using emerging connectomics datasets.

3.4.3 Patterning within the VNC is predictive of behavioral encoding

The morphology of an AN's neurites in the VNC is, to a remarkable degree, predictive of its encoding **(Figure 3.8C-E)**. We observed this in several ways. First, ANs innervating all three leg neuromeres (T1, T2, and T3) encode global self-motion—walking, resting, and gust-like puffs. By contrast, those with more restricted projections to one neuromere (T1 or T2) encode discrete actions—turning, eye grooming, foreleg movements, and PEs. This might reflect the cost of neural wiring, a constraint that may encourage a neuron to sample the minimal sensory and motor information required to compute a particular behavioral state. Second, broadly tuned ANs (encoding walking and foreleg-dependent behaviors) exhibited bilateral projections in the VNC while narrowly tuned ANs (encoding turning and eye grooming) exhibited unilateral and smooth, putatively dendritic projections. This was correlated with the degree of synchrony in the activity of pairs of ANs **(Figure S3.6)**.

Strikingly, for all ANs that we examined in-depth, we found evidence of axon terminals within the VNC. Thus, ANs may not simply relay behavioral state signals to the brain but may also perform other roles. For example, they might contribute to motor control as components of central pattern generators (CPGs) that generate rhythmic movements [4]. Similarly, rest-ANs might control the limb muscle tone needed to maintain a natural resting posture. ANs might also participate in computing behavioral states. For example, here we speculate that recurrent interconnectivity among PE-ANs might give rise to their temporal integration and encoding of PE number [181, 182]. Finally, ANs might contribute to action selection within the VNC. For example, eye groom-ANs might project to the contralateral T1 neuromere to suppress circuits

driving other foreleg-dependent behaviors like walking and foreleg rubbing.

3.4.4 Future work

Here we investigated animals that were generating spontaneous and puff-induced behaviors including walking and grooming. However, ANs likely also encode other behavioral states, unmeasured internal forces like posture-maintaining muscle stiffness, or even metabolic states. This is hinted at by the fact that some ANs' neural activities were not well explained by any of our behavioral regressors (e.g., R87H02, R39G01, R69H10, and SS29633). Additionally, nearly one-third of the ANs we examined were unresponsive, possibly due to the absence of appropriate context. For example, we found that some silent ANs could become very active during leg movements *only* when the spherical treadmill was removed (SS51017 and SS38631)**(Figure S3.7)**. Thus, future work may investigate AN encoding in other contexts including tethered flight. Finally, it would also be interesting to test the degree to which AN encoding is genetically hardwired or capable of adapting during motor learning or after injury [186, 187]. Our finding that ANs encode high-level behaviors and convey these states to integrative sensory and action selection centers in the brain may guide the study of ascending neurons in the mammalian spinal cord [164, 165, 184, 188–190], and also accelerate the development of more effective bioinspired algorithms for robotic sensory contextualization and action selection [153].

3.5 Materials and Methods

3.5.1 Key resource table

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
***Drosophila melanogaster* stocks**		
Split-GAL4 (spGAL4) lines	Laboratory of Barry Dickson and FlyLight project (Janelia Research Campus, Ashburn VA USA, see Table S3.2)	SS*****
MAN-spGAL4 (*; VT50660-AD; VT14014-DBD*)	[53] Laboratory of Michael Dickinson (Caltech, Pasadena CA USA)	N/A
R85A11-GAL4	[128] Bloomington Stock Center	R85A11
R87H02-GAL4	[128] Bloomington Stock Center	R87H02
R36G04-GAL4	[128] Bloomington Stock Center	R36G04
R39G01-GAL4	[128] Bloomington Stock Center	R39G01
R30A08-GAL4	[128] Bloomington Stock Center	R30A08
R70H06-GAL4	[128] Bloomington Stock Center	R70H06
R69H10-GAL4	[128] Bloomington Stock Center	R69H10
R57C10-Flp2::PEST in su(Hw)attP8; ; HA-V5-FLAG	[170] Bloomington Stock Center	MCFO-5
R57C10-Flp2::PEST in attP18; ; HA-V5-FLAG-OLLAS	[170] Bloomington Stock Center	MCFO-7
UAS-syt:GFP (*Pw[+mC]=UAS-syt.eGFP1, w[*]; ;*)	Bloomington Stock Center	N/A
UAS-OpGCaM6f; UAS-tdTomato (*; P20XUAS-IVS-Syn21-OpGCamp6F-p10 su(Hw)attp5 ; Pw[+mC]=UAS-tdTom.S3*)	Laboratory of Michael Dickinson (Caltech, Pasadena CA USA)	N/A
UAS-smFP (*; ; 10xUAS-IVS-myr::smGdP-FLAG (attP2)*)	Laboratory of Brian McCabe (EPFL, Lausanne CH)	N/A
Antibodies		
rabbit anti-GFP	Thermofisher	RRID: AB_2536526
mouse anti-Bruchpilot / nc82	Developmental Studies Hybridoma Bank	RRID: AB_2314866
rabbit anti-HA-tag	Cell Signaling Technology	RRID:AB_1549585
rat anti-FLAG-tag (DYKDDDDK)	Novus	RRID:AB_1625981
rabbit polyclonal anti-DsRed	Takara Biomedical Technology	RRID: AB_10013483
rabbit anti-V5-tag (GKPIPNPLLGLDST) conjugated with DyLight 550	Cayman Chemical	Cat# 11261
goat anti-rabbit secondary antibody conjugated with Alexa 488	Thermofisher	RRID: AB_143165
goat anti-mouse secondary antibody conjugated with Alexa 633	Thermofisher	RRID: AB_2535719
donkey anti-rabbit secondary antibody conjugated with AlexaFluor 594	Jackson ImmunoResearch Labs	RRID:AB_2340621
donkey anti-rat secondary antibody conjugated with AlexaFluor 647	Jackson ImmunoResearch Labs	RRID:AB_2340694
donkey anti-mouse secondary antibody conjugated with AlexaFluor 488	Jackson ImmunoResearch Labs	RRID:AB_2341099

donkey anti-rabbit secondary antibody conjugated with Cy3	Jackson ImmunoResearch Labs	RRID:AB_2307443
Chemicals		
Triton-X	Sigma-Aldrich	
PBS		
Normal goat serum (NGS)		
Slowfade	Thermofisher	Cat#: S36936
Software and digital resources		
Anaconda3	Anaconda	https://www.anaconda.com/products/individual
Python3.6	Python	https://www.python.org/
R 3.6.1	R	https://www.rstudio.com/
DeepLabCut	[39]	https://github.com/DeepLabCut/DeepLabCut
DeepFly3D	[37]	https://github.com/NeLy-EPFL/DeepFly3D
AxoID	This paper	https://github.com/NeLy-EPFL/AxoID
MakeAverageBrain	[191]	https://github.com/NeLy-EPFL/MakeAverageBrain/tree/workstation
Computational Morphometry Toolkit	N/A	https://www.nitrc.org/projects/cmtk
Brain and VNC template	[192]	https://www.janelia.org/open-science/jrc-2018-brain-templates
Fiji	[89]	https://fiji.sc
ThorImage3.1	Thorlabs	https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=9072
Thorsync	Thorlabs	https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=9072
Adobe Illustrator 2021	Adobe	https://www.adobe.com/
Deposited data and code		
Analysis code	This paper	https://github.com/NeLy-EPFL/Ascending_neuron_screen_analysis_pipeline
Data	This paper	https://dataverse.harvard.edu/dataverse/AN

3.5.2 Fly husbandry

Experimental animals were kept on dextrose cornmeal food at 25°C and 70% humidity on a 12-12 h day-light cycle using standard laboratory tools. All strains used are listed in the Key Resource Table. Female flies were subjected to experiments on 3-6days post-eclosion. Crosses used for experiments were flipped every 2-3 days.

3.5.3 *In vivo* two-photon calcium imaging experiments

Two-photon imaging was performed as described in [171] with minor changes in the recording configuration. We imaged coronal sections of AN axons in the cervical connective to avoid having neurons move outside the field of view due to behavior-related tissue deformations. Imaging was performed using a Galvo-Galvo scanning system. Image dimensions ranged

from 256 x 192 pixels (4.3 fps) to 320 x 320 pixels (3.7 fps), depending on the location of axonal regions-of-interest (ROIs) and the degree of displacement caused by animal behavior. During two-photon imaging, a 7-camera system was used to record fly behaviors as described in [37]. Rotations of the spherical treadmill, and the timing of puff stimuli were also recorded. Air or CO_2 puffs (0.08 L/min) were controlled using either a custom Python script, or manually with an Arduino controller. Puffs were delivered through a syringe needle positioned in front of the animal to stimulate behavior in sedentary animals, or to interrupt ongoing behaviors. To synchronize signals acquired at different sampling rates—optic flow sensors, two-photon images, puff stimuli, and videography—signals were digitized using a BNC 2110 terminal block (National Instrument, USA) and saved using ThorSync software (Thorlabs, USA). Sampling pulses were then used as references to align data based on the onset of each pulse. Then signals were interpolated using custom Python scripts.

3.5.4 Immunofluorescence tissue staining and confocal imaging

Fly brains and VNCs were dissected and fixed as described in [171] with small modifications in staining including antibodies and incubation conditions (see details below). Both primary (rabbit anti-GFP at 1:500; mouse anti-Bruchpilot / nc82 at 1:20) and secondary antibodies (goat anti-rabbit secondary antibody conjugated with Alexa 488 at 1:500; goat anti-mouse secondary antibody conjugated with Alexa 633 at 1:500) for smFP and nc82 staining were performed at room temperature for 24h.

To perform high-magnification imaging of MCFO samples, nervous tissues were incubated with primary antibodies: rabbit anti-HA-tag at 1:300 dilution, rat anti-FLAG-tag (DYKDDDDK) at 1:150 dilution, and mouse anti-Bruchpilot/nc82 at 1:20 dilution. These were diluted in 5% normal goat serum in PBS with 1% Triton-X (PBSTS) for 24 h at room temperature. The samples then were rinsed 2-3 times in PBS with 1% Triton-X (PBST) for 15 min before incubation with secondary antibodies: donkey anti-rabbit secondary antibody conjugated with AlexaFluor 594 at 1:500 dilution, donkey anti-rat secondary antibody conjugated with AlexaFluor 647 at 1:200 dilution, and donkey anti-mouse secondary antibody conjugated with AlexaFluor 488 at 1:500 dilution. These were diluted in PBSTS for 24 h at room temperature. Again, samples were rinsed 2-3 times in PBS with 1% Triton-X (PBST) for 15 min before incubation with the last diluted antibody: rabbit anti-V5-tag (GKPIPNPLLGLDST) conjugated with DyLight 550 at 1:300 dilution for another 24 h at room temperature.

To analyze single neuron morphological patterning, we crossed spGAL4 lines with MCFO-7 [170]. Dissection and MCFO staining were performed by Janelia FlyLight according to the FlyLight 'IHC-MCFO' protocol: https://www.janelia.org/project-team/flylight/protocols. Samples were imaged on an LSM710 confocal microscope (Zeiss) with a Plan-Apochromat 20x /0.8 M27 objective.

To prepare samples expressing tdTomato and syt:GFP, we chose to only stain tdTomato to minimize false positive signals for the synaptotagmin marker. Samples were incubated with a

diluted primary antibody: rabbit polyclonal anti-DsRed at 1:1000 dilution in PBSTS for 24 h at room temperature. After rinsing, samples were then incubated with a secondary antibody: donkey anti-rabbit secondary antibody conjugated with Cy3. Finally, all samples were rinsed 2–3 times for 10 min each in PBST after staining and then mounted onto glass slides with bridge coverslips in Slowfade mounting-media.

Confocal imaging was performed as described in [171]. In addition, high-resolution images for visualizing fine structures were captured using a 40x oil-immersion objective lens with an NA of 1.3 (Plan-Apochromat 40x/1.3 DIC M27, Zeiss) on an LSM700 confocal microscope (Zeiss). The zoom factor was adjusted based on the ROI size of each sample between 84.23×84.23 μm^2 and 266.74×266.74 μm^2. For high-resolution imaging, z-steps were fixed at 0.33 μm. Images were denoised, their contrasts were tuned, and standard deviation z-projections were generated using Fiji [140].

3.5.5 Two-photon image analysis

Raw two-photon imaging data were converted to gray-scale *.tiff image stacks for both green and red channels using custom Python scripts. RGB image stacks were then generated by combining both image stacks in Fiji [140]. We used AxoID to perform ROI segmentation and to quantify fluorescence intensities. Briefly, AxoID was used to register images using cross-correlation and optic flow-based warping [171]. Then, raw and registered image stacks underwent ROI segmentation, allowing $\%\Delta F/F$ values to be computed across time from absolute ROI pixel values. Simultaneously, segmented RGB image stacks overlaid with ROI contours were generated. Each frame of these segmented image stacks was visually examined to confirm AxoID segmentation, or to perform manual corrections using the AxoID GUI. In these cases, manually corrected $\%\Delta F/F$ and segmented image stacks were updated.

3.5.6 Behavioral data analysis

To reduce computational and data storage requirements, we recorded behaviors at 30 fps. This is nearly the Nyquist frequency for rapid walking (up to 16 step cycles/s [116]).

3D joint positions were estimated using DeepFly3D [37]. Due to the amount of data collected, manual curation was not practical. We therefore classified points as outliers when the absolute value of any of their coordinates (x, y, z) was greater than 5 mm (much larger than the fly's body size). Furthermore, we made the assumption that joint locations would only be incorrectly estimated for one of the three cameras used for triangulation. The consistency of the location across cameras could be evaluated using the reprojection error. To identify a camera with a bad prediction, we calculated the reprojection error only using two of the three cameras. The outlier was then replaced with the triangulation result of the pair of cameras with the smallest reprojection error. The output was further processed and converted to angles as described in [32].

We classified behaviors based on a combination of 3D joint angle dynamics and rotations of the spherical treadmill. First, to capture the temporal dynamics of joint angles, we calculated wavelet coefficients for each angle using 15 frequencies between 1 Hz and 15 Hz [193, 194]. We then trained a histogram gradient boosting classifier [195] using joint angles, wavelet coefficients, and ball rotations as features. Because flies perform behaviors in an unbalanced way (some behaviors are more frequently than others), we balanced our annotations using SMOTE [196]. The model was validated using 5-fold, three times repeated, stratified cross-validation. Fly speeds and heading directions were estimated using optical flow sensors [171]. To further improve the accuracy of the onset of walking we applied empirically-determined thresholds (pitch: 0.0038; roll: 0.0038; yaw: 0.014) to the rotational velocities of the spherical treadmill. The rotational velocities were smoothed and denoised using a moving average filter (length 81). All frames that were not previously classified as grooming or pushing, and for which the spherical treadmill was classified as moving, were labeled as 'walking'. These were furthered subdivided into forward or backward walking depending on the sign of the pitch velocity. Conversely, frames for which the spherical treadmill was not moving were labeled as 'resting'. To reduce the effect of optical flow measurement jitter, walking and resting labels were processed using a hysteresis filter that only changed state if at least 15 consecutive frames are in a new state. Classification in this manner was generally effective but most challenging for kinematically similar behaviors like eye- and antennal-grooming, or hindleg rubbing and abdominal grooming **(Figure S3.1E)**.

Proboscis extension (PE) events were classified based on the length of the proboscis **(Figure S3.1A-D)**. First, we trained a deep network [39] to identify the tip of the proboscis and a static landmark (the ventral part of the eye) from side-view camera images. Then, the distance between the tip of proboscis and this static landmark was calculated to obtain the PE length for each frame. A semi-automated PE event classifier was made by first denoising the traces of PE distances using a median filter with a 0.3 s running average. Traces were then normalized to be between 0 (baseline values) and 1 (maximum values). Next, PE speed was calculated using a data point interval of 0.1 s to detect significant changes in PE length. This way, only peaks larger than a manually set threshold of 0.03 Δnorm.length/0.1 s were considered. Because the peak speed usually occurred during the rising phase of a PE, a kink in PE speed was identified by multiplying the peak speed with an empirically-determined factor ranging from 0.4 to 0.6, and finding that speed within 0.5 s prior to the peak speed. The end of a PE was the time-point at which the same speed was observed within 2 s after the peak PE speed. This filtered out occasions where the proboscis remained extended for long periods of time. All quantified PE lengths and durations were then used to build a filter to remove false positives. PEs were then binarized to define PE epochs.

To quantify animal movements when the spherical treadmill was removed, we manually thresholded the variance of pixel values from a side view camera within a region of the image that included the fly. Pixel value changes were calculated using a running window of 0.2 s. Next, the standard deviation of pixel value changes was generated using a running window of 0.25 s. This trace was then smoothed and values lower than the empirically-determined

threshold were called 'resting' epochs. The remainder were considered 'movement' periods.

3.5.7 Regression analysis of PE integration time

To investigate the integrative nature of the PE-AN responses, we convolved PE traces with uniform time windows of varying sizes. This convolution was performed such that the fluorescence at each time point would be the sum of fluorescence during the previous 'window_size' frames (i.e., not a *centered* sliding window but one that only uses previous time points), effectively integrating over the number of previous PEs. This integrated signal was then masked such that all time points where the fly was not engaged in PE were set to zero. Then, this trace was convolved with a calcium indicator decay kernel, notably yielding non-zero values in the time intervals between PEs. We then determined the explained variance as described elsewhere and finally chose a window size maximizing the explained variance.

3.5.8 Linear modeling of neural fluorescence traces

Each regression matrix contains elements corresponding to the results of a ridge regression model for predicting the time-varying fluorescence (%ΔF/F) of ANs using specific regressors (e.g., low-level joint angles, or high-level behaviors). To account for slow calcium indicator decay dynamics, each regressor was convolved with a calcium response function. The half-life of the calcium response function was chosen from a range of 0.2 s to 0.95 s [171] in 0.05 s steps, in order to maximize the variance in fluorescence traces that convolved regressors could explain. The rise time was fixed at 0.1415 s [171]. The ridge penalty parameter was chosen using nested 10-fold stratified cross-validation [197]. The intercept and weights of all models were restricted to be positive, limiting our analysis to excitatory neural activity. Values shown in the matrices are the mean of 10-fold stratified cross-validation. We calculated Unique (UEV) and All-Explained Variance (AEV) by temporally shuffling the regressor in question, or all other regressors, respectively [155]. We tested the overall significance of our models using an F-statistic to reject the null hypothesis that the model does not perform better than an intercept-only model. The prediction of individual traces were performed using a single regressor plus intercept. Therefore they were not regularized.

3.5.9 Behavior-based neural activity analysis

For a given behavior, $\Delta F/F$ traces were compiled, cropped, and aligned with respect to their onset times. Mean and 95% confidence intervals for each time point were then calculated from these data. Because the duration of each behavioral epoch was different, we only computed mean and confidence intervals for epochs that had at least five data points.

To test if each behavior-triggered average $\Delta F/F$ was significantly different from the baseline, first, we aligned and upsampled fluorescence data that were normalized between 0 (baseline mean) and 1 (maximum) for each trial. For each behavioral epoch, the first 0.7 s of data were

removed. This avoided contaminating signals with neural activity from preceding behaviors (due to the slow decay dynamics of OpGCaMP6f). Next, to be conservative in judging whether data reflected noisy baseline or real signals, we studied their distributions. Specifically, we tested the normality of twenty resampled groups of 150 bootstrapped datapoints—a size that reportedly maximizes the power of the Shapiro-Wilk test [198]. If a majority of results did not reject the null hypothesis, the entire recording was considered baseline noise and the $\Delta F/F$ for a given behavioral class was not considered significantly different from baseline. On the other hand, if the datapoints were not normally distributed, the baseline was determined using an Otsu filter. For recordings that passed this test of normality, if the majority of six ANOVA tests on the bootstrapped data rejected the null hypothesis and the datapoints of a given behavior were significantly different (***p<0.001, **p<0.01, *p<0.05) from baseline (as indicated by a posthoc Tukey test), these data were considered signal and not noise.

To analyze PE-AN responses to each PE during PE trains, putative trains of PEs were manually identified to exclude discrete PE events. PE trains included at least 3 consecutive PEs in which each PE lasted at least 1 s and there was less than 3 s between each PE. Then, the mean fluorescence of each PE was computed for 25 PE trains (n=11 animals). The median, IQR, and 1.5 IQR were then computed for PEs depending on their ordered position within their PE trains. We focused our analysis on the first 11 PEs because they had a sufficiently large amount of data.

3.5.10 Neural fluorescence-triggered averages of spherical treadmill rotational velocities

A semi-automated neural fluorescence event classifier was constructed by first denoising ΔF/F traces by averaging them using a 0.6 s running window. Traces were then normalized to be between 0 (their baseline values) and 1 (their maximum values). To detect large deviations, the derivative of the normalized ΔF/F time-series was calculated at an interval of 0.1 s. Only peaks greater than an empirically determined threshold of 0.03 *d*norm ΔF/F / 0.1 s were considered events. Because peak fluorescence derivatives occurred during the rising phase of neural fluorescence events, the onset of a fluorescence event was identified as the time where the ΔF/F derivative was 0.4-0.6x the peak within the preceding 0.5 s time window. The end of the event was defined as the time that the ΔF/F signal returned to the amplitude at event onset before the next event. False positives were removed by filtering out events with amplitudes and durations that were lower than the empirically determined threshold. Neural activity event analysis for turn-ANs was performed by testing if the mean normalized fluorescence event for one ROI was larger than the other ROI by an empirically determined factor of 0.2x. Corresponding ball rotations for events that pass this criteria were then collected. Fluorescence events onsets were then set to 0 s and aligned with spherical treadmill rotations. Using these rotational velocity data, we calculated the mean and 95% confidence intervals for each time point with at least five data points. A 1 s period before each fluorescence event was also analyzed as a baseline for comparison.

3.5.11 Brain and VNC confocal image registration

All confocal images, except for MCFO image stacks, were registered based on nc82 neuropil staining. We built a template and registered images using the CMTK munger extension [191]. Code for this registration process can be found at: https://github.com/NeLy-EPFL/MakeAverageBrain/tree/workstation. Brain and VNC of MCFO images were registered to JRC 2018 templates [192] using the Computational Morphometry Toolkit: https://www.nitrc.org/projects/cmtk. The template brain and VNC can be downloaded here: https://www.janelia.org/open-science/jrc-2018-brain-templates.

3.5.12 Analysis of individual AN innervation patterns

Single AN morphologies were traced by masking MCFO confocal images using either active tracing, or manual background removal in Fiji [140]. Axons in the brain were manually traced using the Fiji plugin 'SNT'. Most neurites in the VNC were isolated by (i) thresholding to remove background noise and outliers, and (ii) manually masking debris in images. In the case of ANs from SS29579, a band-pass color filter was applied to isolate an ROI that spanned across two color channels. The boundary of the color filter was manually tuned to acquire the stack for a single neuron mask. After segmentation, the masks of individual neurons were applied across frames to calculate the intersectional pixel-wise sum with another mask containing either (i) neuropil regions of the brain and VNC, (ii) VNC segments, or (iii) left and right halves of the VNC. Brain and VNC neuropil regions and their corresponding abbreviations were according to established nomenclature [114]. Neuropil region masks can be downloaded here: https://v2.virtualflybrain.org. These were also registered to the JRC 2018 template. Masks for T1, T2, and T3 VNC segments were based on previously delimited boundaries [59]. The laterality of a neuron's VNC innervation was calculated as the ratio of the absolute difference between its left and right VNC innervations divided by its total innervation. Masks for the left and right VNC were generated by dividing the VNC mask across its midline.

3.5.13 AxoID: a deep learning-based software for tracking axons in imaging data

AxoID aims to extract the GCaMP fluorescence values for axons present on coronal section two-photon microscopy imaging data. In this manuscript, it is used to record activity from ascending neurons (ANs) passing through the *Drosophila melanogaster* cervical connective. Fluorescence extraction works by performing the following three main steps **(Figure S3.2A)**. First, during a *detection* stage, ROIs corresponding to axons are segmented from images. Second, during a *tracking* stage, these ROIs are tracked across frames. Third, *fluorescence* is computed for each axon over time.

To track axons, we used a two-stage approach: detection and then tracking. This allowed us to improve each problem separately without the added complexity of developing a detector that must also do tracking. Additionally, this allowed us to detect axons without having to know

how many there are in advance. Lastly, significant movement artifacts between consecutive frames pose additional challenges for robustness in temporal approaches while, in our case, we can apply the detection on a frame-by-frame basis. However, we note that we do not leverage temporal information.

Detection

Axon detection consists of finding potential axons by segmenting the background and foreground of each image. An ROI or putative axon is defined as a group of connected pixels segmented as foreground. Pixels are considered connected if they are next to one another.

By posing detection as a segmentation problem, we have the advantage of using standard computer vision methods like thresholding, or artificial neural networks that have been developed for medical image segmentation. Nevertheless, this simplicity has a drawback: if axons appear very close to one another and their pixels are connected, they may be segmented as one ROI rather than two. We try to address this issue using an ROI separation approach described later.

Image segmentation is performed using deep learning on a frame-by-frame basis, whereby a network generates a binary segmentation of a single image. As a post-processing step, all ROIs smaller than a minimum size are discarded. Here, we empirically chose 11 pixels as the minimum size as a trade-off between removing small spurious regions while still detecting small axons.

We chose to use a U-Net model [199] with slight modifications because of its, or its derivatives', performance on recent biomedical image segmentation problems [200–202]. We add zero-padding to the convolutions to ensure that the output segmentation has the same size as the input image, thus fully segmenting it in a single pass, and modify the last convolution to output a single channel rather than two. Batch normalization [203] is used after each convolution and its non-linearity function. Finally, we reduce the width of the network by a factor of 4: each feature map has 4 times fewer channels than the original U-Net, not counting the input or output. The input pixel values are normalized to the range [-1, 1], and the images are sufficiently zero-padded to ensure that the size can be correctly reduced by half at each max-pooling layer.

To train the deep learning network, we use the Adam optimizer [204] on the binary cross-entropy loss with weighting. Each background pixel is weighted based on its distance to the closest ROI, given by $1 + exp(-\frac{d}{3}^2)$ with d the Euclidean distance, plus a term that increases if the pixel is a border between two axons, given by $exp(-\frac{d_1+d_2}{6}^2)$ with d_1 and d_2 as the distances to the two closest ROIs, as in [199]. These weights aim to encourage the network to correctly segment the border of the ROI and to keep a clear separation between two neighboring regions. At training time, the background and foreground weights are scaled by $\frac{b+f}{2b}$ and $\frac{b+f}{2f}$, respectively, to take into account the imbalance in the number of pixels, where b and f are

the quantity of background and foreground (i.e., ROI) pixels in the entire training dataset. To evaluate the resulting deep network, we use the Sørensen-Dice coefficient [205, 206] at the pixel level, which is equivalent to the F1-score. The training is stopped when the validation performance does not increase anymore.

The network was trained on a mix of experimental and synthetic data. We also apply random gamma corrections to the training input images, with γ sampled in [0.7, 1.3] to keep reasonable values, and to encourage robustness against intensity variations between experiments. The target segmentation of the axons on the experimental data was generated with conventional computer vision methods. First, the images were denoised with the non-local means algorithm [207] using the Python implementation of OpenCV [208]. We used a temporal window size of 5, and performed the denoising separately for the red and green channels, with a filter strength $h = 11$. The grayscale result was then taken as the per-pixel maximum over the channels. Following this, the images were smoothed with a Gaussian kernel of standard deviation 2 pixels, and thresholded using Otsu's method [209]. A final erosion was applied and small regions below 11 pixels were removed. All parameter values were set empirically to generate good qualitative results. In the end, the results were manually filtered to keep only data with satisfactory segmentation.

Because the experimental data have a fairly simply visual structure, we constructed a pipeline in Python to generate synthetic images visually similar to real ones. This was achieved by first sampling an image size for a given synthetic experiment, then by sampling 2D Gaussians over it to simulate the position and shape of axon cross-sections. After this, synthetic tdTomato levels were uniformly sampled and GCaMP dynamics were created for each axons by convolving a GCaMP response kernel with Poisson noise to simulate spikes. Then, the image with the Gaussian axons was deformed multiple times to make different frames with artificial movement artifacts. Eventually, we sampled from the 2D Gaussians to make the axons appear pixelated, and added synthetic noise to the images.

In the end, we chose a deep learning-based approach because our computer vision pipeline alone was not be robust enough. Our pipeline is used to generate a target segmentation dataset from which we manually select a subset of acceptable results. These results are then used to train the deep learning model.

Fine-tuning At the beginning of the detection stage, an optional fine-tuning of the network can be applied to try to improve the segmentation of axons. The goal is to have a temporary network adapted to the current data for better performance. To do this, we train the network on a subset of experimental frames using automatically generated target segmentations.

The subset of images is selected by finding a cluster of frames with high cross-correlation-based similarity. For this, we only consider the tdTomato channel to avoid the effects of GCaMP dynamics. Each image is first normalized by its own mean pixel intensity μ and standard deviation σ: $p(i,j) \leftarrow \frac{p(i,j)-\mu}{\sigma}$, where $p(i,j)$ is the pixel intensity p at the pixel location i,j.

The cross-correlation is then computed between each pair of normalized images p_m and p_n as $\sum_{i,j} p_m(i,j) \cdot p_n(i,j)$. Afterwards, we take the opposite of the cross-correlation as a distance measure and use it to cluster the frames with the OPTICS algorithm [210]. We set the minimal number of sample for a cluster to 20, in order to maintain at least 20 frames for fine-tuning, and a maximum neighborhood distance of half the largest distance between frames. Finally, we select the cluster of images with the highest average cross-correlation (i.e., the smallest average distance between its elements).

Then, to generate a target segmentation image for these frames, we take their temporal average and optionally smooth it, if there are less than 50 images, to help remove the noise. The smoothing is done by filtering with a Gaussian kernel of standard deviation 1 pixel, then median filtering over each channel separately. The result is then thresholded through a local adaptive method, computed by taking the weighted mean of the local neighborhood of the pixel, subtracted by an offset. We apply Gaussian weighting over windows of 25×25 pixels, with an offset of -0.05, determined empirically. Finally, we remove regions smaller than 11 pixels. The result serves as a target segmentation image for all of the fine-tuning images.

The model is then trained on 60% of these frames with some data augmentation, while the other 40% are used for validation. The fine-tuning stops automatically if the performance on the validation frames drops. This avoids bad generalization for the rest of the images. The binary cross-entropy loss is used, with weights computed as discussed previously. For the data augmentation, we use random translation ($\pm 20\%$), rotation ($\pm 10°$), scaling ($\pm 10\%$), and shearing ($\pm 5°$).

Tracking

Once the regions of interest are segmented, the next step of the pipeline consists of tracking the axons through time. This means defining which axons exist, and then finding the ROI they correspond to in each frame.

Tracking template To accomplish this, the tracker records the number of axons, their locations with respect to one another, and their areas. It stores this information into what we call the 'tracker template'. Then, for each frame, the tracker matches its template axons to the ROIs to determine which regions correspond to which axons.

The tracker template is built iteratively. It is first initialized and then updated by matching with all experimental data. The initialization depends on the optional fine-tuning in the detection step. If there is fine-tuning, then the smoothed average of the similar frames and its generated segmentation are used. Otherwise, one frame of the experiment is automatically selected. For this, AxoID considers only the frames with a number of ROIs equal to the most frequent number of ROIs, and then selects the image with the highest cross-correlation with the temporal average of these frames. It is then smoothed and taken with the segmentation

produced by the detection network as initialization. The cross-correlation and smoothing are computed identically as in the fine-tuning. Each ROI in the initialization segmentation defines an axon in the tracker template, with its area and position recorded as initial properties.

Afterwards, we update them by matching each experimental frame to the tracker template. It consists of assigning the ROI to the tracker axons, and then using these regions' areas and positions to update the tracker. The images are matched sequentially, and the axons properties are taken as running averages of their matched regions. For example, considering the n^{th} match, the area of an axon is updated as:

$$area \leftarrow \frac{area * n + area_{ROI}}{n+1}$$

Because of this, the last frames are matched to a tracker template that is different from the one used for the first frames. Therefore, we fix the axons properties after the updates and match each frame again to obtain the final identities of the ROIs.

Matching To assign axon identities to the ROIs of a frame, we perform a matching between them as discussed above. To solve it, we define a cost function for matching a template axon to a region which represents how dissimilar they are. Then, using the Hungarian assignment algorithm [211], we find the optimal matching with the minimum total cost **(Figure S3.2B)**.

Because some ROIs in the frame may be wrong detections, or some axons may not be correctly detected, the matching has to allow for the regions and axons to end up unmatched for some frames. Practically, we implement this by adding "dummy" axons to the matching problem with a flat cost. To guarantee at least one real match, the flat cost is set to the maximum between a fixed value and the minimum of the costs between regions and template axons with a margin of 10%: $dummy = max(v, 1.1 \cdot min(\text{costs}))$ with $v = 0.3$ the fixed value. Then, we can use the Hungarian method to solve the assignment, and all ROIs linked to these dummy axons can be considered unmatched.

We define the cost of assigning a frame's ROI i to a tracker template axon k by their absolute difference in area plus the mean cost of an optimal inner matching of the other ROI to the other axons assuming i and k are already matched:

$$\text{cost}(i,k) = w_{area}|area_i - area_k| + \frac{1}{N_{ROI}-1}\sum_{i' \neq i} \text{cost}'(i', k^*_{i'})$$

where $w_{area} = 0.1$ is a weight for balancing the importance of the area, N_{ROI} is the number of ROI in the frame, and $\text{cost}'(i', k^*_{i'})$ is the inner cost of assigning region $i' \neq i$ to axon $k^*_{i'} \neq k$ selected in an "inner" assignment problem, see below. In other words, the cost is relative to how well the rest of the regions and axons match if we assume that i and k are already matched.

The optimal inner matching is computed through another Hungarian assignment, for which we define another cost function. For this "inner" assignment problem, the cost of matching an ROI $i' \neq i$ and a template axon $k' \neq k$ is defined by how far they are and their radial difference with respect to the matched i and k, plus their difference in area:

$$\text{cost}'(i', k') = \left(\frac{w_{dist}}{\eta_{dist}} ||(x_{i'} - x_i) - (x_{k'} - x_k)|| + \frac{w_\theta}{\eta_\theta} |\theta_{i'} - \theta_{k'}| \right) \frac{H}{H + x^y_{k'}} + w_{area} |area_{i'} - area_{k'}|$$

$$\text{with } \eta_\theta = \arctan\left(\alpha_\theta \frac{\eta_{dist}}{||x_{k'} - x_k||} \right)$$

where $w_{dist} = 1.0$, $w_\theta = 0.1$, and $w_{area} = 0.1$ are weighting parameters, $\eta_{dist} = \min(H, W)$ and η_θ are normalization factors with H and W the height and width of the frame and $\alpha_\theta = 0.1$ a secondary normalization factor. The $\cdot^y$ operation returns the height component of a vector, and the $\frac{H}{H + x^y_{k'}}$ term is useful to reduce the importance of the first terms if the axon k' is far from axon k in the height direction. This is needed as the scanning of the animal's cervical connective is done from top to bottom, thus we need to allow for some movement artifacts between the top and bottom of the image. Note that the dummy axons for unmatched regions are also added to this inner problem.

This inner assignment is solved for each possible pair of axon-ROI to get all final costs. The overall matching is then performed with them. Because we are embedding assignments, the computational cost of the tracker increases exponentially with the number of ROIs and axons. It stays tractable in our case as we generally deal with few axons at a time. All parameter values used in the matching were found empirically by trial and error.

Identities post-processing: ROI separation In the case of fine-tuning at the detection stage, AxoID will also automatically try to divide ROIs that are potentially two or more separate axons. We implement this to address the limitation introduced by detecting axons as a segmentation: close or touching axons may get segmented together.

To do this, it first searches for potential ROIs to be separated by reusing the temporal average of the similar frames used for the fine-tuning. This image is initially segmented as described before. Then local intensity maxima are detected on a grayscale version of this image. To avoid small maxima due to noise, we only keep those with an intensity ≥ 0.05, assuming normalized grayscale values in $[0, 1]$. Following this, we use the watershed algorithm, with the scikit-image [212] implementation, to segment the ROI based on the gray level and detected maxima. In the previous stages, we discarded ROIs under 11 pixels to avoid small spurious detections. Similarly, here we fuse together adjacent regions that are under 11 pixels to only output results after the watershedding above or equal to that size. Finally, a border of 1 pixel width is inserted between regions created from the separation of an ROI.

These borders are the divisions separating the ROI, referred to as "cuts". We parameterize each of these as a line, defined as its normal vector $\mathbf{n}$ and distance d to the origin of the image

(top-left). To report them on each frame, we first normalize this line to the current ROI, and then reverse that process with respect to the corresponding regions on the other frames. To normalize the line to an ROI, we fit an ellipse on the ROI contour in a least-square sense. Then the line parameters are transformed into this ellipse's local coordinates following Algorithm 1. It is essentially like transforming the ellipse into a unit circle, centered and axis-aligned, and applying a similar transformation to the cutting line **(Figure S3.2C, middle)**. The choice of fitting an ellipse is motivated by the visual aspect of the axons in the experimental data as they are fairly similar to elongated ellipses. Considering this, a separation between two close ellipses could be simplified to a linear border, motivating the linear representation of the ROI separation.

Because this is done as a post-processing step following tracking, we can apply that division on all frames. To do this, we again fit an ellipse to their ROI contours in the least-squares sense. Then, we take the normalized cutting line and fit it back to each of them according to Algorithm 2. This is similar to transforming the normalized unit circle to the region ellipse and applying the same transform to the line **(Figure S3.2C, right)**.

Finally, a new axon is defined for each cut. In each frame, the pixels of the divided region on the furthest side of the linear separation (with respect to the fitting ellipse center) are taken as the new ROI of that axon for that given frame.

In case there are multiple cuts of the same ROI (e.g., because three axons were close), the linear separations are ordered by distance to the center of the fitting ellipse and are then applied in succession. This is simple and efficient, but assumes there is little to no crossing between linear cuts.

Fluorescence extraction

With the detection and tracking results, we know where each axon is in the experimental data. Therefore, to compute tdTomato and GCaMP fluorophore time-series we take the average of non-zero pixel intensities of the corresponding regions in each frame. We report the GCaMP fluorescence at time t as F_t, and the ratio of GCaMP to tdTomato fluorescence at time t as R_t to gain robustness against image intensity variations.

The final GCaMP fluorescence is reported as in [171]:

$$\Delta F/F = \frac{F_t - F}{F}$$

where F is a baseline fluorescence. Similarly, we report the ratio of GCaMP over tdTomato as in [141, 171]:

$$\Delta R/R = \frac{R_t - R}{R}$$

where R is the baseline. The baseline fluorescences F and R are computed as the minimal

Algorithm 1: Normalize a line with an ellipse

Input: $line$, $ellipse$
Output: normalized line $line'$
/* Initialization */
$\mathbf{n} \leftarrow line.\mathbf{normal}$;
$d \leftarrow line.distance$;
$\mathbf{c} \leftarrow ellipse.\mathbf{center}$;
$w \leftarrow ellipse.width/2$;
$h \leftarrow ellipse.height/2$;
$\theta \leftarrow ellipse.rotation$;
$\mathbf{R}_{-\theta} :=$ rotation matrix of angle $-\theta$;
/* Normalization */
$d' \leftarrow d - \mathbf{c} \cdot \mathbf{n}$;
$\mathbf{n}' \leftarrow \mathbf{R}_{-\theta}\,\mathbf{n}$;
$\mathbf{n}'.x \leftarrow \mathbf{n}'.x/\mathbf{c}.y$;
$\mathbf{n}'.y \leftarrow \mathbf{n}'.y/\mathbf{c}.x$;
$d' \leftarrow d'/(w * h)$;
$line'.distance \leftarrow d'/||\mathbf{n}'||$;
$line'.\mathbf{normal} \leftarrow \mathbf{n}'/||\mathbf{n}'||$;

Algorithm 2: Fit a line to an ellipse

Input: $line$, $ellipse$
Output: fitted line $line'$
/* Initialization */
$\mathbf{n} \leftarrow line.\mathbf{normal}$;
$d \leftarrow line.distance$;
$\mathbf{c} \leftarrow ellipse.\mathbf{center}$;
$w \leftarrow ellipse.width/2$;
$h \leftarrow ellipse.height/2$;
$\theta \leftarrow ellipse.rotation$;
$\mathbf{R}_{\theta} :=$ rotation matrix of angle θ;
/* Fitting */
$\mathbf{n}' \leftarrow \mathbf{n}$;
$\mathbf{n}'.x \leftarrow \mathbf{n}'.x * \mathbf{c}.y$;
$\mathbf{n}'.y \leftarrow \mathbf{n}'.y * \mathbf{c}.x$;
$d' \leftarrow d * (w * h)$;
$d' \leftarrow d'/||\mathbf{n}'||$;
$\mathbf{n}' \leftarrow \mathbf{n}'/||\mathbf{n}'||$;
$line'.\mathbf{normal} \leftarrow R_{\theta}\,\mathbf{n}'$;
$line'.distance \leftarrow d' + \mathbf{c} \cdot \mathbf{n}'$;

temporal average over windows of 10 s of the fluorophore time series F_t and R_t, respectively. Note that axons can be missing in some frames. For instance, if they were not detected or leave the image during movement artifacts. In this case, the fluorescence of that axon will have missing values at the time index t in which it was absent.

3.5.14 Overall workflow

To improve the performance of AxoID, the fluorescence extraction pipeline is applied three times: once over the raw data, once over the data registered using cross-correlation, and once over the data registered using optic-flow warping. Note that the fine-tuning in the detection stage is not used with the raw experimental data as it is based on the cross-correlation between the frames and would therefore lead to worse or redundant results with the data registered using cross-correlation. Eventually, the three fluorescence results can be visualized, chosen from, and corrected by a user through a GUI **(Figure S3.2D)**.

Data registration

Registration of the experimental frames consists in transforming each image to make them similar to a reference image. The goal is to reduce the artifacts introduced by animal movements and to align axons across frames. This should help to improve the results of the detection and tracking.

Cross-correlation Cross-correlation registration consists of translating an image so that its correlation to a reference is maximized. Note that the translated image wraps-around (e.g., pixels disappearing to the left reappear on the right). This aims to align frames against translations, but is unable to counter rotations or local deformations. We used the single step Discrete Fourier Transform (DFT) algorithm [150] to find the optimal translation of the frame. It first transforms the images into the Fourier domain, computes an initial estimate of the optimal translation, and then refines this result using a DFT. We based our Python implementation on previous work [213].

For each experiment, the second frame is taken as the reference frame to avoid recording artifacts that sometimes appear on the first recorded image.

Optic-flow registration Optic flow-based registration was previously published [171]. Briefly, this approach computes an optic flow from the frame to a reference image, then deforms it by moving the pixels along that flow. The reference image is taken as the first frame of the experiment. This method has the advantage of being able to compute local deformations, but at a high computational cost.

AxoID GUI

Finally, AxoID contains a GUI where a user can visualize the results, select the best one, and manually correct it.

First, the user is presented with three outputs of the fluorescence extraction pipeline from the raw and registered data with the option of visualizing different information to select the one to keep and correct. Here, the detection and tracking outputs are shown, as well as other information like the fluorescence traces in $\Delta F/F$ or $\Delta R/R$. One of the results is then selected and used throughout the rest of the pipeline.

Following this, the user can edit the tracker template, which will then automatically update the ROI identities across frames. The template and the identities for each frame are shown, with additional information like the image used to initialized the template. The user has access to different tools: axons can be fused, for example, if they actually correspond to a single real axon that was incorrectly detected as two, and, conversely, one axon can be manually separated in two if two close ones are detected together. Moreover, useless axons or wrong detections can be discarded.

Once the user is satisfied with the overall tracker, they can correct individual frames. At this stage, it is possible to edit the detection results by discarding, modifying, or adding ROIs onto the selected image. Then, the user may change the tracking results by manually correcting the identities of these ROIs. In the end, the final fluorescence traces are computed on the selected outputs including user corrections.

3.6 Code availability

Analysis code is available at:
https://github.com/NeLy-EPFL/Ascending_neuron_screen_analysis_pipeline
AxoID code is available at:
https://github.com/NeLy-EPFL/AxoID

3.7 Data availability

Data are available at:
https://dataverse.harvard.edu/dataverse/AN

3.8 Acknowledgments

We thank the Janelia Research Campus FlyLight project for generating Ascending Neuron split-GAL4 driver lines. PR acknowledges support from an SNSF Project grant (175667) and an SNSF Eccellenza grant (181239). FA acknowledges support from a Boehringer Ingelheim Fonds PhD stipend.

3.9 Author contributions

C-L.C. - Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Validation, Writing – Original Draft Preparation, Writing – Review & Editing, Visualization.
F.A. - Methodology, Software, Formal Analysis, Investigation, Data Curation, Validation, Data Curation, Writing – Original Draft Preparation, Writing - Review & Editing.
R.M. - Methodology, Investigation, Data Curation, Validation. Writing - Review & Editing
V.M. - Investigation, Data Curation, Visualization. Writing - Review & Editing
N.T. - Methodology, Software, Formal Analysis, Data Curation Visualization. Writing - Review & Editing
S.G. - Methodology, Software, Formal Analysis, Data Curation, Visualization. Writing - Review & Editing
B.D. - Resources, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition. Writing - Review & Editing
P.R. - Conceptualization, Methodology, Resources, Writing – Original Draft Preparation, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition.

3.10 Supplementary Information

3.10.1 Supplementary Figures

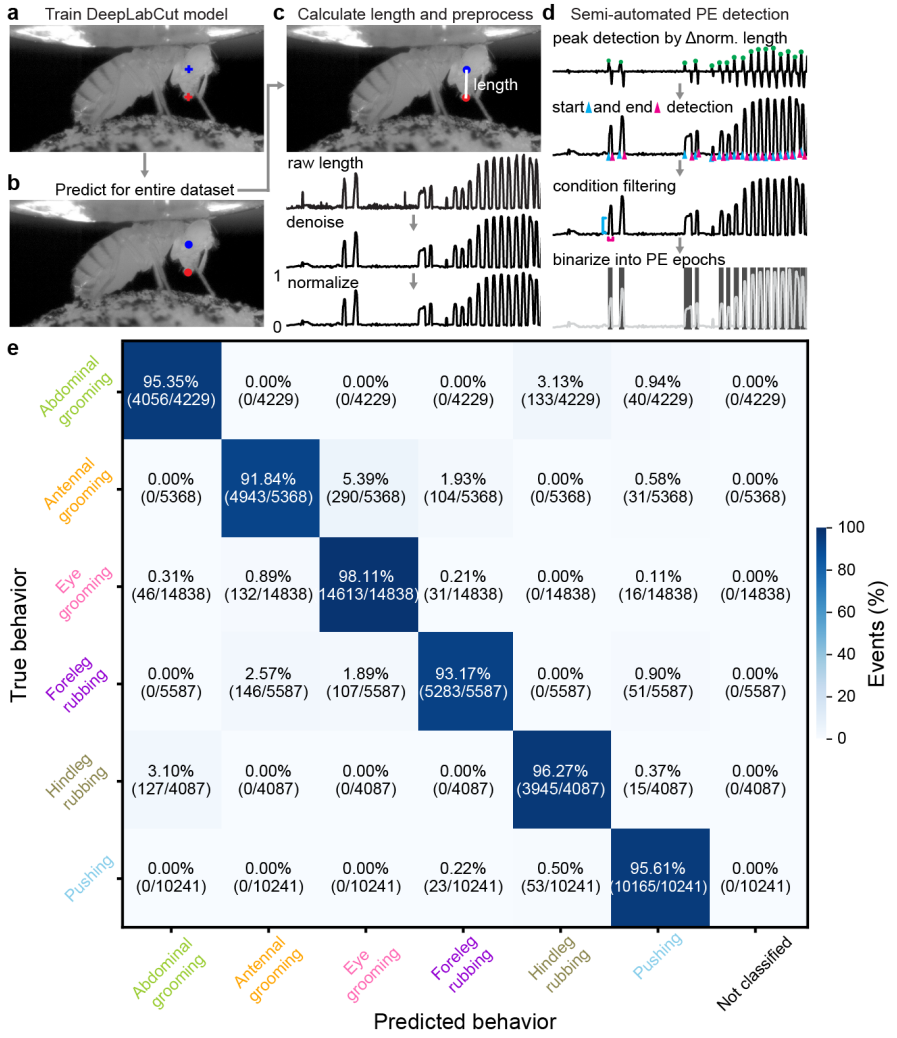

Figure S3.1: **Semi-automated tracking of proboscis extensions, and the accuracy of the behavioral classifier. Related to Figure 3.1**. We detected proboscis extensions using side-view camera images. (**a**) First, we trained a deep neural network model with manual annotations of landmarks on the ventral eye (blue cross) and distal proboscis tip (red cross). (**b**) Then we applied the trained model to estimate these locations throughout the entire dataset. (**c**) Proboscis extension length was calculated as the denoised and normalized distance between landmarks. (**d**) Using these data, we performed semi-automated detection of PE epochs by first identifying peaks from normalized proboscis extension lengths. Then we detected the start (cyan triangle) and end (magenta triangle) of these events. We removed false-positive detections by thresholding the amplitude (cyan line) and duration (magenta line) of events. Finally, we generated a binary trace of PE epochs (shaded regions). (**e**) A confusion matrix quantifies the accuracy of behavioral state classification using 10-fold, stratified cross-validation of a

histogram gradient boosting classifier. Walking and resting are not included in this evaluation because they are predicted using spherical treadmill rotation data. The percentage of events in each category ('predicted' behavior versus ground-truth, manually-labelled 'true' behavior) is color-coded.

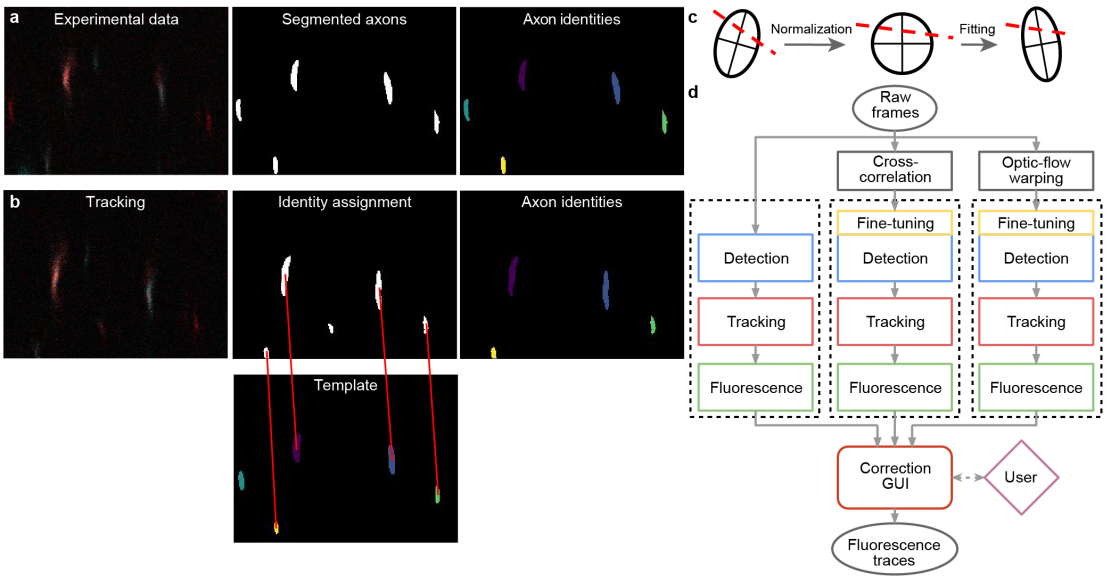

Figure S3.2: **AxoID, a deep learning-based algorithm that detects and tracks axon cross-sections in two-photon microscopy images. Related to Figure 3.1.** (**a**) Pipeline overview: a single image frame (left) is segmented (middle) during the detection stage with potential axons shown (white). Tracking identities (right) are then assigned to these ROIs. (**b**) To track ROIs across time, ROIs in a tracker template (bottom-middle) are matched (red lines) to ROIs in the current segmented frame (top-middle). An undetected axon in the tracker template (cyan) is left unmatched. (**c**) ROI separation is performed for fused axons. An ellipse is first fit to the ROI's contour and a line is fit to the separation (dashed red line). For normalization, the ellipse is transformed into an axis-aligned circle and the linear separation is transformed accordingly. For another frame, a transformation of the circle into a newly fit ellipse is computed and applied to the line. The ellipse's main axes are shown for clarity. (**d**) The AxoID workflow. Raw experimental data is first registered via cross-correlation and optic flow warping. Then, raw and registered data are separately processed by the fluorescence extraction pipeline (dashed rectangles). Finally, a GUI is used to select and correct the results.

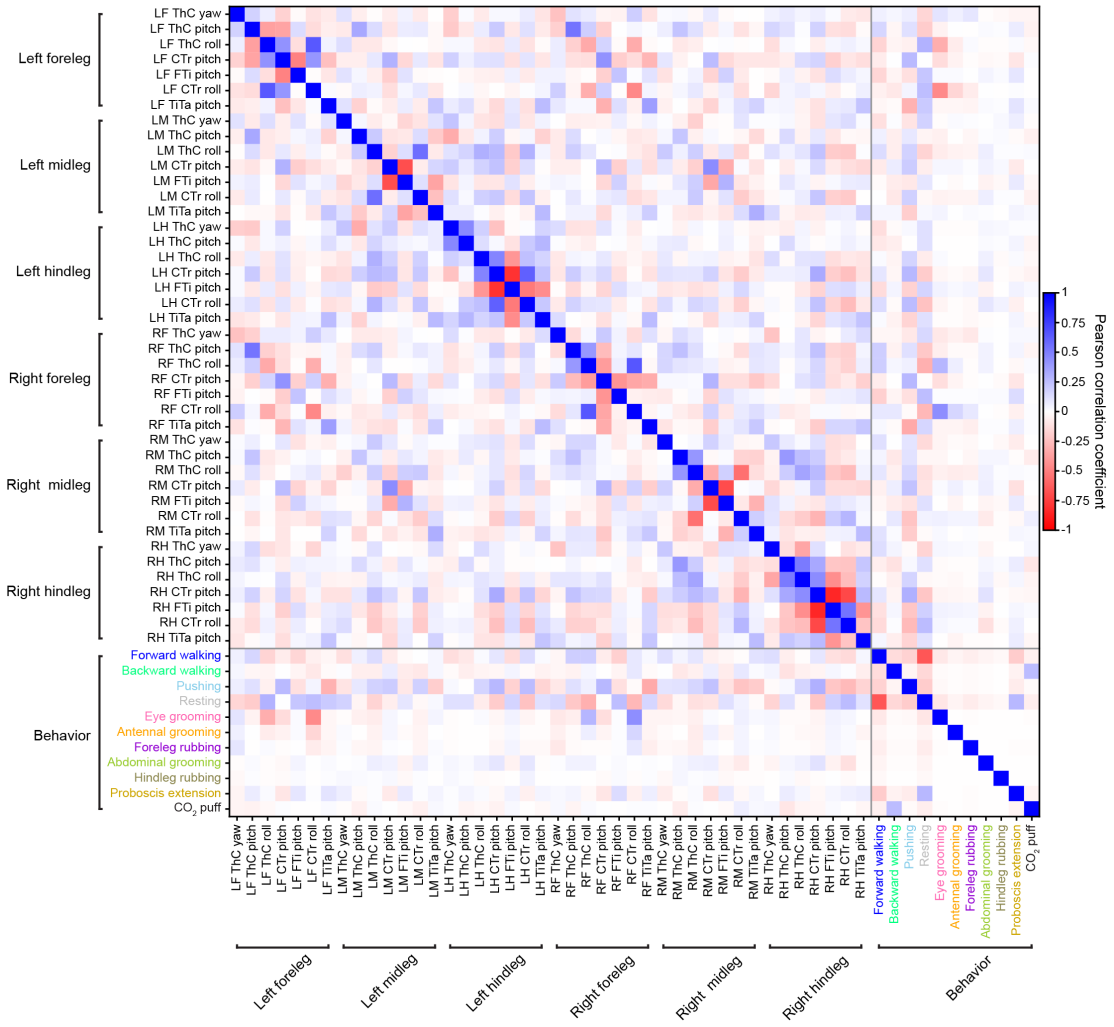

Figure S3.3: **Correlations among and between low-level joint angles and high-level behavioral states. Related to Figure 3.2.** Pearson correlation coefficients (color-coded) for joint angles, high-level behavioral states, proboscis extensions, and puffs.

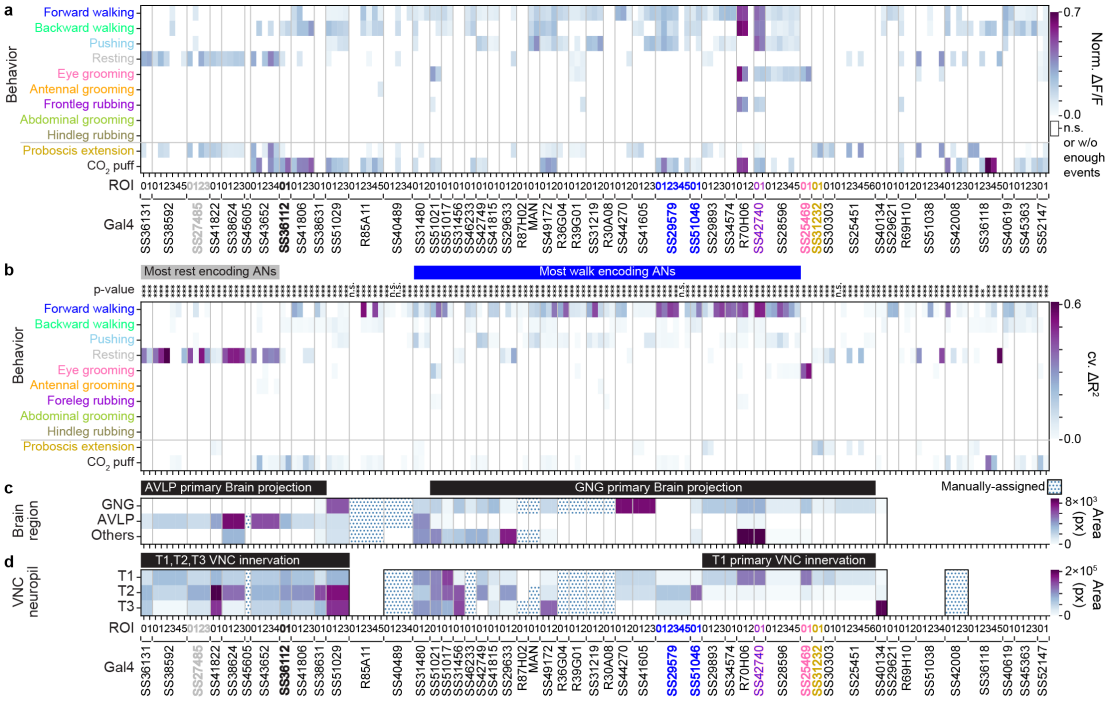

Figure S3.4: **Normalized mean activity ($\Delta F/F$) of ascending neurons during high-level behaviors, and a summary of their behavioral encoding, brain targeting, and VNC patterning. Related to Figure 3.2 and Figure 3.6.** (**a**) Normalized mean $\Delta F/F$, normalized between 0 and 1, for a given AN across all epochs of a specific high-level behavior. Analyses were performed for 157 ANs recorded from 50 driver lines. Note that fluorescence for non-orthogonal behaviors/events may overlap (e.g., for backward walking and puff, or resting and proboscis extensions). Conditions with less than ten epochs longer than 0.7 s are masked (white). ANOVA and posthoc Tukey tests to correct for multiple comparisons were performed to test if values are significantly different from baseline. Non-significant samples are also masked (white). **(b)** Variance in AN activity that can be uniquely explained by a regressor (cross-validated ΔR^2) for high-level behaviors. Non-orthogonal regressors (PE and CO_2 puffs) are separated from the others. *P*-values report the F-statistic of overall significance of the complete regression model with no regressors shuffled (*p<0.05, **p<0.01, and ***p<0.001). (**c,d**) The most substantial AN (**c**) targeting of brain regions, or (**d**) patterning of VNC regions, as quantified by pixel-based analysis of MCFO labelling. Driver lines that were manually quantified are indicated (dotted cells). Projections that could not be unambiguously identified are left blank. Notable encoding and innervation patterns are indicated by bars above each matrix. Lines (and their corresponding ANs) selected for more in-depth analysis are color-coded by the behavioral class that best explains their neural activity: SS27485 (resting), SS36112 (puff responses), SS29579 (walking), SS51046 (turning), SS42740 (foreleg-dependent behaviors), SS25469 (eye grooming), and SS31232 (proboscis extensions).

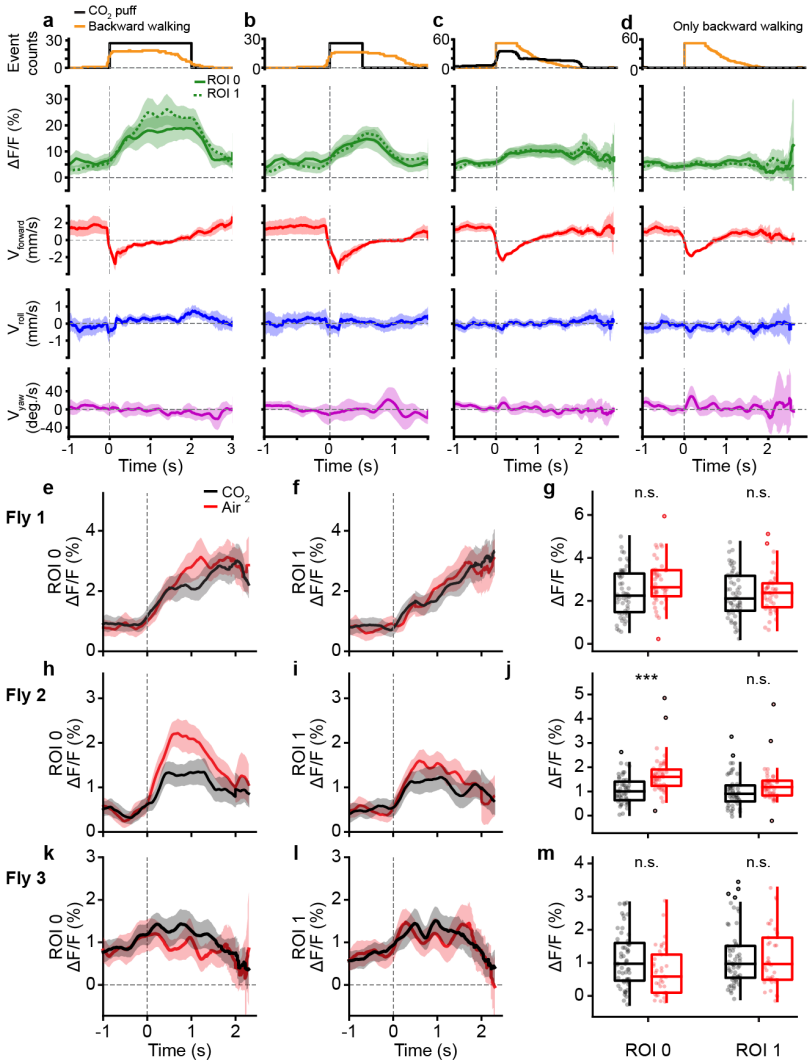

Figure S3.5: **Puff-ANs do not encode backward walking and respond similarly to puffs of air, or CO_2. Related to Figure 3.4. (a-d)** Puff-ANs (SS36112) activity (green) and corresponding spherical treadmill rotational velocities (red, blue, and purple) during (**a**) long, 2 s CO_2-puff stimulation (black) and associated backward walking (orange), (**b**) short, 0.5 s CO_2-puff stimulation, (**c**) periods with backward walking, and (**d**) the same backward walking events as in **c** but only during periods without coincident puff stimulation. Shown are the mean (solid and dashed lines) and 95% confidence interval (shaded areas) of multiple $\Delta F/F$ and ball rotation time-series. (**e-m**) Activity of puff-ANs (SS36112) from three flies (**e-g**, **h-j**, and **k-m**, respectively) in response to puffs of air (red), or CO_2 (black). (**e-f, h-i, k-l**) Shown are mean (solid and dashed lines) and 95% confidence interval (shaded areas) $\Delta F/F$ for ROIs (**e,h,k**) 0 and **(f,i,l)** 1. (**g,j,m**) Mean fluorescence (circles) of traces for ROIs 0 (left) or 1 (right) from 0.7 s after puff onset until the end of stimulation. Overlaid are box plots representing the median, interquartile range (IQR), and 1.5 IQR. Outliers beyond 1.5 IQR are indicated (opaque circles). A Mann-Whitney test (*** p<0.001, ** p<0.01, * p<0.05) was used to compare responses to puffs of CO_2 (red), or air (black).

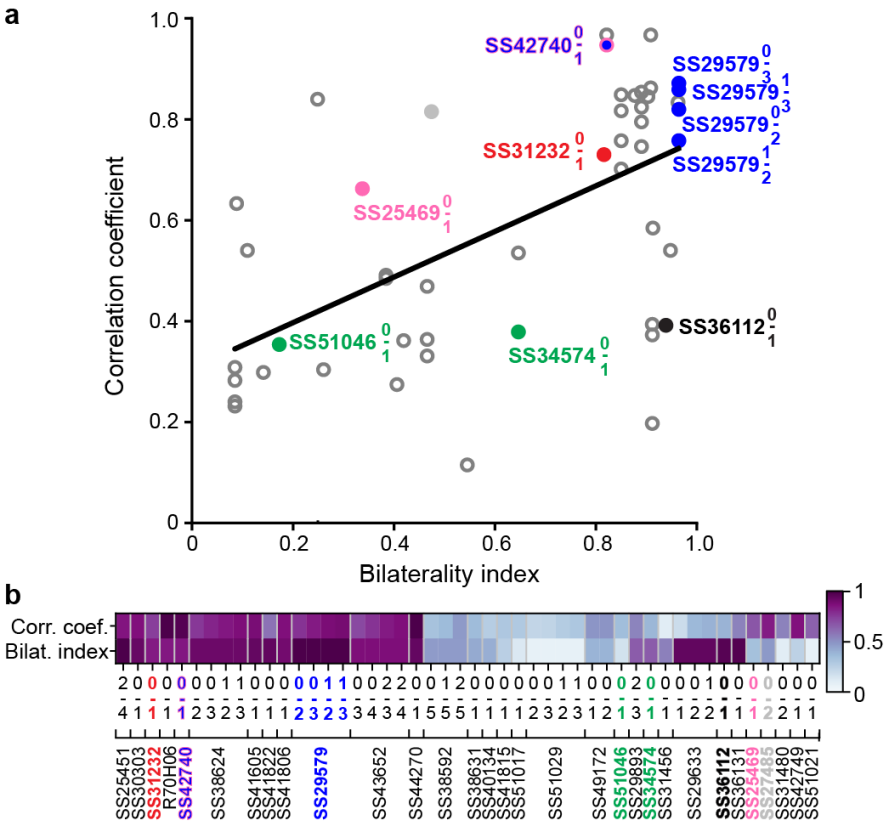

Figure S3.6: **The bilaterality of an ascending neuron pair's VNC patterning correlates with the synchrony of their activity. Related to Figure 3.5.** (**a**) A bilaterality index, quantifying the differential innervation of the left and right VNC (without distinguishing between axons and dendrites) is compared with the Pearson correlation coefficient computed for the activity of left and right ANs for a driver line pair (R^2 = 0.31 and p<0.001 using an F-test). (**b**) Bilaterality index and Pearson correlation coefficient values for each AN pair.

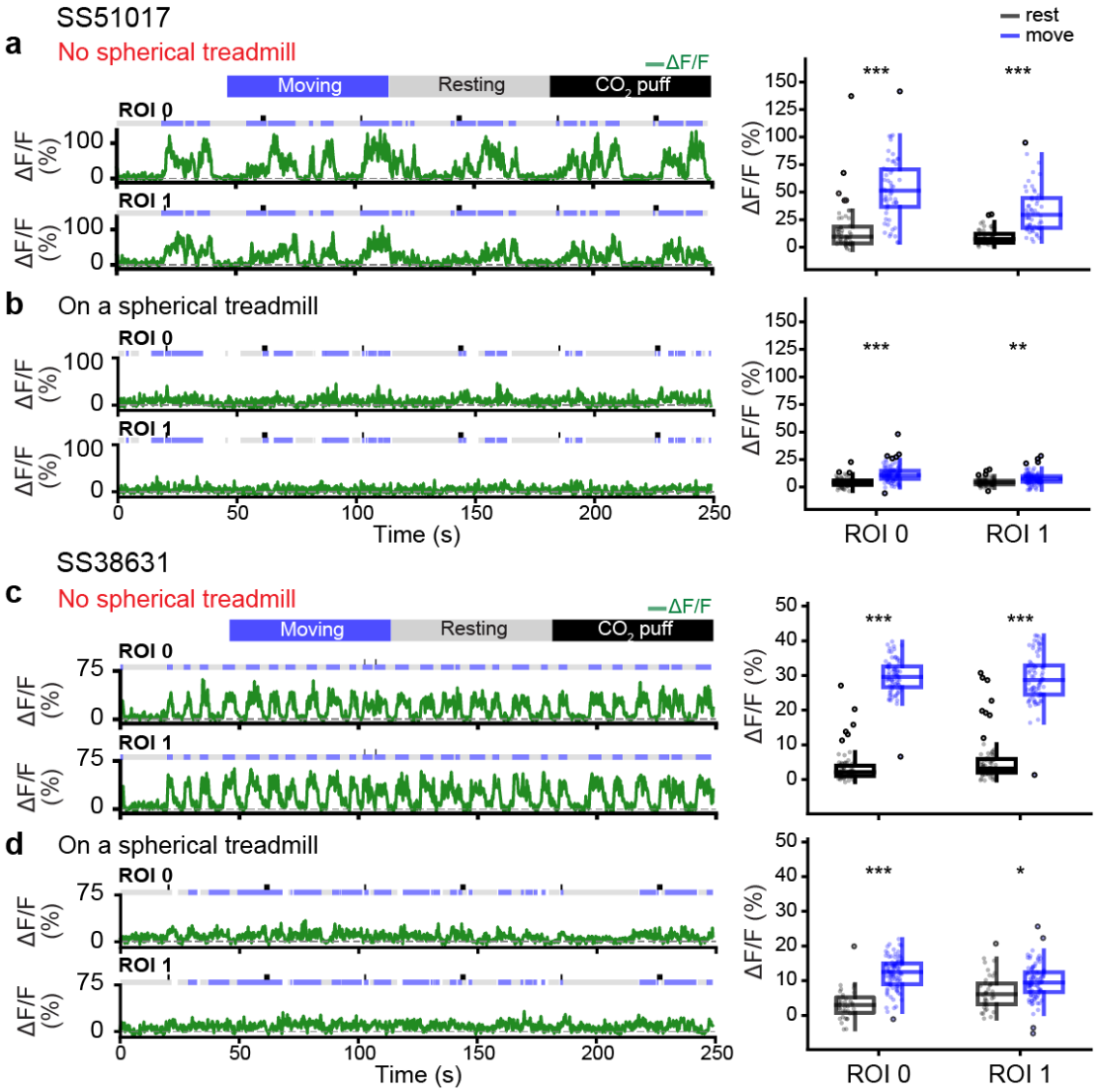

Figure S3.7: **Ascending neurons that become active only when the spherical treadmill is removed.** Representative AN recordings from ROIs 0 and 1 for an (**a,b**) SS51017-spGAL4, or (**c,d**) SS38631-spGAL4 animal measured when it is (**a,c**) suspended without a spherical treadmill, or (**b,d**) in contact with the spherical treadmill. Moving, resting, and puff stimulation epochs are indicated. Shown are (**left**) representative neural activity traces and (**right**) summary data including the median, interquartile range (IQR), and 1.5 IQR of AN $\Delta F/F$ values when the animals are resting (black), or moving (blue). Outliers (values beyond 1.5 IQR) are indicated (opaque circles). Statistical comparisons were performed using a Mann-Whitney test (*** p<0.001, ** p<0.01, * p<0.05).

3.10.2 Supplementary Tables

Table S3.1: Sparse AN driver lines and associated properties.

No.	Gal4	Confocal SNR of smFP	2P SNR of OpGCaMP6f and tdTomato	Number ROIs	ROI#	Encoding	Level synchronous activity (ROI# vs ROI#: Corr. coef.)	Redundant Gal4	Supp. Video#
1	SS36131	strong	strong	2	0 1	rest rest	0 vs 1: 0.7	SS36132 SS36133	3
2	SS38592	strong	strong	6	0 1 2 3 4 5	rest rest rest unclear unclear rest	0 vs 5: 0.4 1 vs 5: 0.36 2 vs 5: 0.49	SS38598	4
3	SS27485	strong	strong	4	0 1 2 3	rest unclearrest rest	0 vs 2: 0.85	-	5
4	SS41822	strong	strong	2	0 1	rest rest	0 vs 1: 0.71	SS41808 SS41809 SS41820 SS41821	6
5	SS38624	strong	strong	4	0 1 2 3	rest rest rest rest	0 vs 2: 0.84 0 vs 3: 0.7 1 vs 2: 0.9 1 vs 3: 0.87	-	7
6	SS45605	strong	strong	1	0	unclear	-	-	8
7	SS43652	strong	medium	5	0 1 2 3 4	rest puff rest rest rest	0 vs 3: 0.91 0 vs 4: 0.84 2 vs 3: 0.77 2 vs 4: 0.89	-	9
8	SS36112	strong	strong	2	0 1	puff puff	0 vs 1: 0.51	-	10
9	SS41806	strong	strong	4	0 1 2 3	unclear unclear puff puff	2 vs 3: 0.67	-	11
10	SS38631	strong	strong	2	0 1	off ball movement off ball movement	0 vs 1: 0.88	-	12
11	SS51029	strong	medium	4	0 1 2 3	puff puffpuff puff	0 vs 2: 0.32 0 vs 3: 0.33 1 vs 2: 0.37 1 vs 3: 0.4	SS51024	13
12	R85A11	strong	strong	6	0 1 2 3 4 5	unresponsive unresponsive walk unclear walk unclear	2 vs 4: 0.87	-	15
13	SS40489	-	strong	5	0 1 2 3 4	unclear unclear unclear unclear unclear	0 vs 3: 0.38 0 vs 4: 0.43 1 vs 3: 0.53 1 vs 4: 0.53	-	16
14	SS31480	strong	strong	3	0 1 2	walk unclear walk	0 vs 2: 0.61	-	17

Table S3.1 continued from previous page

No.	GAL4	Confocal SNR of smFP	2P SNR of OpGCaMP6f and tdTomato	Number ROIs	ROI#	Encoding	Level synchronous activity (ROI# vs ROI#: Corr. coef.)	Redundant GAL4	Supp. Video#
15	SS51021	strong	strong	2	0 1	foreleg movement foreleg movement	0 vs 1: 0.69	-	18
16	SS51017	strong	strong	2	0 1	off ball movement off ball movement	0 vs 1: 0.67	-	19
17	SS31456	strong	strong	2	0 1	off ball movement off ball movement	0 vs 1: 0.20	-	20
18	SS46233	strong	medium	2	0 1	walk walk	0 vs 1: 0.63	-	21
19	SS42749	strong	strong	2	0 1	push push	0 vs 1: 0.89	-	22
20	SS41815	strong	medium	2	0 1	unclear unclear	0 vs 1: 0.42	-	23
21	SS29633	strong	strong	3	0 1 2	unclear unclear unclear	0 vs 1: 0.11 0 vs 2: 0.41 1 vs 2: 0.42	-	24
22	R87H02	strong	strong	2	0 1	unclear unclear	0 vs 1: 0.48	-	25
23	MAN	strong	strong	2	0 1	push push	0 vs 1: 0.82	-	26
24	SS49172	strong	strong	3	0 1 2	walk walk walk	0 vs 1: 0.54 0 vs 2: 0.53	-	27
25	R36G04	strong	strong	2	0 1	walk walk	0 vs 1: 0.55	-	28
26	R39G01	strong	strong	3	0 1 2	unclear unclear unclear	0 vs 1: 0.09 0 vs 2: 0.82 1 vs 2: 0.19	-	29
27	SS31219	strong	strong	3	0 1 2	walk walk walk	0 vs 1: 0.67 0 vs 2: 0.74	-	30
28	R30A08	strong	medium	2	0 1	unclear unclear	0 vs 1: 0.22	-	31
29	SS44270	strong	strong	3	0 1 2	walk walk walk	0 vs 1: 0.98	SS41605	32
30	SS41605	strong	strong	4	0 1 2 3	unclear unclear push push	2 vs 3: 0.94	SS44270	33
31	SS29579	strong	strong	6	0 1 2 3 4 5	walk walk walk walk unresponsive walk	0 vs 2: 0.84 0 vs 3: 0.89 0 vs 3: 0.89 1 vs 2: 0.78 1 vs 3: 0.87	-	34
32	SS51046	medium	strong	2	0 1	turn turn	0 vs 1: 0.48	-	35
33	SS29893	strong	strong	4	0 1 2 3	unclear unclear turn turn	2 vs 3: 0.62	SS34574	36
34	SS34574	strong	strong	2	0 1	turn turn	0 vs 1: 0.44	SS29893	37

Table S3.1 continued from previous page

No.	GAL4	Confocal SNR of smFP	2P SNR of OpGCaMP6f and tdTomato	Number ROIs	ROI#	Encoding	Level synchronous activity (ROI# vs ROI#: Corr. coef.)	Redundant GAL4	Supp. Video#
35	R70H06	strong	strong	3	0 1 2	foreleg movement foreleg movement unresponsive	0 vs 1: 0.98	SS42740 SS42707	38
36	SS42740	strong	strong	2	0 1	foreleg movement foreleg movement	0 vs 1: 0.97	R70H06 SS42707	39
37	SS25469	strong	strong	2	0 1	eye groom eye groom	0 vs 1: 0.75	SS52106 SS52107 SS52108	40
38	SS31232	strong	strong	2	0 1	proboscis extension proboscis extension	0 vs 1: 0.81	SS30303 SS25451	41
39	SS30303		strong	2	0 1	proboscis extension proboscis extension	0 vs 1: 0.89	SS31232 SS25451	42
40	SS25451	strong	strong	7	0 1 2 3 4 5 6	unresponsive unresponsive proboscis extension unresponsive proboscis extension unresponsive unresponsive	2 vs 4: 0.93	SS31232 SS30303	43
41	SS28596	strong	strong	6	0 1 2 3 4 5	foreleg movement foreleg movement foreleg movement foreleg movement foreleg movement foreleg movement	0 vs 2: 0.4 0 vs 3: 0.65 0 vs 4: 0.61 0 vs 5: 0.21 1 vs 2: 0.52 1 vs 3: 0.5 1 vs 4: 0.46 1 vs 5: 0.28	R86H08	44
42	SS40134	strong	medium	2	0 1	unclear unclear	0 vs 1: 0.34	-	45
43	SS29621	strong	strong	2	0 1	walk walk	0 vs 1: 0.74	-	46
44	R69H10	strong	strong	3	0 1 2	unclear unclear unclear	0 vs 1: 0.21 0 vs 2: 0.07 1 vs 2: 0.02	-	47
45	SS51038	strong	strong	5	0 1 2 3 4	rest unclear unclear unresponsive rest	0 vs 2: -0.41 0 vs 4: 0.86 2 vs 4: -0.43	-	48
46	SS42008	strong	strong	4	0 1 2 3	walk unclear walk unclear	0 vs 2: 0.53	SS42007	49
47	SS36118	strong	strong	6	0 1 2 3 4 5	unclear unclear unresponsive puff puff rest	2 vs 4: 0.25 2 vs 5: 0.12 3 vs 4: 0.75 3 vs 5: 0.27	-	50
48	SS40619	medium	strong	2	0	walk	0 vs 1: 0.74	-	51
49	SS45363	strong	strong	4	0 1 2 3	puff puff unclear puff	0 vs 3: 0.83 1 vs 3: 0.85 2 vs 3: 0.71	-	52

Table S3.1 continued from previous page

No.	GAL4	Confocal SNR of smFP	2P SNR of OpGCaMP6f and tdTomato	Number ROIs	ROI#	Encoding	Level synchronous activity (ROI# vs ROI#: Corr. coef.)	Redundant GAL4	Supp. Video#
50	SS52147	medium	medium	2	0 1	puff puff	0 vs 1: 0.32	-	53
51	R38F09	-	strong	10	-	unresponsive	-	-	-
52	SS46269	strong	strong	12	-	unresponsive	-	-	-
53	SS25470	strong	strong	5	-	unresponsive	-	SS48406	-
54	SS25478	strong	strong	5	-	unresponsive	-	-	-
55	SS28382	strong	strong	6	-	unresponsive	-	-	-
56	SS29574	strong	strong	2	-	unresponsive	-	-	-
57	SS31899	strong	strong	8	-	unresponsive	-	-	-
58	SS33380	strong	strong	7	-	unresponsive	-	-	-
59	SS33433	strong	strong	2	-	unresponsive	-	-	-
60	SS38012	strong	strong	6	-	unresponsive	-	SS43528	-
61	SS38386	medium	strong	3	-	unresponsive	-	-	-
62	SS38687	strong	strong	3	-	unresponsive	-	-	-
63	SS46290	medium	medium	2	-	unresponsive	-	-	-
64	SS46300	strong	medium	2	-	unresponsive	-	-	-
65	SS48406	strong	strong	5	-	unresponsive	-	SS25470	-
66	SS48409	strong	strong	2	-	unresponsive	-	SS48632	-
67	SS49982	strong	medium	4	-	unresponsive	-	-	-
68	SS50004	strong	strong	2	-	unresponsive	-	-	-
69	SS50013	medium	medium	1	-	unresponsive	-	-	-
70	SS50652	strong	medium	3	-	unresponsive	-	-	-
71	SS36132	strong	-	-	-	not imaged (redundant)	-	SS36131 SS36133	-
72	SS36133	strong	-	-	-	not imaged (redundant)	-	SS36131 SS36132	-
73	SS38598	strong	-	-	-	not imaged (redundant)	-	SS38592	-
74	SS41808	strong	-	-	-	not imaged (redundant)	-	SS41822 SS41809 SS41820 SS41821	-
75	SS41809	strong	-	-	-	not imaged (redundant)	-	SS41822 SS41808 SS41820 SS41821	-
76	SS41820	strong	-	-	-	not imaged (redundant)	-	SS41822 SS41808 SS41809 SS41821	-
77	SS41821	strong	-	-	-	not imaged (redundant)	-	SS41822 SS41808 SS41809 SS41820	-
78	SS42007	strong	-	-	-	not imaged (redundant)	-	SS42008	-
79	SS42707	medium	-	-	-	not imaged (redundant)	-	SS42740 R70H06	-
80	SS43528	strong	weak	-	-	not imaged (undetectable in 2P; redundant)	-	SS38012	-
81	SS48632	medium	-	-	-	not imaged (redundant)	-	SS48409	-
82	SS51024	strong	-	-	-	not imaged (redundant)	-	SS51029	-

Table S3.1 continued from previous page

No.	GAL4	Confocal SNR of smFP	2P SNR of OpGCaMP6f and tdTomato	Number ROIs	ROI#	Encoding	Level synchronous activity (ROI# vs ROI#: Corr. coef.)	Redundant GAL4	Supp. Video#
83	SS52108	strong	weak	-	-	not imaged (redundant)	-	SS25469 SS52106 SS52107	-
84	SS52106	medium	-	-	-	not imaged (redundant)	-	SS25469 SS52107 SS52108	-
85	SS52107	medium	-	-	-	not imaged (redundant)	-	SS25469 SS52106 SS52108	-
86	R86H08	medium	weak	-	-	not imaged (redundant)		SS28596	-
87	SS29889	strong (indistinguishable brain neurons)	-	-	-	not imaged (redundant)	-	SS29890	-
88	SS29890	strong (indistinguishable brain neurons)	-	-	-	not imaged (redundant)	-	SS29889	-
89	SS29605	-	unreliable expression	-	-	not imaged (unreliable expression)	-	-	-
90	SS31246	unreliable expression	unreliable expression	-	-	not imaged (unreliable expression)	-	-	-
91	SS46696	unreliable expression	unreliable expression	-	-	not imaged (unreliable expression)	-	-	-
92	R75E01	strong but with glia	-	-	-	not imaged (glia included)	-	-	-
93	SS37652	medium	weak	-	-	not imaged (undetectable)	-	-	-
94	SS41602	strong	weak	-	-	not imaged (undetectable)	-	-	-
95	SS43651	strong	weak	-	-	not imaged (undetectable)	-	-	-
96	SS44305	strong	weak	-	-	not imaged (undetectable)	-	-	-
97	SS46255	strong	weak	-	-	not imaged (undetectable)	-	-	-
98	SS41824	strong	weak	-	-	not imaged (undetectable)	-	-	-
99	SS25488	-	weak	-	-	not imaged (undetectable)	-	-	-
100	R81G07	weak	weak	-	-	not imaged (undetectable)	-	-	-
101	SS45635	weak	weak	-	-	not imaged (undetectable)	-	-	-
102	SS45648	weak	weak	-	-	not imaged (undetectable)	-	-	-
103	SS46290	weak	weak	-	-	not imaged (undetectable)	-	-	-
104	SS46847	weak	weak	-	-	not imaged (undetectable)	-	-	-
105	SS47868	weak	weak	-	-	not imaged (undetectable)	-	-	-

Table S3.1 continued from previous page

No.	GAL4	Confocal SNR of smFP	2P SNR of OpGCaMP6f and tdTomato	Number ROIs	ROI#	Encoding	Level synchronous activity (ROI# vs ROI#: Corr. coef.)	Redundant GAL4	Supp. Video#
106	SS50282	weak	weak	-	-	not imaged (undetectable)	-	-	-
107	SS50829	weak	weak	-	-	not imaged (undetectable)	-	-	-
108	R88C08	weak	weak	-	-	not imaged (undetectable)	-	-	-

Table S3.2: Activation (AD) and DNA-binding Domains (DBD) of split-GAL4 lines used in this study.

	Driver line	AD	DBD
1	SS36131	R70D06	VT033054
2	SS38592	VT016458	VT012410
3	SS27485	R75E01	R18B05
4	SS41822	VT033054	VT026646
5	SS38624	VT002081	R85H01
6	SS45605	R15E01	R41E03
7	SS43652	VT026477	R38E07
8	SS36112	VT026646	VT028606
9	SS41806	VT060737	VT028606
10	SS38631	R72A10	VT038208
11	SS51029	VT034810	VT004985
12	R85A11	-	-
13	SS40489	R36B06	VT007767
14	SS31480	R68C10	VT008150
15	SS51021	VT027767	VT027005
16	SS51017	VT005404	VT027767
17	SS31456	VT013500	VT012768
18	SS46233	VT029814	VT028464
19	SS42749	R66A06	VT056770
20	SS41815	VT043377	VT014669
21	SS29633	R33F06	R76E11
22	R87H02	-	-
23	MAN	VT50660	VT14014
24	SS49172	VT049120	VT008188
25	R36G04	-	-
26	R39G01	-	-
27	SS31219	VT045153	VT019074
28	R30A08	-	-
29	SS44270	VT058560	VT033054
30	SS41605	R80A11	VT038205
31	SS29579	VT023828	VT059224
32	SS51046	VT007177	VT057280
33	SS29893	R67F03	VT050658
34	SS34574	VT008537	VT050658
35	R70H06	-	-
36	SS42740	VT037865	VT061717
37	SS25469	VT027704	VT044958
38	SS31232	VT063643	VT059781
39	SS30303	VT063643	VT018278
40	SS25451	VT063643	VT059224
41	SS28596	R94B04	R86H08
42	SS40134	VT028320	R49A01
43	SS29621	R22E07	R30E10
44	R69H10	-	-
45	SS51038	VT030558	VT001497
46	SS42008	VT033469	VT043682
47	SS36118	VT060737	VT026477
48	SS40619	VT021853	VT050234
49	SS45363	VT062587	VT043920
50	SS52147	VT044164	VT040034
51	R38F09	-	-
52	SS46269	VT023490	VT016254
53	SS25470	VT063643	VT048352
54	SS25478	VT025966	VT013121
55	SS28382	R18G02	R49C03
56	SS29574	VT008660	VT043400

Table S3.2 continued from the previous page

	split-GAL4 line	AD	DBD
57	SS31899	R26H04	R46A10
58	SS33380	R19F01	R60A06
59	SS33433	R94D12	VT060731
60	SS38012	R48E02	R93B07
61	SS38386	VT016966	VT046334
62	SS38687	R30A02	VT015159
63	SS46290	VT029750	VT043288
64	SS46300	VT043146	VT000254
65	SS48406	VT048352	VT039769
66	SS48409	VT036302	VT049125
67	SS49982	R77D08	VT029514
68	SS50004	VT017645	VT049348
69	SS50013	VT008992	VT039485
70	SS50652	R60C01	R80B01
71	SS36132	R70D06	VT025996
72	SS36133	R70D06	VT026646
73	SS38598	VT024634	VT016458
74	SS41808	VT060737	VT033054
75	SS41809	R20E05	VT033054
76	SS41820	VT060737	VT025996
77	SS41821	R20E05	VT025996
78	SS42007	VT033469	VT026646
79	SS42707	VT061717	VT045101
80	SS43528	VT025966	R93B07
81	SS48632	VT036302	R93B07
82	SS51024	VT004985	VT034810
83	SS52108	VT063231	R69H06
84	SS52106	VT063231	VT063626
85	SS52107	VT063231	VT021731
86	R86H08	-	-
87	SS29889	R64G04	VT008537
88	SS29890	R64G04	VT050658
89	SS29605	VT019902	VT048942
90	SS31246	VT038171	VT021780
91	SS46696	VT046782	VT008483
92	SS22675	R44G08	R81A04
93	SS22721	R92D09	R92A07
94	R75E01	-	-
95	SS37652	VT040698	VT023490
96	SS41602	R75E01	R74C01
97	SS43651	VT026477	VT039361
98	SS44305	R21E09	VT016966
99	SS46255	R24H02	VT037862
100	SS41824	R20E05	VT026646
101	SS25488	VT029593	VT020527
102	R81G07	-	-
103	SS45635	VT008882	VT014208
104	SS45648	VT008808	VT029814
105	SS46290	VT029750	VT043288
106	SS46847	VT023490	VT039485
107	SS47868	R24H02	VT002064
108	SS50282	VT037554	VT012768
109	SS50829	VT033290	VT027767
110	R88C08	-	-
111	SS22647	R94B04	R66B12
112	SS22377	R12H12	R92A07

3.10.3 Supplement Videos

Video S3.1: **High-level behaviors, their associated 3D poses, and spherical treadmill rotational velocities.** Behaviors were captured from six camera views. Illuminated text (top) indicates the behavioral class being illustrated. Also shown are corresponding 3D poses (bottom-left) and spherical treadmill rotational velocities, proboscis extension (PE) lengths, and puff stimulation periods (bottom-right).
Download Video S3.1

Video S3.2: **Representative data for 50 comprehensively analyzed, AN-targeting sparse driver lines.** Shown are: (**a**) spFP staining, (**b**) a representative two-photon microscope image, (**c**) outline of the associated cervical connective after filling the surrounding bath with fluorescent dye, (**d**) and PE length, puff stimuli, spherical treadmill rotational velocities, and AN (ROI) $\Delta F/F$ traces. Indicated above are regressors for forward walking ('F.W.'), backward walking ('B.W.'), resting ('Rest'), eye grooming ('Eye groom'), antennal grooming ('Ant. groom'), foreleg rubbing ('Fl. rub'), abdominal grooming ('Abd. groom'), hindleg rubbing ('Hl. rub'), and proboscis extension ('PE'). For each driver line, the title indicates 'date-Gal4-reporters-fly#-trial#'.
Download Video S3.2

Videos S3.3 - S3.52: **Representative behavioral videos and AN two-photon imaging data for 50 comprehensively analyzed, AN-targeting sparse driver lines.**
Download Video S3.3 - S3.52

4 Descending neuron population dynamics during odor-evoked and spontaneous limb-dependent behaviors

Florian Aymanns, Chin-Lin Chen, Pavan Ramdya. "Descending neuron population dynamics during odor-evoked and spontaneous limb-dependent behaviors", *bioRxiv*, 2022.06.30.497612, 2022. The article can be found here: https://doi.org/10.1101/2022.06.30.497612

My contribution:

This manuscript was the main focus of my work in the lab. I collected all of the population data, conceptualized the data analysis approach and analyzed all of the data. My contribution includes developing a pipeline that allows us to record from populations of neurons in the cervical connective, making the driver line to target brain neurons, recording from this driver line, development of the olfactometer, and analysis of the data using linear models among other approaches. Furthermore, I wrote the first draft of the manuscript and made the figures. Chin-Lin Chen performed the R65D11 recordings, which were then analyzed by me. Jasper Phelps traced DNx01 in the electron microscopy dataset. Pavan Ramdya conceptualized this project and helped with writing the manuscript and the design of the figures.

4.1 Abstract

Deciphering how the brain regulates motor circuits to control complex behaviors is an important, long-standing challenge in neuroscience. In the fly, *Drosophila melanogaster*, this is accomplished by a population of ~ 1100 descending neurons (DNs). Activating only a few DNs is known to be sufficient to drive complex behaviors like walking and grooming. However, what additional role the larger population of DNs plays during natural behaviors remains largely unknown. For example, they may modulate core behavioral commands, or comprise parallel pathways that are engaged depending on sensory context. We evaluated these possibilities by recording populations of nearly 100 DNs in individual tethered flies while they generated limb-dependent behaviors. We found that the largest fraction of recorded DNs encode walking while fewer are active during head grooming and resting. A large fraction of walk-encoding DNs encode turning and far fewer weakly encode speed. Although odor context does not determine which behavior-encoding DNs are recruited, a few DNs encode odors rather than behaviors. Lastly, we illustrate how one can identify individual neurons from DN population recordings by analyzing their spatial, functional, and morphological properties. These results set the stage for a comprehensive, population-level understanding of how the brain's descending signals regulate complex motor behaviors.

4.2 Introduction

The richness of animal behaviors depend on the coordinated actions of many individual neurons within a population. For example, individual neurons or small networks may compete in a winner-take-all manner to select the next most appropriate motor action [214]. The brain can then convey these decisions to motor circuits via a population of descending neurons (DNs) projecting to the spinal cord of vertebrates, or ventral nerve cord (VNC) of invertebrates. There, DNs axons impinge upon local circuits including central pattern generators (CPGs) that transform DN directives into specific limb or body part movements [101, 105, 215, 216]. DNs make up only about 1 % of brain neurons. Thus, DN population activity represents a critical information bottleneck: high-dimensional brain dynamics must be compressed into low-dimensional commands that efficiently interface with and can be read out by motor circuits. The information carried by individual DNs has long been a topic of interest [63, 65, 96]. Through electrophysiological recordings in large insects, the activities of individual DNs have been linked to behaviors like walking and stridulation [65]. For some DNs links between firing rate and behavioral features like walking speed have also been established [96]. However, how the larger population of DNs coordinate their activities remains unknown.

The fruit fly, *Drosophila melanogaster*, is an excellent model for investigating how DNs regulate behavior. Flies are genetically-tractable, have a rich behavioral repertoire, and also have a numerically small and compact nervous system [217]. There are only thought to between ~ 350 [59] and ~ 550 [79] pairs of DNs in *Drosophila*. Sparse sets of these DNs can be experimentally targeted using transgenic driver lines [59] for functional recordings [76, 87,

90, 95, 171], or optogenetic perturbations [74, 77, 78, 95, 161]. Their functional properties can be understood within a circuit context using emerging connectomics datasets [85, 218]. Thus, building upon foundational work in other insects [63, 65, 96], studies in *Drosophila* can ultimately reveal how identified DNs work collectively to regulate complex behaviors.

Until now, investigations of *Drosophila* have focused on individual or small sets of of DNs. These studies have demonstrated that artificial activation of DN pairs is sufficient to drive complex behaviors including escape [219](giant fiber neurons, GF), antennal grooming [93](antennal Descending Neurons, aDN), backward walking [53](Moonwalker Descending Neurons, MDN), forward walking [75](DNp09), and landing [77](DNp10 and DNp07). These results also suggest a command-like role for some DNs, in that they are both necessary and sufficient to drive particular actions [60].

Although a command-like role for individual DNs is intuitively easy to grasp, it may not translate well toward understanding how natural behavior is coordinated by large DN populations. Notably, a behavioral screen revealed that optogenetic activation of most DNs drives only small changes in locomotor and grooming behaviors [74], rather than a large variety of distinct actions as might be expected if each DN was a command-like neuron. Therefore, DN populations are likely employing additional, alternative control approaches. For example, some groups of DNs might modulate or fine-tune actions primarily driven by other, command-like DNs. The balance between command-like and modulatory roles of DNs may differ for stereotyped versus more flexible behaviors. In line with their role in behavioral modulation, studies in crickets and locusts have demonstrated that changes in the firing rates of some DNs [66, 96] correlate with walking speed and turning [65]. Alternatively, DN subpopulations may represent parallel pathways that are recruited depending on sensory context. For example, different groups of DNs may be differently engaged during odor-evoked versus spontaneously-generated walking [63, 65]. Finally, some DNs may convey raw sensory information to enable feedback control by downstream motor circuits. For example, in addition to discovering DNs active during steering, one recent study also observed DNa02 activity in response to fictive odors in immobile flies [89]. Furthermore, it has been shown that aversive odor can trigger backward walking mediated by moon walker DNs [220]. DNp07 activity has also been observed in non-flying flies in response to visual stimulation [77].

To resolve the degree to which DNs (i) modulate ongoing behaviors, (ii) are recruited depending on sensory context, and/or (iii) convey raw sensory information from the environment, one would need to simultaneously record large numbers of DNs in a behaving animal. Until now, this has been technically difficult for several reasons. First, there has been an absence of tools for selectively genetically targeting DN populations. Additionally, because DN cell bodies and neurites are distributed across the brain [59], relatively invasive [221] volumetric imaging would be required to simultaneously record the activity of many at once. We previously developed a thoracic dissection approach that enables the optical recording of DN axons within the cervical connective in tethered, behaving animals [171, 187]**(Figure 4.1a)**. Here, we combined this imaging approach with genetic tools [222] that restrict the expression

of neural activity reporters to the brain to record populations of nearly 100 DNs in individual tethered, behaving flies. During these recordings, we presented olfactory stimuli and acquired behavioral data for 3D pose estimation [37], as well as fictive locomotor trajectories [36].

Using these tools, we could test the extent to which DN population activity patterns are consistent with roles in behavior modulation **(Figure 4.1b)**, context-dependence recruitment **(Figure 4.1c)**, and/or raw sensory signaling **(Figure 4.1d)**. We observed that the largest fraction of DNs are active during walking. Principal component analysis revealed diverse neural dynamics across epochs of walking. This variability reflected partially overlapping subsets of DNs that encode turning and, to a far weaker extent, speed. These data support a role for DN populations in behavioral modulation. DNs are active during walking or grooming irrespective of whether the behavior was generated spontaneously or during olfactory stimulation. These data rule out a strong context dependence for the recruitment of DN subpopulations. However, we did find that some DNs are specifically responsive to odors, revealing that motor circuits have access to surprisingly unfiltered sensory information. Finally, we illustrate how one can identify DNs from population recordings **(Figure 4.1e)**. We studied a prominent pair of DNs that are asymmetrically active during antennal grooming. By using their topological and encoding properties, we could identify a sparse driver line that targets these neurons used morphological analysis (MCFO and connectomics) to confirm their identities as DNx01 neurons originating from the antennae [59]. These data provide a first, global view of DN population activity during natural behaviors and open the door to a comprehensive mechanistic understanding of how the brain's descending signals regulate motor control.

4.3 Results

4.3.1 Recording descending neuron population activity in tethered, behaving *Drosophila*

To selectively record the activity of populations of DNs, we devised an intersectional genetic and optical approach. First, we restricted the expression of GCaMP6s (a fluorescent indicator of neural activity [19]) and tdTomato (an anatomical fiduciary) to the supraesophageal zone of the brain [222](*;otd-nls:FLPo ; R57C10-GAL4, tub>GAL80>*). We confirmed that transgene expression was restricted to cell bodies in the brain **(Figure 4.1f)**. The axons of DNs could be seen passing through the cervical connective, targeting motor circuits within the ventral nerve cord (VNC) **(Figure 4.1g)**. Thus, although our driver line lacks expression in the subesophageal zone (SEZ) **(Figure S4.1c)**—a brain region known to house at least 41 DNs [59, 83] some of which can drive grooming [78](DNg11, DNg12)—we could still capture the activities of a large population of DNs. Second, by performing coronal (x-z) two-photon imaging of the thoracic cervical connective [171], we could exclusively record DN axons **(Figure 4.1h)** in tethered animals that were behaving on a spherical treadmill. This imaging approach could also compensate for image translations during animal behavior, keeping regions-of-interest (ROIs) within the field-of-view (FOV). We then applied image registration [171]—to correct

for translations and deformations—as well as image denoising [223]—to obtain higher signal-to-noise images. We confirmed that denoising does not systematically delay or prolong the temporal dynamics of neural activity **(Figure S4.1a,b)**. Resulting images clearly showed elliptical ROIs that likely represent individual large axons or possibly also tightly packed groups of smaller axons **(Figure 4.1i)**. From here on we will interchangeably refer to ROIs as DNs or neurons. From these data, we calculated $\Delta F/F$ images **(Figure 4.1j)** within which we could manually label 75-95 of the most distinct and clearly visible ROIs in each animal. This resulted in high-dimensional neural time-series **(Figure 4.1k)**.

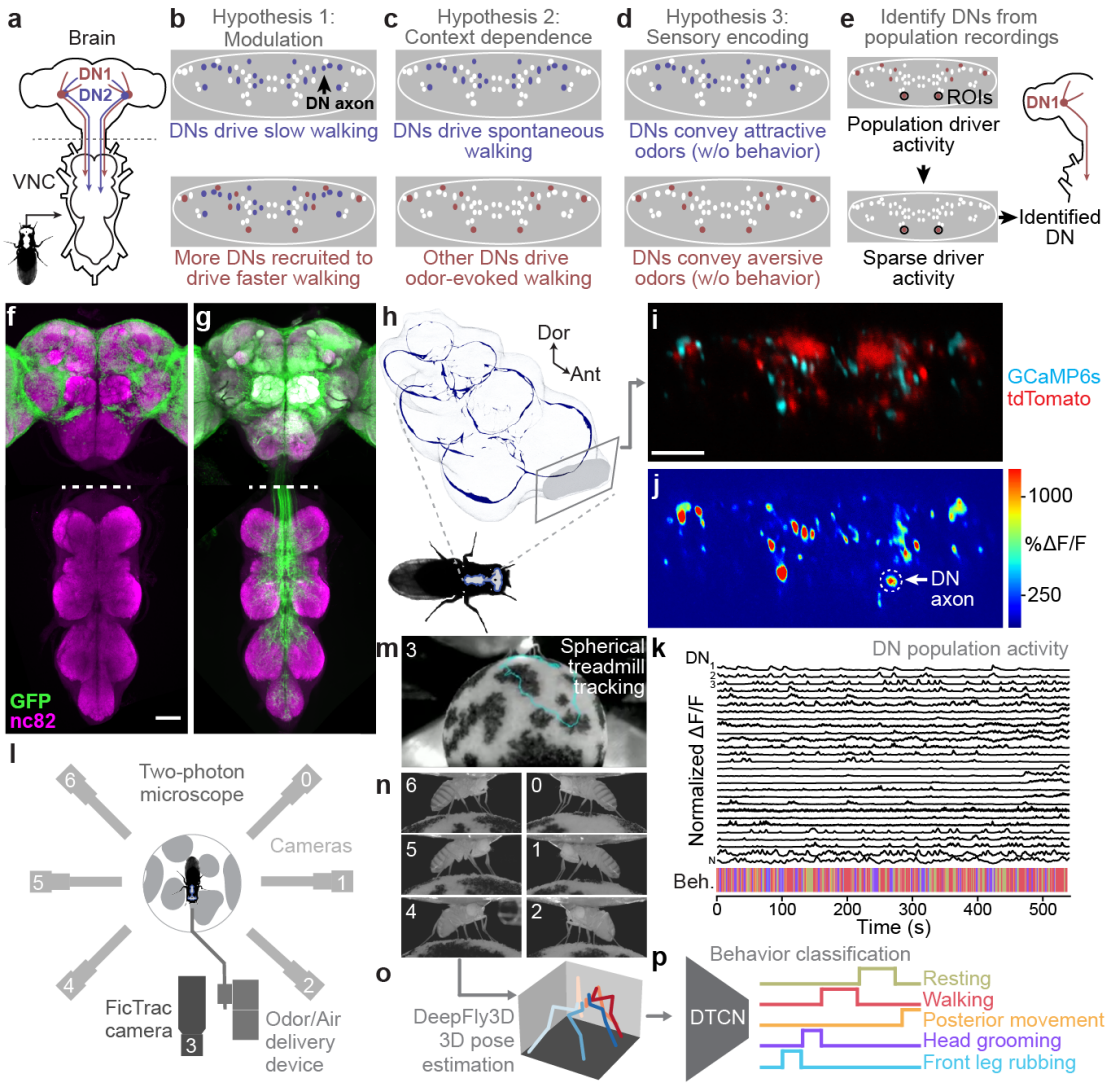

Figure 4.1: **Recording descending neuron population activity and animal behavior. (a)** Schematic of the *Drosophila* nervous system showing descending neurons (DNs) projecting from the brain to motor circuits in the ventral nerve cord (VNC). Only two pairs of DNs (red and blue) are shown for clarity. Indicated (dashed gray line) is the coronal imaging region-of-interest in the thoracic cervical connective. *(legend continued on next page)*

(b) In a 'modulation' framework for DN population control, new DNs (red) may be recruited to modulate ongoing behaviors primarily driven by core DNs (blue). Each ellipse is an individual DN axon (white, blue, and red). **(c)** In a 'context dependence' framework for DN population control, different DNs may be recruited to drive identical behaviors depending on sensory context. **(d)** Alternatively, in a 'sensory encoding' framework, many DNs may not drive behaviors but rather transmit sensory signals to the VNC. **(e)** An approach for identifying DNs from population recordings. One may first identify sparse transgenic strains labeling specific neurons from DN populations (circled in black) using their functional attributes/encoding, positions within the cervical connective, and axon shapes. Then, one can use sparse morphological data to find corresponding neurons in the brain and VNC connectome. **(f,g)** Template-registered confocal volume z-projections illustrating a 'brain only' driver line (otd-nls:FLPo; R57C10-GAL4,tub>GAL80>) expressing **(f)** a nuclear (histone-sfGFP) or **(g)** a cytosolic (smGFP) reporter. Scale bar is 50 µm. Location of two-photon imaging plane in the thoracic cervical connective is indicated (white dashed lines). Tissues are stained for GFP (green) and neuropil ('nc82', magenta). **(h)** Schematic of the VNC illustrating the coronal (x-z) imaging plane. Dorsal-ventral ('Dor') and anterior-posterior ('Ant') axes are indicated. **(i)** Denoised two-photon image of descending neuron (DN) axons passing through the thoracic cervical connective. Scale bar is 10 µm. **(j)** Two-photon imaging data from panel **i** following motion correction and $\%\Delta F/F$ color-coding. An ROI (putative DN axon) is indicated (white dashed circle). **(k)** Sample normalized $\%\Delta F/F$ time-series traces for 28 (out of 95 total) DN axons recorded from one animal. Behavioral classification at each time-point is indicated below, color-coded is as in panel **p**. **(l)** Schematic of system for recording behavior and delivering odors during two-photon imaging (not to scale) while a tethered fly walks on a spherical treadmill. **(m)** Spherical treadmill ball rotations (fictive walking trajectories) are captured using the front camera and processed using FicTrac software. Overlaid (cyan) is a sample walking trajectory. **(n)** Video recording of a fly from six camera angles. **(o)** Multiview camera images are processed using DeepFly3D to calculate 2D poses and then triangulated 3D poses. These 3D poses are further processed to obtain joint angles. **(p)** Joint angles are input to a dilated Temporal Convolutional Network (DTCN) to classify behaviors including walking, resting, head (eye and antennal) grooming, front leg rubbing, or posterior (abdominal and hindleg) movements.

To test the context-dependence and sensory feedback encoding of DNs, we built an olfactometer that could sequentially present humidified air, and one of two alternating odors: apple cider vinegar ('ACV', an attractive odorant [224]), or methyl salicylate ('MSC', a putatively aversive odorant [173]). During experiments, we alternated presentation of ACV and MSC with humid air. We performed photoionization detector (PID) measurements to confirm that our olfactometer could deliver a steady flow of air/odor with minimal mechanical perturbations during switching **(Figure S4.1d-f)**.

Along with neural recordings and odor delivery, we quantified limb and joint positions by recording tethered animals from six camera angles synchronously at 100 frames-per-second (fps) **(Figure 4.1l)**. A seventh, front-facing camera recorded spherical treadmill rotations that could be converted into fictive locomotor trajectories using FicTrac [36] **(Figure 4.1m)**. Multi-

view camera images **(Figure 4.1n)** were post-processed using DeepFly3D [37] to estimate 3D joint positions [32] **(Figure 4.1o)**. These data were used to train a dilated temporal convolutional neural network (DTCN) [42] that could accurately classify epochs of walking, resting, head (eye and antennal) grooming, front leg rubbing, and posterior (hindleg and abdominal) movements **(Figure 4.1p, Figure S4.1g)**. Animals predominantly alternated between resting, walking, and head grooming with little time spent front leg rubbing or moving their posterior limbs and abdomen **(Figure S4.1h)**. Notably, we also observed structure in our behavioral data: flies were more likely to walk after resting or generating posterior movements **(Figure S4.1i)**. Flies also frequently performed front leg rubbing following head grooming [121] **(Figure S4.1i)**. Taken together, this experimental and computational pipeline yielded a rich dataset of DN population activity and associated odor-evoked and spontaneous behaviors **(Video 1)**.

4.3.2 Encoding of behavior in descending neuron populations

With these data, we first asked to what extent DN populations encode—and presumably regulate—each of our classified behaviors: resting, walking, head grooming, posterior movements, and front leg rubbing. We addressed this question in two ways. First, we asked how well each behavior could be predicted based on the activity of each neuron. Specifically, we used a linear model to quantify the extent to which a given DN's activity could explain the variance of (i.e., 'encode') each behavior **(Figure 4.2a,b)**. The largest fraction (~60%) of DNs encode walking. The second largest group of DNs encode head grooming (~15%). Only a very small fraction of DNs encode resting and no neurons encode front leg rubbing, or posterior movements **(Figure 4.2c)**. However, some of these behaviors were very infrequent (posterior movements) or of short duration (front leg rubbing) **(Figure S4.1h)**, weakening the power of our analysis in these cases. As well, although none of the DNs *best* explained posterior movements and front leg rubbing out of all behaviors, we observed that DNs encoding walking also encoded posterior movements. Similarly, DNs encoding head grooming also encoded front leg rubbing **(Figure 4.2b)**. This may be due to the strong sequential occurrence of these pairs of behaviors **(Figure S4.1i)** and the long decay time constant of GCaMP6s (~1 s [19]) resulting in elevated calcium signals for the subsequent behavior in the pair. To resolve the extent to which these DNs truly encode one behavior versus the other, we performed a more narrow linear regression analysis. We used equal amounts of data from this pair of sequential behaviors, and calculated the neural variance (rather than the behavioral variance) uniquely explained by these two behaviors alone. These analyses confirmed that DNs predominantly encode walking rather than posterior movements **(Figure S4.2a)**, and head grooming rather than front leg rubbing **(Figure S4.2b)**.

We next asked to what extent DNs encode—and presumably drive—the kinematics of joints, limbs, or limb pairs rather than behaviors. To test this possibility, we quantified how much better neural activity could be predicted from joint angles rather than from behavior. Specifically, we computed the amount of variance in DN activity that could be uniquely explained by subgroups of joint angles, but not behavior categories or any of the remaining joint angles.

Separating joint angles into groups (all, pairs, or individual legs) allowed us to probe the possibility that some neurons might control pairs of legs or individual legs and helped to mitigate the effect of correlations between joint angles within individual legs. Because of the rapid movements of each leg (5 Hz to 20 Hz [116, 225]) and the long decay time of our calcium indicator, we also convolved joint angle and behavior regressors with a calcium response function (crf) kernel. We found that joint angles can only very marginally improve the prediction of neural activity beyond simply using behavior regressors **(Figure S4.4)**. Thus, DN populations largely encode high-level behaviors, suggesting that they delegate low-level kinematic control to downstream circuits in the VNC.

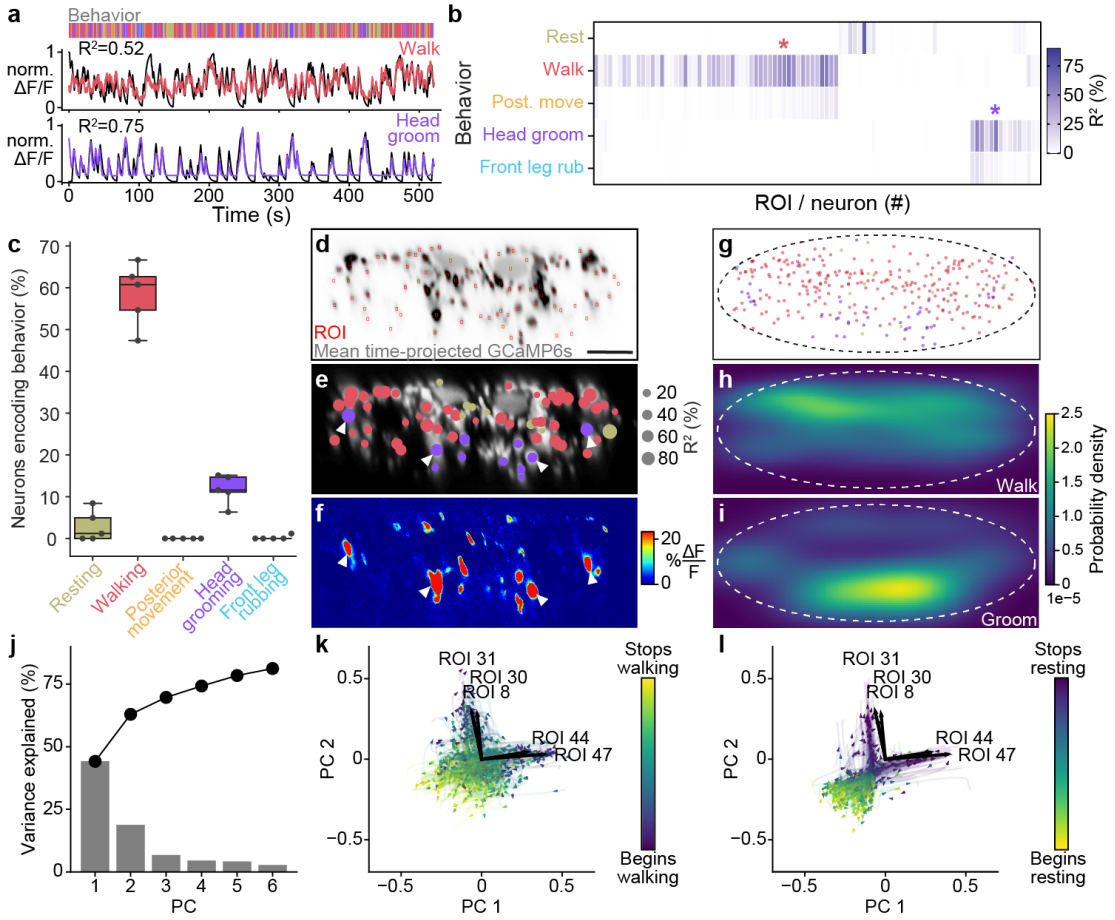

Figure 4.2: **Encoding of behavior in descending neuron populations. (a)** Shown for walking (top) and head grooming (bottom) are the activity (normalized and cross-validation predicted $\Delta F/F$) of individual walk- and head-groom encoding DNs (red and purple lines), as well as predicted $\Delta F/F$ traces from convolving binary behavior regressors with a crf (black lines). The output of the behavior classifier is shown (color bar). **(b)** The cross-validation mean of behavioral variance explained by each of 95 DNs from one animal. Colored asterisks are above the two DNs illustrated in panel **a**. **(c)** The percentage of DNs encoding each classified behavior across five animals. Box plots indicate the median, lower, and upper quartiles. Whiskers signify 1.5 times the interquartile range. *(legend continued on next page)*

(d) Mean time projection of GCaMP6s fluorescence over one nine minute recording. Image is inverted for clarity (high mean fluorescence is black). Manually identified DN regions-of-interest (ROIs) are shown (red rectangles). Scale bar is 10 µm. Panels **d-i** share the same scale. **(e)** DNs color-coded (as in panel **c**) by the behavior their activities best explain. Radius scales with the amount of variance explained. Prominent head groom-encoding neurons that are easily identified across animals are indicated (white arrowheads). **(f)** Behavior-triggered average $\Delta F/F$ image for head grooming. Prominent head grooming DNs identified through linear regression in panel **e** are indicated (white arrowheads). **(g)** Locations of DNs color-coded by the behavior they encode best. Data are from five animals. **(h,i)** Kernel density estimate based on the locations of **(h)** walking or **(i)** head grooming DNs in panel **g**. **(j)** Amount of variance explained by the principal components (PCs) of neural activity derivatives during walking. **(k,l)** Neural activity data during **(k)** walking and **(l)** resting evolve on two lobes. The PC embedding was trained on data taken during walking only. Colored lines indicate individual epochs of **(k)** walking and **(l)** resting. Time is color coded and the temporal progressions of each epoch is indicated (arrowheads). Note that color scales are inverted to match the color at transitions between walking and resting. Black arrows indicate ROIs with high PC loadings. ROI number corresponds to the matrix position in panel **b**. For their locations within this fly's cervical connective, see **Figure S4.5d**.

4.3.3 The spatial organization of descending neuron encoding

Previous morphological analyses demonstrated a clear organization of *Drosophila* DN projections within the VNC [59]. This is likely linked to the distinct functional partners of DNs which regulate limb-dependent (e.g., walking and grooming) versus wing-dependent (e.g., flight and courtship display) behaviors. To further explore the relationship between function and topology, we next asked to what extent we might observe a relationship between a DN's encoding of behavior and its axon's position within the cervical connective **(Figure 4.2d)**. We found that DNs encoding walking are spread throughout the dorsal connective **(Figure 4.2e,g,h)**. On the other hand, head groom-encoding DNs are predominantly in the ventral connective **(Figure 4.2g,i)** including two prominent pairs—lateral and medial-ventral **(Figure 4.2e, white arrowheads)**—whose activities explain the largest amount of variance in head grooming across animals **(Figure S4.3c)**. Surprisingly, we also observed DNs which encode resting. These were located medially, close to the giant fibers, as well as in the lateral extremities of the connective **(Figure 4.2e, olive circles)**. We speculate that rest-encoding DNs may suppress other behaviors or could actively drive the tonic muscle tone required to maintain a natural posture.

We next performed a complementary analysis to further examine the functional-topological organization of DNs in the connective. Specifically, we generated behavior-triggered averages of $\Delta F/F$ images. The results of this approach were only interpretable for frequently occurring behaviors so here we focused on walking, resting, and head grooming **(Figure S4.1h)(Videos 2-4)**. Across animals, we consistently observed DNs encoding walking in the dorsal connective,

and two pairs of ventral DNs encoding head grooming **(Figure 4.2f; Figure S4.3d)**. By contrast, rest-encoding DNs were located in less consistent locations within the connective. These findings confirm that DN populations for walking and head grooming are largely spatially segregated, a feature that may facilitate the identification of specific cells from DN population recordings across animals.

4.3.4 Descending neuron population dynamics suggest more nuanced feature encoding

Thus far we have observed that DN subpopulations encode distinct behaviors and that, by far, the largest fraction encode walking. However, locomotion is not monolithic but rather continuously varies in speed and direction within a single walking trajectory. Thus, the large number of DNs active during walking may represent an aggregate of subpopulations that are differently engaged during distinct locomotor modes. To address this hypothesis, we first closely examined the temporal structure of DN population activity dynamics only during walking epochs. We asked to what extent there is variability and structure in population activity that could potentially support the modulatory encoding of walking speed, forward/backward, or turning.

As in a similar analysis of *C. elegans* population dynamics [226], we calculated the temporal derivative of each DN's activity and then performed principal component analysis (PCA) on these time-series. We found that the first two PCs can explain upwards of 60 % of the variance in DN population activity during walking **(Figure 4.2j; Figure S4.5a)**. Visualizing 2D trajectories (PC 1 and PC 2 subspace) of DN activity during individual walking bouts revealed that it moved primarily along two directions **(Figure 4.2k; Figure S4.5b)**. These directions were even more clear for resting data (embedded within the same PC space) just before the fly began to walk **(Figure 4.2l; Figure S4.5c)**.

To identify individual DNs that most heavily influence this dynamical divergence, we next found those with the largest PC loadings. These neurons' activities most strongly influence the position of population activity in the PC space **(Figure 4.2k,l; Figure S4.5b,c)**. Consistently, also in flies with a less clear divergence in neural trajectories, we found subsets of DNs whose activities correspond to one of these two directions. By examining the positions of their axons, we observed that they are spatially segregated on opposite sides of the connective **(Figure S4.5d, cyan arrowheads)**.

4.3.5 Descending neurons that encode walking include spatially segregated turn-encoding clusters

The divergence of population dynamics and spatial segregation of associated neurons led us to hypothesize that subsets of walk-encoding DNs might be preferentially become active during left and right turning. Alternatively, they might encode fast versus slow walking speeds [75].

Studies in other insects and vertebrates have shown that DNs can play a modulatory role by regulating turning and speed during locomotion [65, 101]. As well, in *Drosophila*, the activation of DNp09 neurons can drive forward walking [75]. Recordings from sparse sets of DNa01 [89, 171] and DNa02 [89] neurons have also shown turn encoding (i.e., steering).

Therefore, we next tested if variability in the activity of DNs might reflect fine-grained encoding of turning and speed during forward walking. We did not analyze backward walking due to its scarcity **(Figure S4.6a)**, brevity **(Figure S4.6b)** and minimal dynamic range **(Figure S4.6c)** in our experimental data. Specifically, we quantified the degree to which DN population activity could uniquely explain yaw (turning) or pitch (speed) angular velocity of the spherical treadmill. Both of these time-series data were convolved with a crf to account for slow calcium indicator decay dynamics. To capture information about turning and walking speed that could not simply be explained by whether the fly was walking or not, we compared the explained variance of our neuron-based ridge regression to a model predicting these features from just a binary walking regressor and shuffled neural data **(Figure 4.3a)**. Neural activity could explain a great deal of variance in turning and, to a lesser extent, speed. Here, the absence of speed encoding may be because the binary walking regressor alone can partially predict speed variance from transitions between resting and walking ($R^2 = 25\,\%$). Therefore, we refined our analysis by only using data from walking epochs and calculating the amount of turning and speed variance that could be explained using neural activity. In this manner, we confirmed that neural activity can uniquely explain turning to a greater extent ($\sim 60\,\%$) than it can explain walking speed ($\sim 30\,\%$) (**Figure 4.3b**).

We next investigated which individual or groups of neurons contribute to the prediction of turning and walking speed. To do this, we only used the activity of one neuron at a time in our model (by contrast, in the paragraph above we used all neurons). Among DNs that encode walking, we found specific DNs that strongly explain right or left turning. By contrast, a more distributed set of DNs weakly encode walking speed **(Figure 4.3c; Figure S4.7a)**. Having identified clusters of DNs encoding right ('red') and left ('blue') turning, we next investigated whether there might be groups encoding other walking features. Among neurons that best explain walking, we again observed clusters for turning but no prominent clusters for speed **(Figure 4.3d; Figure S4.7b)**. This was also reflected in the amount of variance explained: across animals, some walk-encoding DNs also strongly encoded turning **(Figure 4.3e)**, whereas these DNs only encoded a tiny fraction of the variance in walking speed **(Figure 4.3f)**.

Simple models for locomotor control [227] suggest that turning can be controlled by the relative activities of DNs on one side of the brain versus the other. This is supported by a study showing asymmetric activation of *Drosophila* MDNs [131]. Alternatively, the spatial asymmetry in neural activity required for turning might arise in circuits downstream of DNs within the VNC (i.e., with no spatial asymmetry in DN activity). To distinguish between these possibilities, we examined the spatial location of turn-encoding DNs in the cervical connective. We found both ipsi- and contralateral turn-encoding DNs on both sides of the connective **(Figure 4.3g; Figure S4.7c)**, but a clear ipsilateral enrichment **(Figure 4.3h; Figure S4.7d)**.

Many of the DNs encoding turning had high PC loading **(Figure S4.5d)** revealing that turning contributes heavily to the variance in DN population dynamics during walking. By contrast, DNs encoding walking speed were more homogeneously distributed across the connective with no clear spatial enrichment **(Figure 4.3i,j; Figure S4.7e,f)**. Overall, these data support the notion that, during walking, DN population activity largely varies due to turn-related modulation rather than shifts in walking speed.

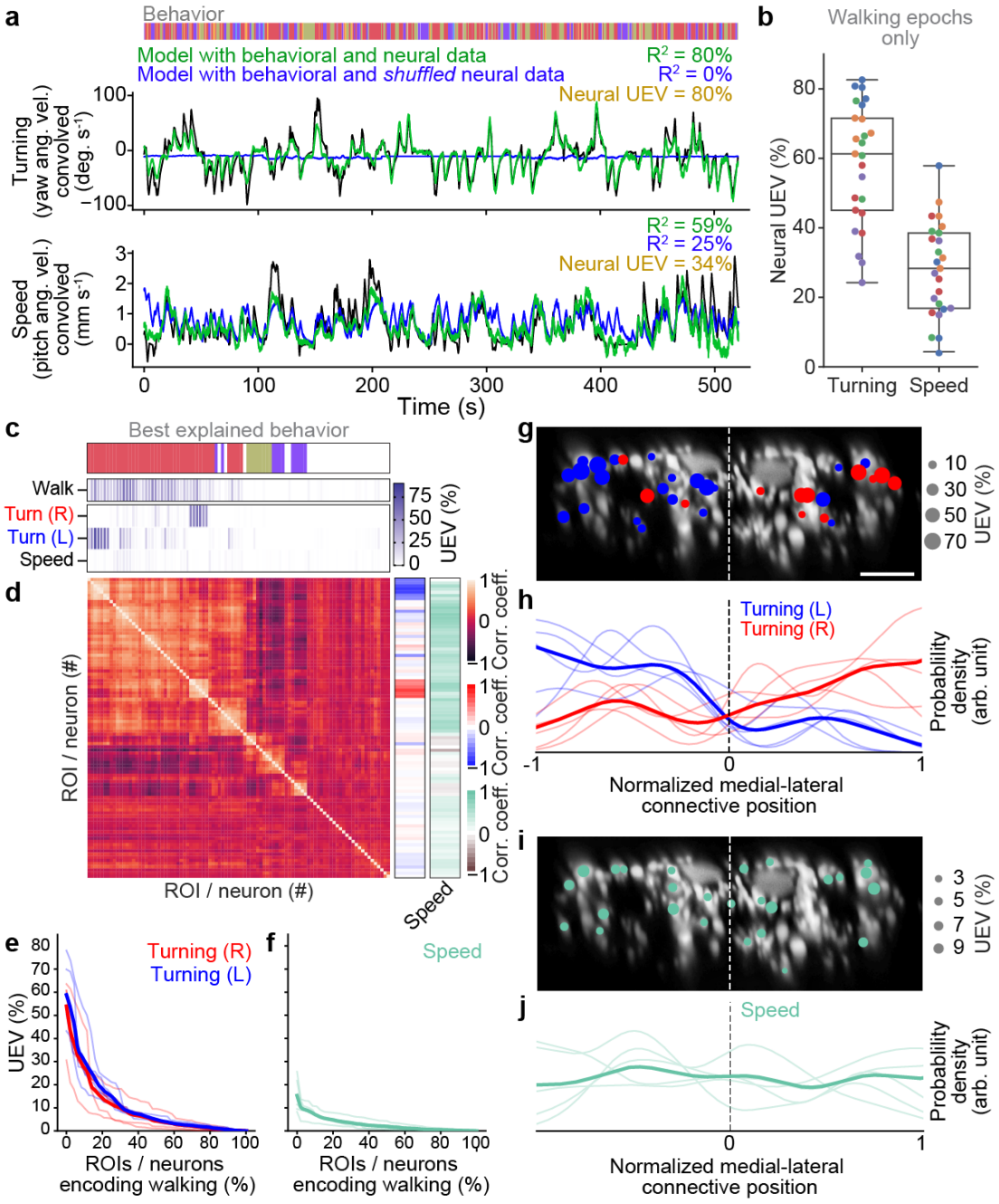

Figure 4.3: See Figure Legend on next page.

Turning and speed encoding in descending neuron populations. **(a)** Predictions of (top) turning and (bottom) walking speed modeled using convolved behavior regressors and all neurons in one animal. Shown are predictions (green) with all regressors intact or (blue) with neural data shuffled across time. Indicated are R^2 values obtained by comparing predicted and real (black) turning and walking speed. These are subtracted to obtain neural unique explained variance (UEV). The fly's behavior throughout the recording is indicated (color bar). **(b)** UEV obtained only using data taken during walking, thus accounting for trivial explanations of speed and turning variance resulting from transitions between resting and walking. Shown are data from five trials each for five flies (color-coded). **(c)** UEV of each DN from one animal for walking speed or left and right turning ordered by clustering of Pearson's correlation coefficients in panel **d**. Walking R^2 values are the same as in **Figure 4.2b** but reordered according to clustering. The models for turning and walking speed were obtained using behavior regressors as well as neural activity. To compute the UEV, activity for a given neuron was shuffled temporally. The behavior whose variance is best explained by a given neuron is indicated (color bar). **(d)** Pearson's correlation coefficient matrix comparing neural activity across DNs ordered by clustering. Shown as well are the correlation of each DN's activity with right, left, and forward walking (right). **(e,f)** UEV for **(e)** turning or **(f)** speed for DNs that best encode walking. Neurons are sorted by UEV. Shown are the distributions for individuals (translucent lines), and the mean across all animals (opaque line). **(g)** Locations of turn-encoding DNs (UEV > 5%), color-coded by preferred direction (left, blue; right, red). Circle radii scale with UEV. Dashed white line indicates the approximate midline of the cervical connective. Scale bar is 10 µm for panels **g** and **i**. **(h)** Kernel density estimate of the distribution of turn encoding DNs. Shown are the distributions for individuals (translucent lines), and the mean distribution across all animals (opaque lines). Probability densities are normalized by the number of DNs along the connective's medial-lateral axis. **(i)** Locations of speed encoding DNs (UEV > 2%). Circle radii scale with UEV. Dashed white line indicates the approximate midline of the cervical connective. **(j)** Kernel density estimate of the distribution of speed encoding DNs. Shown are the distributions for individuals (translucent lines), and the mean distributions across all animals (opaque lines). Probability densities are normalized by the number of DNs along the connective's medial-lateral axis.

4.3.6 Descending neurons are active during behaviors irrespective of olfactory context

Beyond a modulatory role, the large number of DNs active during walking could reflect context dependence: specific subpopulations may only be engaged as a function of sensory context. The possibility of recruiting separate pools of DNs for walking is supported by the observation that an attractive odor, apple cider vinegar (ACV), decreases resting and increases walking **(Figure S4.8a)**, increases forward walking speed **(Figure S4.8b)**, and reduces turning **(Figure S4.8c,d)**.

To address the extent to which subgroups of DNs are recruited depending on olfactory context, we studied the amount of walking or head grooming variance explained by each DN using

only data acquired during exposure to either humidified air, apple cider vinegar (ACV), or methyl salicylate (MSC)—rather than analyzing all walking epochs as in our previous analysis. Humidified air data were subsampled to match the smaller amount of data available for ACV and MSC presentation. We found that largely the same DNs were highly predictive of walking **(Figure S4.9a)** and head grooming **(Figure S4.9b)** irrespective of olfactory context. Only a very small fraction of DNs were differentially recruited during odor presentation **(Figure S4.9a,b, black asterisks)**. Of these four DNs with different recruitment, three achieve significance because they have only a few values distinct from zero in a single trial. The overall explained variance is also very small **(Figure S4.9c)**. These data suggest that changing odor context alters action selection and locomotor kinematics but does not shift the identity of active DN subpopulations driving behavior.

4.3.7 Descending neurons exhibit raw odor encoding

Although walk- and head groom-encoding DN populations are recruited irrespective of odor context, it has been shown that a fictive odor (i.e., optogenetic activation of *Orco>CsChrimson*) can activate DNa02 neurons in immobile animals [89]. This implies that DNs may encode the presence and identity of real odors. To examine the extent of this raw sensory encoding, we trained and cross-validated linear discriminant odor classifiers using neural residuals (i.e., the neural activity remaining after subtracting activity that could be predicted using a model based on crf convolved behavior regressors). These residuals allowed us to control for the fact that odor presentation also modulates behavioral statistics **(Figure S4.8a-c)**.

Classification using neural residuals performed significantly better than classification using behavioral information alone ($p < 0.0001$ for a two-sided Mann-Whitney U test)**(Figure 4.4a)**. This reveals raw odor encoding within DN populations. However, this might result from many neurons with weak, complementary odor encoding or a few neurons with strong odor encoding. To distinguish between these possibilities, we next identified which DNs encode olfactory signals. We predicted each neuron's activity using regressors for behavior and the presence of each odor. The more intuitive approach of modeling neural activity by just using odor regressors does not account for the confound that behaviors are also modulated by specific odors. Therefore, we computed the amount of neural variance that could uniquely be explained by the presence of an odor and none of the behavior variables. We found that the activity of a few DNs could be uniquely explained by each of the odors **(Figure 4.4b, asterisks)**. In these neurons, clear activity peaks coincided with the presence of MSC **(Figure 4.4c)** or ACV **(Figure 4.4d)**. Notably, there appears to be no overlap between the DNs encoding MSC or ACV within individual animals **(Figure S4.8e)**. MSC encoding neurons were found dorsally on the lateral sides of the giant fibers in the connective while ACV encoding neurons were more broadly dispersed **(Figure 4.4e,f)**.

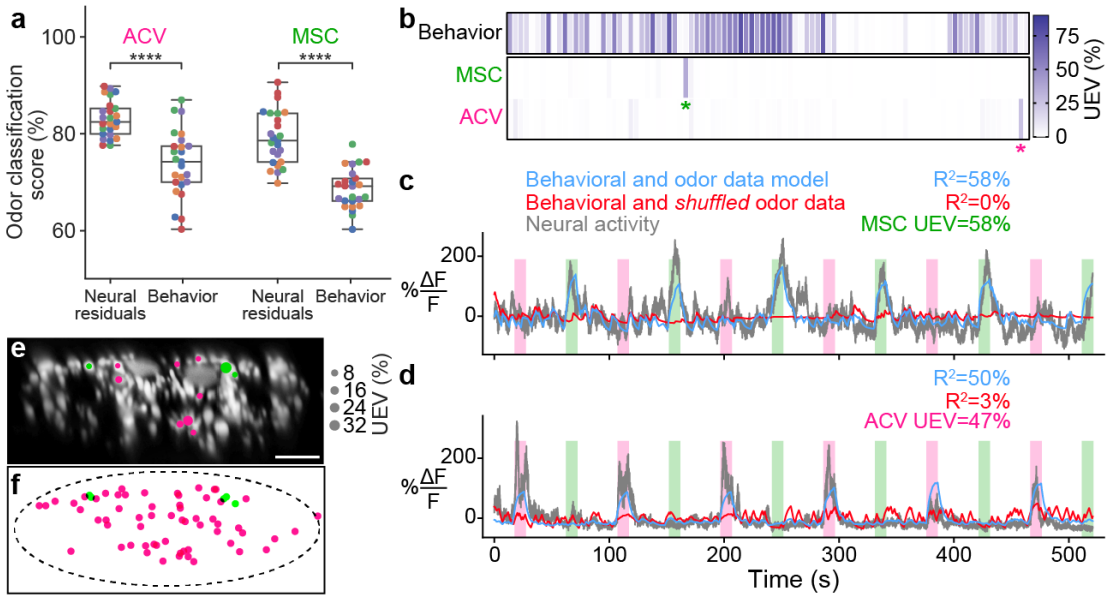

Figure 4.4: **Odor encoding in descending neuron populations. (a)** Neural residuals—obtained by subtracting convolved behavior regressors from raw neural data—can predict the presence of an odor significantly better than behavior regressors convolved with a calcium response function ('Behavior'). Two-sided Mann-Whitney U test (**** indicates $p < 10^{-4}$). The classification score was obtained using a linear discriminant classifier with cross-validation. Shown are five trials for five animals (color-coded). **(b)** Matrix showing the cross-validated ridge regression unique explained variance (UEV) of a model that contains behavior and odor regressors for one animal. The first row ('Behavior') shows the composite R^2 for all behavior regressors with odor regressors shuffled. The second and third rows show the UEVs for regressors of the odors methyl salicylate (MSC) or apple cider vinegar (ACV), respectively. Colored asterisks indicate neurons illustrated in panel **a**. **(c,d)** Example DNs best encoding **(c)** MSC or **(d)** ACV, respectively. Overlaid are traces of neural activity (gray), row one in the matrix (blue), and row one with odor data shuffled (red). **(e,f)** Locations of odor encoding neurons in **(e)** one individual, and **(f)** across all five animals. Scale bar is 10 µm.

4.3.8 Identifying individual descending neurons from population recordings

Until now, we have demonstrated that DN populations exhibit heterogeneous encoding: large, distributed groups encode walking and, by contrast, a few prominent pairs encode head grooming. Determining how these subpopulations control adaptive behavior is an important future challenge that will require a comprehensive approach examining phenomena ranging from global DN population dynamics down to the synaptic connectivity of individual DNs. The recent production of hundreds of sparse transgenic driver lines [59, 128] and connectomics datasets [85] suggest that this bridging of mechanistic scales may soon be within reach in *Drosophila*.

To illustrate how this might be accomplished, we aimed to identify specific DNs within our

population imaging dataset. Specifically, while analyzing head grooming DNs, we noticed a large pair of ventral neurons **(Figure 4.5a)** that sometimes exhibited asymmetric activity **(Figure 4.5b, gray arrowheads)** when flies appeared to touch one rather than both antennae **(Video 5)**. To quantify this observation, we replayed limb 3D kinematics in NeuroMechFly, a biomechanical simulation of *Drosophila* [32] **(Figure 4.5c)**. By detecting leg-antennal collisions as a proxy for antenna deflection, we found that occasional asymmetries **(Figure 4.5d)** did coincide with asymmetric activity in corresponding neural data **(Figure 4.5b, purple traces)**. These results suggested that this pair of DNs encodes mechanosensory signals associated with antennal deflections.

To further reveal the identity of these DNs, we examined data from our functional screen of sparse Gal4 and split Gal4 driver lines [228]. In this dataset, we observed similar asymmetric activity during antennal grooming in R65D11-GAL4. Coronal (x-z) two-photon imaging in R65D11 animals expressing OpGCaMP6f and tdTomato, shows axons that are similarly large and ventromedially located within the cervical connective **(Figure 4.5e)**. These also produce asymmetric activity during antennal grooming **(Figure 4.5f)**. This suggests that these neurons may report something unique to head grooming (e.g., coincident front limb movements) or simply antennal deflection. To distinguish between these possibilities, we analyzed neural responses to CO_2 puff stimulation of the antennae, while discarding data with resulting head grooming or front leg rubbing to ensure that the antennae were not touched by the legs. We measured an increase in the activity of both DNs upon puff stimulation **(Figure 4.5g,h)** suggesting that, like the neurons recorded in DN populations, R65D11 neurons also encode sensory signals—antennal deflection—rather than behavior.

To confirm that these sparse neurons are DNs, we next performed MultiColor FlpOut (MCFO) [170] and confocal imaging of their morphologies. R65D11 drives expression in several neurons. However, we found similarly large axonal projections from only one set of neurons that descend from the brain to the VNC **(Figure 4.5i, cyan)**. Close examination of these neurites in the brain **(Figure 4.5j)** and VNC **(Figure 4.5k)** revealed a striking resemblance to the reported structure of DNx01 neurons [59] with cell bodies outside of the brain—putatively in the antennae and enabling antennal mechanosensing.

These results enable the analysis of synaptic connectivity in identified DNs. To illustrate this, based on their unique location and size, we identified DNx01s in a VNC electron microscopy dataset [85] **(Figure 4.5l)** via manual reconstruction and striking morphological similarity to R65D11 DNs **(Figure 4.5m)**. From this reconstruction, once the full VNC connectome becomes available, one may identify synaptic partners of DNx01 to further understand how they contribute to controlling antennal grooming and other behaviors. Taken together, these data suggest a possible road map for using functional, topological, and morphological data to decipher the cellular identity of individual DNs from population recordings.

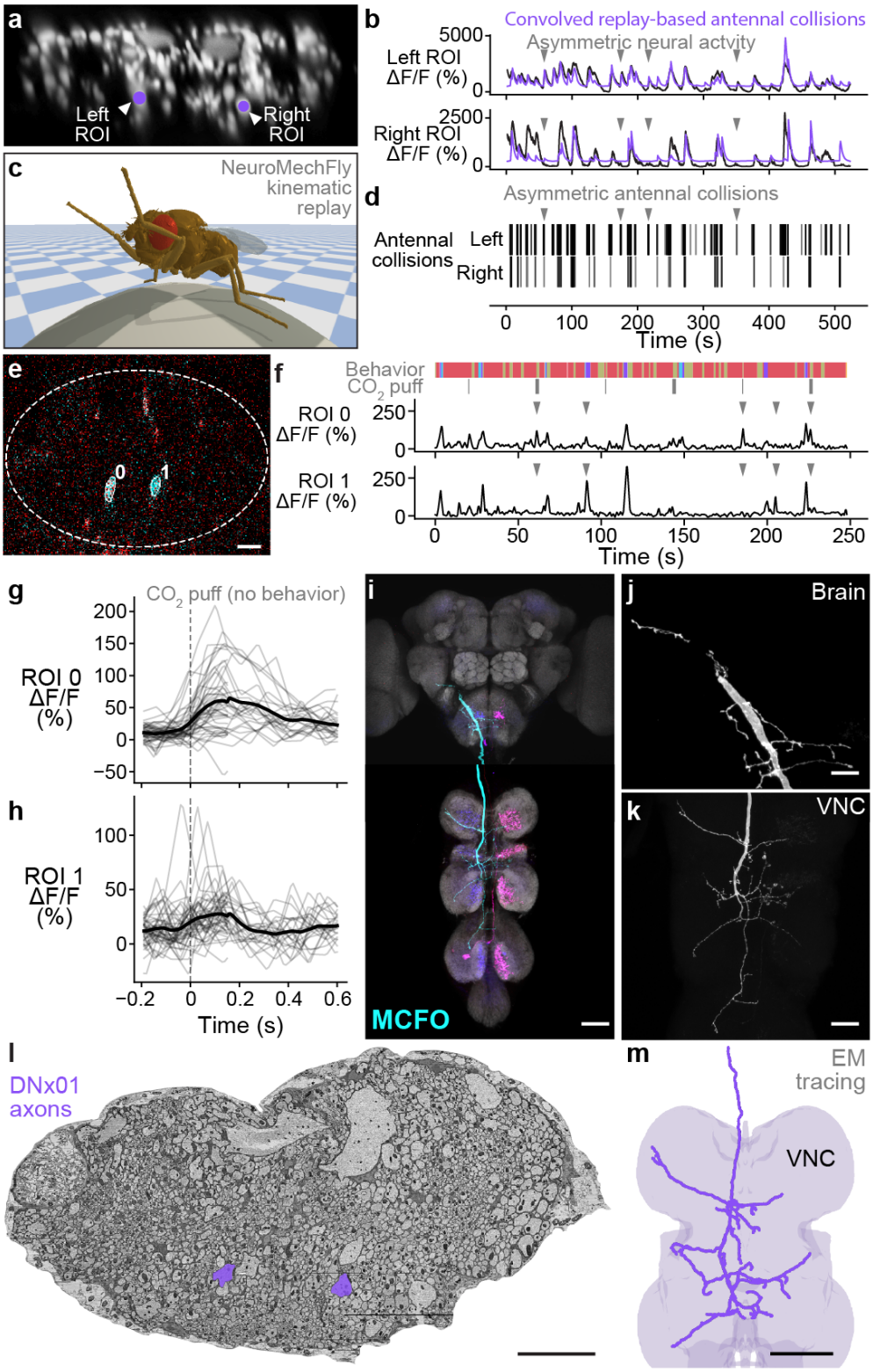

Figure 4.5: See Figure Legend on next page.

Identifying a pair of antennal deflection-encoding DNs from population recordings. **(a)** A pair of head groom-encoding DNs (purple circles and white arrowheads) can be identified from DN population recordings based on their shapes, locations, and activity patterns. **(b)** Example $\Delta F/F$ traces (black) of DNs highlighted in panel **a**. Sample time points with bilaterally asymmetric neural activity are indicated (gray arrowheads). Overlaid is a prediction of neural activity derived by convolving left and right antennal collisions measured through kinematic replay in the NeuroMechFly physics simulation (purple). **(c)** Kinematic replay of recorded joint angles in NeuroMechFly allow one to infer antennal collisions from real, recorded head grooming. **(d)** Left and right antennal collisions during simulated replay of head grooming shown in panel **b**. Sample time points with bilaterally asymmetric collisions are indicated (gray arrowheads). **(e)** Two-photon image of the cervical connective in a R65D11>OpGCaMP6f,tdTomato animal. Overlaid are ROIs identified using AxoID. The pair of axonal ROIs are in a similar ventral location and have a similarly large relative size like those seen in DN population recordings. Scale bar is 5 µm. **(f)** Sample neural activity traces from ROIs 0 and 1. Bilaterally asymmetric neural activity events (gray arrowheads), behaviors (color bar), and CO_2-puffs directed at the antennae (gray bars) are indicated. **(g,h)** CO_2 puff-triggered average of neural activity for ROIs **(g)** 0 and **(h)** 1. Only events in which animals did not respond with head grooming or front leg rubbing were used. Stimuli were presented at $t = 0$. Shown are individual responses (gray lines) and their means (black lines). **(i)** Confocal volume z-projection of MCFO-expression in an R65D11-GAL4 animal. Cyan neuron morphology closely resembles DNx01 [59]. Scale bar is 50 µm. **(j,k)** Higher magnification MCFO image, isolating the putative DNx01 from panel **i**, of the **(j)** brain and **(k)** VNC. Scale bars are 20 µm. **(l)** The locations of axons in the cervical connective (purple) from neurons identified as DNx01. Scale bar is 10 µm. **(m)** Manual reconstruction of a DNx01 from panel **l**. Scale bar is 50 µm.

4.4 Discussion

Here, by combining genetic and optical imaging approaches, we recorded the behavioral and sensory encoding of DN populations in behaving *Drosophila*. Across individual animals, we found that most recorded DNs encode walking. A smaller number are active during head grooming and resting. We did not find DNs encoding posterior movements, possibly due to the infrequency of this behavior. We also did not identify neurons that are active during multiple behaviors. This suggests that each ROI consists of individual neurons or that, if an ROI contains many neurons, they all show similar encoding. Subsets of walk-encoding DNs were also strongly active during turning: they were at higher density on the ipsilateral half of the cervical connective with respect to turn direction. By contrast, DNs distributed throughout the connective very weakly encoded walking speed. However, we caution that the small range of walking speeds in our data—flies accelerate rapidly from resting to walking and vice-versa—may mask stronger speed encoding. The partial overlap between turn- and speed-encoding DNs leaves open the possibility that neurons simultaneously encode these two properties in a differential steering fashion. Notably, we did not observe any DNs that are only active during transitions between multiple behaviors. However, the fly makes fast transitions. Thus, the

signal-to-noise and temporal resolution of our approach may not be sufficient to identify such neurons—higher temporal resolution electrophysiological recordings would be required to confirm the absence of DN encoding for behavioral transitions or for precise limb kinematics.

The encoding—and presumptive control—of walking by large numbers of DNs support the notion that the brain tightly regulates locomotion. In contrast to walking, head grooming is encoded by fewer neurons in our data. This may be because grooming limb movements are more stereotyped and thus may only rely on controllers within the VNC (a notion that is supported by the ability of headless flies to perform spontaneous grooming [45]). Other studies have also shown brain-wide activity during walking but not during other behaviors [159, 229, 230]. This difference may arise because adaptive locomotion—to avoid obstacles [231], cross gaps [232], and court potential mates [233]—depends heavily on the brain's descending signals. Thus, we hypothesize that, although a core set of command neurons can drive both walking and grooming, more DNs are additionally engaged during walking to allow for more flexible navigation in continuously changing and complex environments.

Because of the large number of DNs involved in walking, we hypothesized that subsets might represent parallel channels which are recruited depending on sensory context. For example, there may be DN subpopulations which drive walking in the presence of attractive odors and others engaged in the presence of aversive odors. This notion is supported by studies showing that optogenetic activation of a large variety of DNs elicits only a small set of stereotyped behaviors [74]. However, our population imaging data do not support this notion: largely the same DNs encode walking and head grooming irrespective of olfactory context.

A non-behavioral role for DNs has also been suggested by previous studies showing that the perception of a fictive odor modulates the activity of specific DNs involved in steering in immobile flies [89]. In line with this, we also identified DNs encoding odors and not behavior. Notably, the two odors we presented—apple cider vinegar and methyl salicylate—were encoded by distinct DNs. Extrapolating beyond ACV and MSC, it seems unlikely that DNs encode many individual odors with high specificity: the number of DNs is far smaller than what would be required to cover an enormous olfactory space. Instead, we speculate that DN odor-encoding may represent classes like attractive versus aversive odors or the valence of sensory inputs in general. Where odor information is conveyed within downstream motor circuits and for what purpose is a fascinating subject for future study.

Many fewer neurons encode head grooming as opposed to walking. Because our transgenic strain does not drive expression in SEZ neurons, we could not record several DNs whose activation has been shown to be sufficient to drive antennal grooming (aDN), front leg rubbing (DNg11), or both head grooming and front leg rubbing (DNg12) [78]. Thus, we expect that with the addition of these SEZ DNs the apparent dichotomy that many neurons encode walking and fewer encode grooming may become less pronounced. Nevertheless, groom-encoding DNs were notable in that they could be more reliably identified across individual animals. Among our head groom-encoding DNs, a pair passing through the ventral cervical connective

appear to encode limb contact during antennal grooming as well as puff-dependent antennal deflections. We speculate that mechanosensory signals from the antennae may be used for feedback control in the VNC: being aware of whether the antennae are touched while grooming may allow for a continuous modulation of grooming kinematics and force application by the front legs. This may be a conserved control mechanism as similar DNs have been identified in the blow fly [234].

Our morphological and physiological evidence suggest that these antennal mechanosensory signals are provided by DNx01s, a subset of bilateral campaniform sensillum (bCS) neurons that are also found on the legs and form major presynaptic connections to fast motor neurons [85]. Thus, DNx01s may have a role beyond feedback control during grooming. This is also implied by the large size of DNx01 axons and their high sensitivity to puff-mediated antennal deflection. Because other DNs with large axons (e.g., giant fiber neurons) are often implicated in fast reflexive movements [235, 236], DNx01s may be used to respond to, for example, strong gusts of wind that initiate a stance stabilization reflex.

Our work sets the stage for a more comprehensive, multi-scale investigation of how the brain regulates complex limb-dependent motor behaviors. Nevertheless, overcoming several technical limitations in our study should also be a focus of future work. First, although we could achieve precise limb kinematic measurements at 100 Hz, it will be critical to record neural data at equally high temporal resolution. The fly can walk with stride frequencies of up to 20 Hz [116] and leg movements during grooming occur at up to 7 Hz [225]. This currently makes it difficult to relate neural activity—read out by the relatively slow fluorescence fluctuations of a genetically-encoded calcium indicator—to individual joint angles and limb positions. To address this challenge, one might use faster indicators of neural activity (e.g., newer variants of GCaMP [20] or voltage indicators [237]). Additionally, coronal section two-photon imaging in the thoracic cervical connective with a piezo-driven objective lens limited our neural data acquisition to ~16 Hz. One might perform data acquisition at a higher rate using more advanced imaging methods including single-objective light-sheet microscopy [238]. Alternatively, the fly could be forced to generate slower limb movements using a motorized treadmill [239]. Another challenge is that DNs from the subesophageal zone (SEZ) are absent in our driver line. The SEZ is considered a center for action section [54, 82] and is known to house numerous DNs [59, 79]. Thus, complementing our driver line with an SEZ-expressing transgene [240] would enable a fully comprehensive recording of DN population dynamics. Nevertheless, we expect our observation to remain intact: locomotion is regulated by large DN populations in a distributed manner, while more stereotyped grooming behaviors engage fewer DNs. This would suggest a dichotomy in DN population control for flexible versus stereotyped motor behaviors. Future studies may test if this holds true as well for wing-dependent behaviors like continuous steering during flight [87, 95] versus stereotyped wing displays during courtship [119].

4.5 Materials and Methods

4.5.1 Fly husbandry and stocks

All experiments were performed on female *Drosophila melanogaster* raised at 25 °C and 50 % humidity on a 12 h light-dark-cycle. Flies were 10 days post-eclosion (dpe) for experiments, and had been starved overnight on a wet precision wipe (Kimtech Science, 05511, USA). Sources of each genotype used are indicated in **Table S4.1**.

4.5.2 Olfactometer

Mass flow controllers (MFC) were used to regulate air flow (Vögtlin, GSC-B4SA-BB23 (2 Lmin^{-1}), GSC-A3KA-BB22 (100 mLmin^{-1})). The larger MFC, set to 42 mLmin^{-1}, was used to continuously bubble odor vials, maintaining a constant head space odorant concentration. The smaller MFC was used to stimulate the fly at 41 mLmin^{-1}. We directed air flow using six solenoid valves (SMC, S070C-6AG-32, Japan) controlled by an Arduino Mega (Arduino, A000067, Italy). One valve in front of each of the three odor vials was used to switch between inputs from each MFC. A second valve after each odor vial was used to direct air flow either towards the fly or into an exhaust. The second valve was placed as close to the fly as possible (~10 cm) to minimize the delay between solenoid switching and odor stimulation. To direct air flow to the fly's antennae, we used a glass capillary held in place by a Sensapex zero-drift micro-manipulator (Sensapex, uMp-3, Finland). We minimized mechanical perturbations by blowing humidified (non-odorized) air onto the fly between odor trials. We used a PID (Aurora Scientific, miniPID, Canada) and visual assessment of animal behavior to confirm that switching generated minimal mechanical perturbations.

4.5.3 Two-photon microscopy

We performed neural recordings using a ThorLabs Bergamo two-photon microscope (Thorlabs, Bergamo II, USA) connected to a Mai Tai DeepSee laser (Spectra Physics, Mai Tai DeepSee, USA). To perform coronal section imaging, we operated the microscope in kymograph mode using the galvo-resonant light path. Images were acquired at a magnification of 7.4 X, resulting in a 82.3 µm wide field-of-view (FOV). To prevent the cervical connective from leaving the FOV, the full range of a 100 µm piezo collar was used to scan the objective lens (Olympus XLUMPlanFLN 20 X, 1.0 NA with 2 mm working distance) in the z-axis. We could achieve a frame rate of approximately 16 Hz by sampling 736 x 480 pixel (x- and z-axes, respectively) images and by enabling bidirectional scanning, with only one fly-back frame.

4.5.4 Neural recordings

Descending neuron population recordings

Flies were dissected to obtain optical access to the thoracic cervical connective, as described in [171]. Briefly, we opened the cuticle using a syringe and waited for the flight muscles to degrade before resecting trachea, the proventriculus, and the salivary glands. After removing these tissues covering the VNC, an implant [187] was inserted into the thoracic cavity to prevent inflation of the trachea and to ensure a clear view of the cervical connective for extended periods of time. Flies were then given several minutes to adapt to positioning over a spherical treadmill in the two-photon microscope system. During this adaptation period, the nozzle of the olfactometer was positioned approximately 2 mm in front of the the fly's head. As well, the thoracic cervical connective was brought into the imaging FOV. Following alignment, their position was further adjusted to maximize fluorescence signal and to minimize the possibility of neurites leaving the imaging FOV. For each fly, a minimum of five 9 min trials were recorded using a laser power of 9.25 mW at 930 nm.

Sparse DNx01 recordings

Recordings of DNx01s in the R65D11-GAL4 driver line were performed as described in [228]. This differed only slightly from DN population recordings in the following ways. First, flies were not starved. Second, the Gal4 drove expression of OpGCaMP6f rather than GCaMP6s. Third, a thoracic implant was not used. Fourth, instead of being presented with a constant flow of air and odors, animals were stimulated with CO_2 puffs of alternating length (0.5 s, 2 s) spaced 40 s apart. A higher pixel dwell time was used to achieve acceptably high imaging signal-to-noise. This resulted in a slower two-photon image acquisition rate (4.3 fps) which was matched by a slower behavior recording frequency (30 Hz). For additional details and a description of the stimulation system see [228]. Note that images of the connective in **Figure 4.5a** and **Figure 4.5e** appear to have different heights due to a difference in z-step size during image acquisition.

4.5.5 Post-processing of two-photon imaging data

Descending neuron population recordings

Binary output files from ThorImage (Thorlabs, ThorImage 3.2, USA) were converted into separate tiff files for each channel using custom Python code (https://doi.org/10.5281/zenodo.5501119). Images acquired from the red channel were then denoised [241] and two-way alignment offset was corrected (https://doi.org/10.5281/zenodo.6475468) based on denoised images. We then used optic flow estimation to correct image translations and deformations based on the denoised red channel images (https://doi.org/10.5281/zenodo.6475525). The green channel was then corrected based on the motion field estimated from the red channel. Finally, we trained a DeepInterpolation network [223] for each fly using the first 500 motion

corrected green channel images from each experimental trial (batch size=20; epochs=20; pre-post frames=30). The first and the last trials were used as validation datasets. The trained network was then used to denoise green channel images.

Baseline fluorescence was then determined on a fly-wise and pixel-wise basis across all trials. The baseline of a pixel was defined as the minimum 'mean of 15 consecutive values' across all experimental trials. Motion correction introduces zeroes to two-photon images in background regions that were out of the FOV prior to warping. Therefore, values close to zero (floating point inaccuracy) were set to the maximum of the datatype of the array. This means essentially ignoring these pixels and their immediate surroundings for baseline computations. We used the baseline image F_0 to calculate $\%\frac{F-F_0}{F_0}$ images. These images were only used for visualization. To identify ROIs/neurons, we created a maximum intensity projection of $\%\frac{F-F_0}{F_0}$ images and manually annotated ROIs. The $\%\frac{F-F_0}{F_0}$ of each ROI was computed by first spatially averaging it's raw pixel values. We then calculated the baseline of this average as described for a single pixel above. For brevity and readability, we refer to $\%\frac{F-F_0}{F_0}$ as $\%\Delta F/F$ throughout the manuscript.

Sparse DNx01 descending neuron recordings

ROIs were detected using AxoID. Raw, non-denoised traces were used for analysis. For more details concerning data post processing see [228].

4.5.6 Behavior classification and quantification

Behavior measurement system

The behavior of tethered animals was recorded using a previously described [37] 7-camera (Basler, acA1920-150um, Germany) system. Animals were illuminated using an infrared (850 nm) ring light (CSS, LDR2-74IR2-850-LA, Japan). To track the joint positions of each leg, six cameras were equipped with 94 mm focal length 1.00 X InfiniStix lenses (Infinity, 94mm/1.00x, USA). All cameras recorded data at 100 fps and were synchronized using a hardware trigger. For more details see [37].

Inferring fictive locomotor trajectories

Video data from the front camera were processed using FicTrac [36] to track ball movements. This camera was outfitted with a lens allowing adjustable focus and zoom (Computar, MLM3X-MP, 0.3X-1X, 1:4.5, Japan). To improve tracking accuracy, the quality factor of FicTrac was set to 40. The circumference of the ball was detected automatically using a Hough circle transform on the mean projection of all images for a given experimental trial. To determine the vertical angular FOV, α, the value given in the specifications of the Computar lens (8.74°) had to be adjusted, accommodating a smaller sensor size. We first determined the focal length to be

43.18 mm using Equation 4.1, where H is the height of the sensor. This was set to 6.6 mm for a 2/3" sensor.

$$f = \frac{H}{2tan(\frac{\alpha}{2})} \tag{4.1}$$

The ROI of the Basler camera was set to 960x480 pixels, reducing the effective sensor height from 5.8 mm to 2.32 mm. Rearranging Equation 4.1 and plugging in f and H yields a vertical angular FOV of 3.05°. Since the camera was already aligned with the animal, the camera-to-animal transform was set to zero. To obtain the fly's trajectory, we developed custom code that integrates rotational velocities (https://github.com/NeLy-EPFL/utils_ballrot).

Postprocessing of 3D pose estimates

Outliers in 3D pose data were corrected as described in [228]. Briefly, we detected outliers based on changes in leg segment length and chose the pair of cameras with minimal reprojection error for triangulation. After outlier correction, the data were aligned and joint angles were computed using published code [32]: https://github.com/NeLy-EPFL/df3dPostProcessing/tree/outlier_correction.

Classification of behaviors

Behaviors were classified based on limb joint angles using the approach described in [42]. Briefly, a network was trained using 1 min of annotations for each fly and heuristic labels. Motion energy, ball rotations, and joint positions where used to generate the heuristic labels. To compute the motion energy, each joint position was convolved with the finite difference coefficients of length nine, estimating the first derivative. After computing the L^1-norm, the signal was filtered with a 10th order low-pass Butterworth filter of critical frequency 4 Hz. The total, front, and hind motion energy were computed by summing over all joints, all front leg joints, and all hindleg joints, respectively. Forward ball rotation speeds were processed using the same Butterworth filter described above. First, we assigned heuristic labels for walking by thresholding the filtered forward walking velocity at $0.5\,\mathrm{mm\,s^{-1}}$. The remaining frames with a motion energy smaller than 0.3 were then classified as resting. Next, heuristic labels for front movements (front motion energy > 0.2 and hind motion energy < 0.2) and posterior movements (front motion energy < 0.2 and hind motion energy > 0.2) were assigned to all remaining frames. The front movement labels were further split into head grooming and front leg rubbing by thresholding the height of the front leg tarsi (the average between left and the right tarsi) at 0.05. After each step, a hysteresis filter was applied. This filter only changes state when at least 50 consecutive frames are in a new state. Based on a hyperparameter search using 'leave one fly out' cross validation on the hand labels **(Figure S4.1g)**, we selected the

weights of $\lambda_{pred} = 0$ and $\lambda_{weak} = 1$ for the loss.

Biomechanical simulation and antennal collision detection

To infer limb-antennal collisions, we performed kinematic replay using the NeuroMechFly physics simulation framework as described in [32]. We used joint angles to replay real limb movements in the simulation. To avoid model explosion and accumulating errors over long simulation times, we ran kinematic replay on time segments shorter than the full experimental trials. These segments consisted of individual head grooming and front leg rubbing events. The default head angle was fixed to 20°. The aristae yaw values were set to −32° and 32° and pedicel yaw values were set to −33° and 33° for the left and right sides, respectively. To speed up the simulation, we only detected collisions between the front legs and head segments.

4.5.7 Confocal imaging of the brain and ventral nerve cord

Confocal images were acquired using a Zeiss LSM700 microscope. These images were then registered to a brain and VNC template described in [228] using the method from [191]. Brain and VNC sample preparation was performed as described in [228]. Both primary and secondary antibodies were applied for 24 h and the sample was rinsed 2-3 times after each step. Antibodies and concentrations used for staining are indicated in **Table S4.2**.

4.5.8 Electron microscopy identification and tracing

Within an electron microscopy dataset of the ventral nerve cord and neck connective [85], we identified the pair of DNx01s based on their large-caliber axons in the cervical connective positioned ventral to the giant fiber neurons axons (**Figure 4.5l**). We then manually reconstructed all branches of one DNx01 neuron using CATMAID [242, 243] as described in [85]. The reconstructed neuron was then registered to the female ventral nerve cord standard template [192] using an elastix-based atlas registration as described in [85].

4.5.9 Data analysis

Animals were excluded from analysis if they produced irregular behavior, or fewer than 75 neuronal ROIs could be manually identified in two photon imaging data.

Linear regression modeling

We relied primarily on regression techniques to link behavioral and neural data. Here we first describe the general approaches used and then discuss details and modifications for individual figure panels. To evaluate the success of regression models, we calculated explained variance in the form of the coefficient of determination (R^2) and unique explained variance

(UEV) [155]. The explained variance is a measure of how much additional variance is explained by the model compared to an intercept only model (i.e., approximating the data by taking the mean). A definition of the coefficient of determination can be found in Equation 4.2, where SSE is the sum of squares of the error, SST is the total sum of squares, y_i is a data point, $\hat{y}_i$ is the prediction of y_i, and $\bar{y}$ is the mean of all data points.

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{4.2}$$

Note that, R^2 becomes negative when SSE is larger than SST. This is the case when the model prediction introduces additional variance due to overfitting. UEV is a measure for the importance of individual or a subset of regressors. It is computed as the reduction in R^2 when a particular subset of regressors is randomly shuffled (Equation 4.3).

$$\text{UEV} = R^2_{\text{intact}} - R^2_{\text{shuffled}} \tag{4.3}$$

We performed 5-fold cross-validation for all of our regression results to ensure good generalization. Non-regularized linear regression sometimes led to overfitting and negative R^2 values. Therefore, we used ridge regression. The ridge coefficient was determined using 5-fold nested cross-validation on the training data set. In some cases we still observed small negative R^2 values after applying ridge regularization. These were set to zero, in particular to avoid problems when computing UEVs. To account for the long decay dynamics of our calcium indicator we convolved behavior variables with an approximation of the calcium response function (crf) (Equation 4.4).

$$\text{crf}(t) = -e^{-at} + e^{-bt} \tag{4.4}$$

We used $a = 7.4$ and $b = 0.3$ to approximate the rise and decay times, respectively, as reported in [19]. We also normalized the function to integrate to one on the interval 0 s to 30 s.

Figure 4.2 & **Figure S4.3**:
In these figures we predicted behavior from neural activity. To accomplish this, we trained models for all pairs of behaviors and neurons (e.g., walking and ROI 41 for the upper plot of **Figure 4.2a**). The target variable is a binary variable indicating whether the fly is walking or not. This was convolved with the crf (black line in panel **a**). The single regressor besides the intercept in the model is the $\Delta F/F$ of a single neuron. **Figure 4.2b** and **Figure S4.3c** show the R^2 values for all models. Each neuron was then assigned to be encoding the behavior with the maximum R^2. For neurons with maxima smaller than 5%, no behavior was assigned.

Figure S4.2:
Using the approach described in the previous section, we observed that some neurons appear to predict both walking and posterior movements, while others predict both head grooming and front leg rubbing. Because fluorescence decays slowly following the cessation of neural ac-

tivity, this may be caused by the frequent sequential occurrence of two behaviors. For instance, if front leg rubbing often occurs after head grooming, the $\Delta F/F$ of a head groom encoding neuron may still be elevated during front leg rubbing. This can lead to false positives in our analysis. To address this potential artifact, we predicted neural activity from multiple behavior regressors (i.e., binary behavior indicators convolved with the crf). We then calculated the UEV for each behavior regressor. For example, when the front leg rubbing regressor is shuffled the R^2 will not decrease by much because the head grooming regressor includes the expected decay through convolution with a crf. The model of **Figure S4.2a** includes a walking and a posterior movements regressor. For **Figure S4.2b** the model includes a head grooming and a front leg rubbing regressor. No other regressors were included in these models and for a given model, only data during one of these two behaviors were used.

Figure 4.3 & **Figure S4.7**:
For **Figure 4.3a** we used all behavior regressors and the $\Delta F/F$ of all neurons to predict ball rotation speeds convolved with a crf. We then temporally shuffled each of the neural regressors to calculate their UEVs. Since knowing whether the fly is walking or not provides some information about forward speed, the R^2_{intact} did not decrease to zero in the speed prediction (panel **a**, 2nd row). We address this issue in panel **b** by only including walking frames but otherwise using the same approach as in panel **a**. To pinpoint the encoding of ball rotations to individual neurons **(Figure 4.3c)**, we made two changes to our approach. First, instead of predicting turning in general, we split the turning velocity into right and left turning. Second, we only included the $\Delta F/F$ of a single neuron in the model rather than the data from all neurons. As before, the target variables were spherical treadmill rotation speeds—right turning, left turning, and forward walking—convolved with the crf.

Figure 4.4 & **Figure S4.8**:
To model odor encoding, we predicted the activity of a single neuron using behavior regressors, including crf-convolved spherical treadmill rotation speeds, and odor regressors constructed by convolving the crf with a binary variable indicating whether a given odor was present or not. To determine how much neural variance can be explained by the behavior regressors in our model, we shuffled both odor regressors **(Figure 4.4b, top row)**. We then calculated the UEV for each odor by first computing the R^2 of the model with all regressors intact, and then subtracting the R^2 of the model after shuffling the odor regressor in question **(Figure 4.4b bottom)**. In panels **c** and **d**, the intact model's prediction (blue) and the shuffled model's prediction (red) are shown.

Figure S4.9:
To see if a neuron's behavior encoding could change depending on the context of which odor is present, we only examined the most commonly encoded behaviors: walking and head grooming. Each row is equivalent to the the walking or head grooming rows in **Figure S4.3b**. However, here we only used subsets of the data to train and validate our models. Each row only uses data when one odor (or humidified air) is present. If there is no context dependence,

each of the three rows should look identical. However, we note that due to the significant reduction in the amount of data for each model, noise can introduce variation across conditions. Humidified air data was subsampled to match the amount of data available for the odors MSC and ACV. To test whether the encoding was significantly different across contexts, we used a two-sided Mann-Whitney U test. The data points are individual cross-validation folds from each trial. Elsewhere in this study we report cross-validation means.

Figure 4.5:
In panel **b** we perform a regression to predict one neuron's activity using the intercept as well as a crf-convolved antennal collision regressor derived from data in panel **d**.

Kernel density estimation

We performed 2D kernel density estimation **(Figure 4.2h,i)** using SciPy's gaussian_kde [244]. We performed 1D kernel density estimation **(Figure 4.3h,j; Figure S4.7d,f)** using sklearn [245]. We normalized kernel density estimates to correct for the variable density of neurons across the connective. The normalization factor was computed as a kernel density estimate of all annotated neuron locations. The kernel bandwidth was determined using leave-one-out cross-validation. This maximizes the log-likelihood of each sample under the model. Data from all flies were used to determine the bandwidth.

Principal component analysis

We performed principal component analysis (PCA) on DN population data. First, as in [226], we observed that PCs from time derivatives of fluorescence traces produce more organized state space trajectories. Therefore, we calculated the derivative of the $\Delta F/F$ traces for each neuron (extracted from non-denoised imaging data) using total variation regularization [246]. Empirically, we found that a regularization parameter of 5000 strikes a good balance between bias and variance in the derivatives. We then performed PCA on the derivatives of all neural data during walking only. This allowed us to specifically ask whether, during walking, neural activity largely remained constant or diverged as a function of specific locomotor features. We then embedded walking and resting neural data epochs into the same PC space (i.e., we did not refit the PCs for resting). We visualized the loadings of individual ROIs/neurons using vectors to illustrate how these neurons influence the position in PC space. For clarity we only show vectors for whom $w_{pc1}^2 + w_{pc2}^2 > 0.1$, where w_{pc1} and w_{pc2} are the loadings of the first and second principal components.

Linear discriminant analysis

We trained linear discriminant models to distinguish between ACV and humidified air, as well as MSC and humidified air **(Figure 4.4a)**. To evaluate classification accuracy, the classification

score was cross-validated. Input data to the model were either behavior regressors or neural residuals. The residuals were computed using ridge regression (as described above) with behavior regressors as input. In both cases the behavior regressors were convolved with the crf.

Event-triggered averaging

Here, 'events' describe individual epochs of a particular behavior. To perform event-triggered averaging of images/videos **(Video 2-4)**, we first identified all events that had no similar event in the previous 1 s. Raw microscope recordings were then chopped into blocks starting 1 s prior to event onset and lasting 4 s after event onset or until the end of the event, whichever was shorter. Each block was then converted into $\Delta F/F$ using the mean of the first five frames as a baseline. All blocks were then temporally aligned and averaged frame-by-frame. We discarded behavior videos with fewer than 50 events 1 s after event onset. Event-triggered averaging of neural traces **(Figure 4.5g,h)** was performed in a similar fashion. However, instead of using raw images, $\Delta F/F$ traces were used and no block-wise $\Delta F/F$ was computed. The values were then averaged one time-point at a time.

Correlation coefficients

We calculated the Pearson's correlation coefficient for each trial between the raw $\Delta F/F$ trace and a time shifted $\Delta F/F$ trace calculated using denoised images **(Figure S4.1b)**. We then either grouped the values by time lag and averaged each group (bottom), or we found the time shift with maximal cross-correlation. In **Figure 4.3d** and **Figure S4.7b** we perform several types of correlation analyses. First, the larger matrices show the correlation between neurons. This is computed as the Pearson's correlation coefficient between the $\Delta F/F$ values of a pair of neurons. Correlations between neural activity and turning or walking speed were calculated as the Pearson's correlation coefficient between the $\Delta F/F$ values of a neuron and corresponding spherical treadmill rotations, using only data when the fly was classified as walking. All of the above calculations were performed on individual trials and then averaged across all trials.

Hierarchical clustering

We used Ward's method [247] to hierarchicaly cluster and sort neurons based on their correlation **(Figure 4.3d)**. The distance between pairs of neurons was set to $1 - r$, where r is the Pearson's correlation coefficient for the two neurons.

4.6 Data and code availability

Data are available at:
https://dataverse.harvard.edu/dataverse/DNs

Analysis code is available at:
https://github.com/NeLy-EPFL/DN_population_analysis

4.7 Funding

PR acknowledges support from an SNSF Project Grant (175667) and an SNSF Eccellenza Grant (181239). FA acknowledges support from a Boehringer Ingelheim Fonds PhD stipend.

4.8 Acknowledgments

We thank K. Asahina (Salk Institute, San Diego, USA) and B. McCabe (EPFL, Lausanne, Switzerland) for transgenic *Drosophila* strains. We thank J. Phelps for discussions and assistance with manual annotation of the EM dataset.

4.9 Author Contributions

F.A. - Methodology, Software, Formal Analysis, Investigation, Data Acquisition, Data Curation, Writing – Original Draft Preparation, Writing - Review & Editing.
C.L.C. - Investigation, Data Acquisition, Data Curation, Writing - Review & Editing.
P.R. - Conceptualization, Methodology, Resources, Writing – Original Draft Preparation, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition.

4.10 Competing interests

The authors declare that no competing interests exist.

4.11 Supplementary Information

4.11.1 Supplementary Tables

Genotype	Source	Reference
w; $\frac{\text{tub>stop>GAL80}}{\text{(CyO)}}$; $\frac{\text{MKRS}}{\text{TM3, Sb (21-21)}}$	Asahina lab (Salk Institute, San Diego, CA USA)	[222]
w; $\frac{\text{Otd-nls:FLP (attP40)}}{\text{(CyO)}}$; $\frac{\text{TM2}}{\text{TM6B(26-27)}}$	Asahina lab (Salk Institute, San Diego, CA USA)	[222]
w; tubP-(FRT.GAL80);	Bloomington Drosophila Stock Center (#62103)	
w; + ; R57C10-GAL4	Bloomington Drosophila Stock Center (#39171)	[128]
w; + ; 10xUAS-IVS-myr::smGFP-FLAG (attP2)	Bloomington Drosophila Stock Center (#62147)	[170]
+[HCS]; P20XUAS-IVS-GCaMP6s attP40; Pw[+mC]=UAS-tdTom.S3	Dickinson lab (Caltech, Pasadena, CA USA)	
;P20XUAS-IVS-Syn21-OpGCaMP6f-464 p10 su(Hw)attp5; Pw[+mC]=UAS-tdTom.S3	Dickinson laboratory (Caltech, Pasadena, CA USA)	
R57C10-446 Flp2::PEST in su(Hw)attP8; ; HA-V5-FLAG	Bloomington Drosophila Stock Center (#64089)	[170]
w;;P{w[+mC]=20xUAS-DSCP> H2A::sfGFP-T2A-mKOk::Caax} JK66B	McCabe lab (EPFL, Lausanne, Switzerland)	[248]
;Otd-nls:FLPo (attP40)/(CyO); R57C10-GAL4, tub>GAL80>/(TM6B)	This paper (available upon request)	

Table S4.1: Sources of transgenic *Drosophila melanogaster* strains.

Produced in	Antibody	Dilution	Manufacturer	RRID
rabbit	anti-GFP	1:500	Thermofisher	AB 2536526
mouse	anti-Bruchpilot	1:20	Developmental Studies Hybridoma Bank	AB 2314866
goat	anti-rabbit conjugated with Alexa 488	1:500	Thermofisher	AB 143165
goat	anti-mouse conjugated with Alexa 633	1:500	Thermofisher	AB 2535719
rabbit	anti-HA-tag	1:300	Cell Signaling Technology	AB 1549585
rat	anti-FLAG-tag DYKDDDDK	1:150	Novus	AB 1625981

Table S4.2: Antibodies and concentrations used to stain *Drosophila melanogaster* nervous tissues.

4.11.2 Supplementary Figures

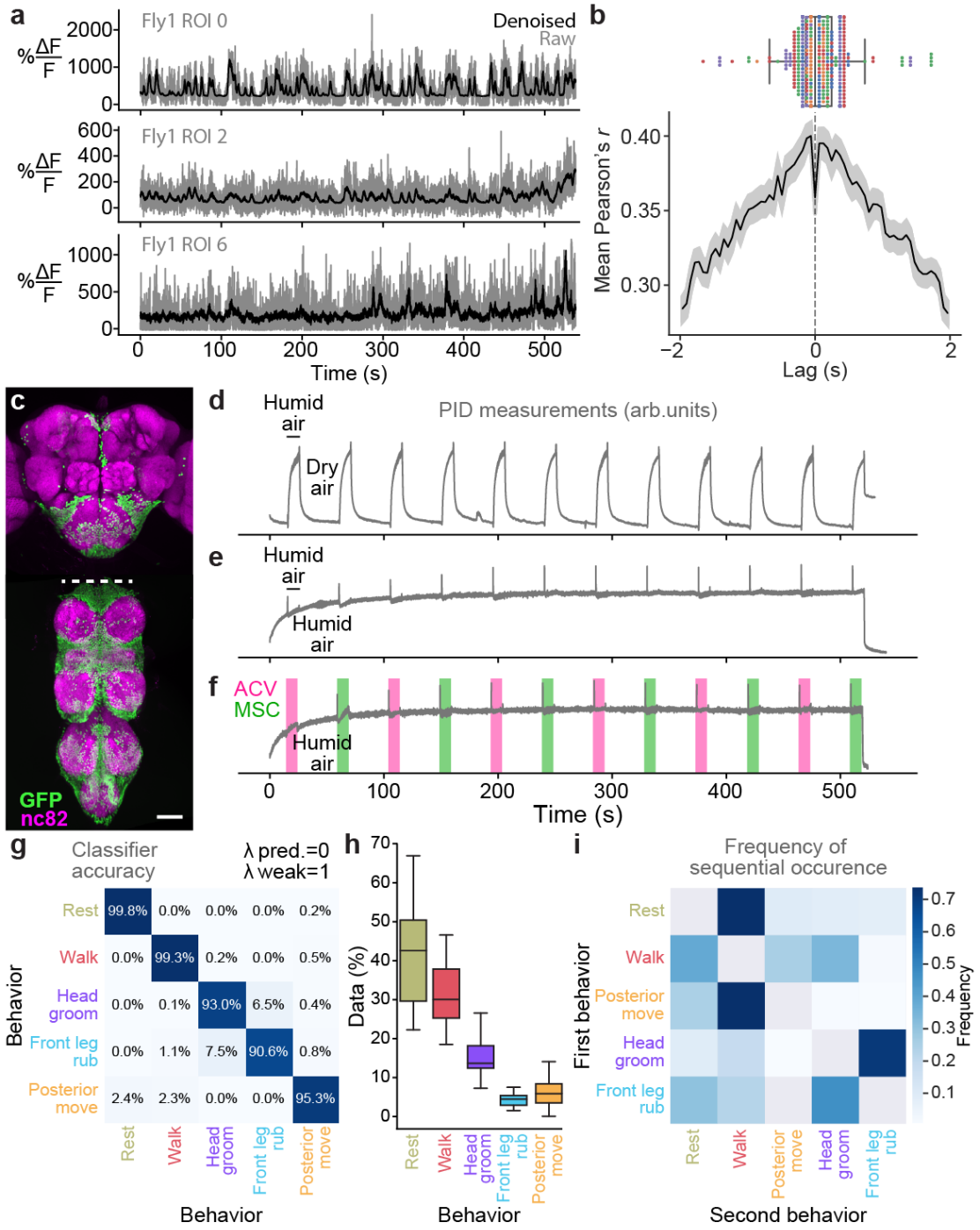

Figure S4.1: See Figure Legend on next page.

Figure S4.1: **Supporting details for neural denoising, driver line expression, odor stimulation, and behavior quantification.** **(a)** Example $\Delta F/F$ traces extracted from optic-flow registered 'Raw' (gray) and corresponding 'Denoised' (black) images. **(b, top)** Lag between denoised and raw $\Delta F/F$ traces with maximal cross-correlation. Overlaid are individual data points from each fly (color-coded). **(b, bottom)** Average cross-correlation for all ROIs and flies (solid line) and corresponding 95% confidence interval (shaded region). **(c)** Z-projected confocal image of the genetic complement of our 'brain only' driver line (otd-nls:FLPo,tub>stop>GAL80; R57C10-GAL4) expressing nuclear GFP. Shown are neuropil ('nc82', magenta) and GFP (green) staining. Scale bar is 50 µm. **(d-f)** Photoionization detector (PID) measurements during our odor delivery protocol switching between **(d)** humid and dry air, **(e)** humid and humid air (to measure valve-related transients) or **(f)** humid air, MSC, and ACV odors. **(g)** Results of leave-one-fly-out cross-validation hyperparameter search for our behavior classifier. The values $\lambda_{weak} = 1$ and $\lambda_{pred} = 0$ yield the highest classification accuracy. **(h)** Relative frequency of classified behaviors in our dataset (n=5 animals). **(i)** The frequency of transitions between sequential behaviors.

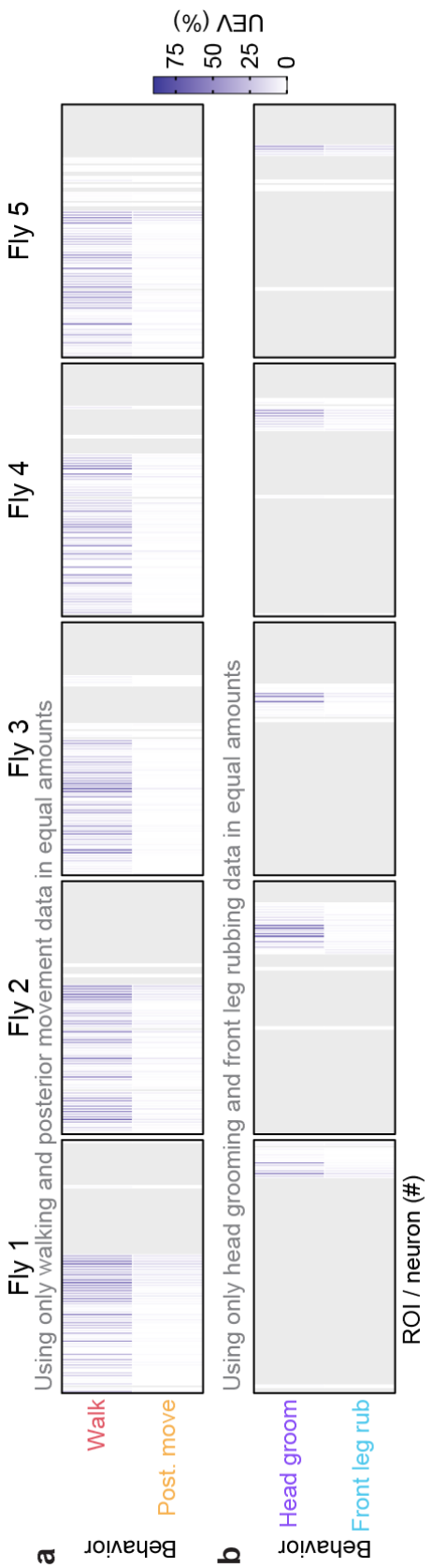

Figure S4.2: **Disentangling the relative encoding of frequently sequential behavior pairs. (a-b)** Cross-validation mean of neural variance uniquely explained by **(a)** walking versus posterior movements or **(b)** head grooming versus front leg rubbing. In both cases only data acquired during the two compared behaviors were analyzed. Additionally, data were balanced to have an equal amount across both behaviors.

Figure S4.3: **Encoding of behavior in descending neuron populations across individual animals. (a)** Mean time projections of GCaMP6s fluorescence over a nine minute recording for five animals. Images are inverted for clarity, illustrating high mean fluorescence (black). Manually identified DN regions-of-interest (ROIs) are shown (red rectangles). Scale bar is 10 µm. All subpanels and panels **c-d** share the same scale. **(b)** The cross-validation mean of behavioral variance explained by DNs for each animal (Fly 1: n=95 ROIs; Fly 2: n=86 ROIs; Fly 3: n=75 ROIs; Fly 4: n=81 ROIs; Fly 5: n=79 ROIs). **(c)** DNs color-coded by the behavior their activities best explain. Radius scales with the amount of variance explained. **(d)** Behavior-triggered average $\Delta F/F$ images for the most common behaviors—resting, walking, and head grooming. **(e)** Locations of DNs for the classified behavior they encode best. Data are pooled across five animals.

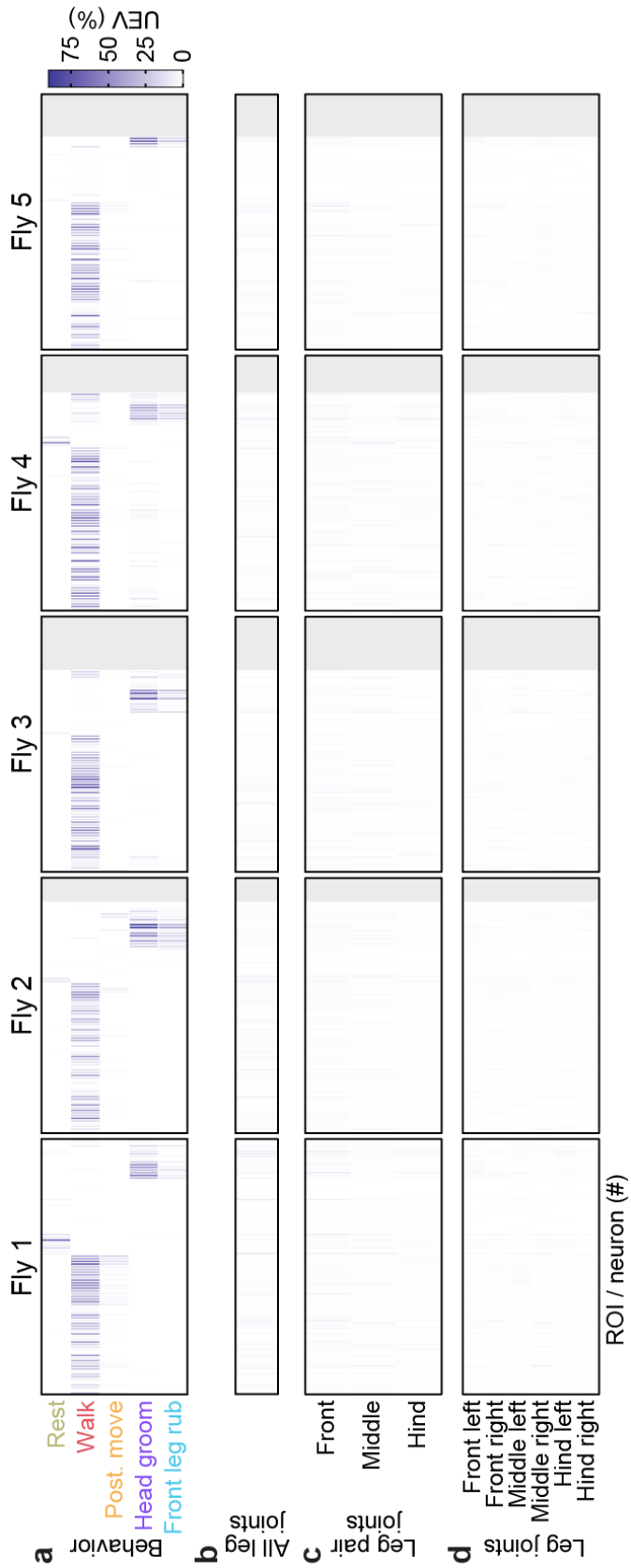

Figure S4.4: **Neural variance explained by distinct kinematic features. (a-d)** Amount of neural variance that can be uniquely explained by **(a)** classified behaviors (taken from the previous figure) **(b)** all joint movements **(c)**, leg pair movements, or **(d)** individual leg movements.

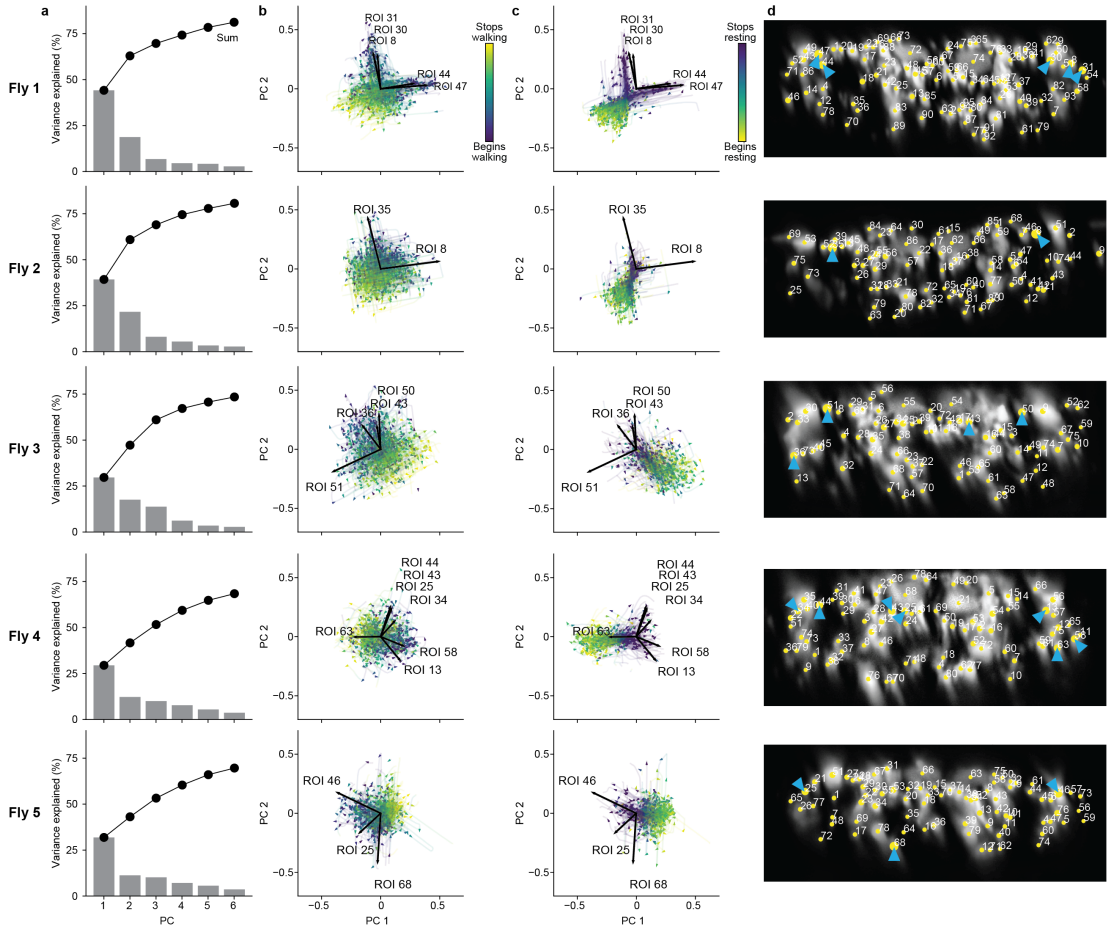

Figure S4.5: **Principal component analysis of neural activity during walking and resting across individual animals.** **(a)** Amount of variance explained by six principal components (PCs) of neural activity derivatives during walking for five individual flies. **(b,c)** The derivative of neural activity during **(b)** walking and **(c)** resting. PC embeddings were trained on data taken during walking only. Colored trajectories are individual epochs of **(b)** walking and **(c)** resting. Time is color coded and the temporal progression of each epoch is indicated (arrowheads). Note that color scales are inverted to match the color at transitions between walking and resting. Black arrows indicate PC loadings for DNs with vectors longer than $\sqrt{0.1}$. ROI numbers correspond to the connective image in panel **d**. **(d)** Locations of ROIs for each individual animal (yellow circles). Numbers are based on the order in **Figure S4.3b**. Circle radii indicate the norm of the loadings for PCs 1 and 2 (i.e., the lengths of the vectors in panels **b** and **c**). ROIs with the largest loadings are indicated (cyan arrowheads).

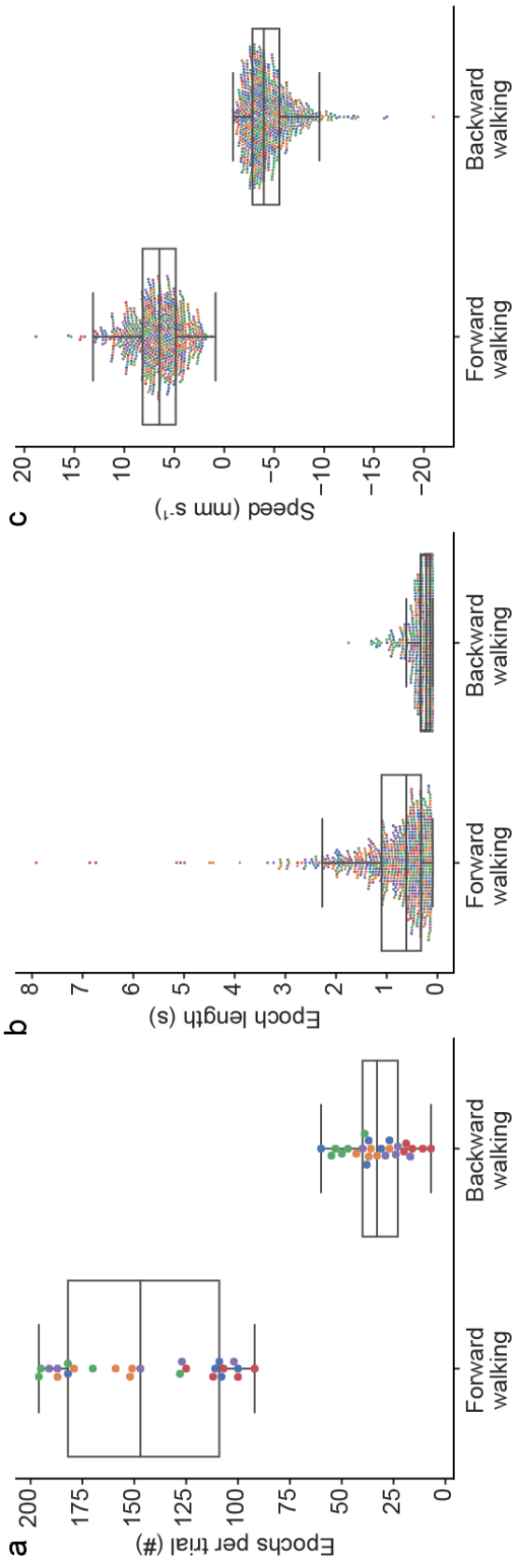

Figure S4.6: **Backward walking is infrequent and brief.** Compared to forward walking epochs, in our data we measured **(a)** fewer and **(b)** shorter backward walking epochs. **(c)** There was also a limited dynamic range in backward walking speed. In all panels, colors indicates fly identities. In panels **b** and **c** circles indicate the values of 700 randomly selected epochs (subsampled for clarity). Box plots indicate the median, lower, and upper quartiles. Whiskers signify 1.5 times the interquartile range.

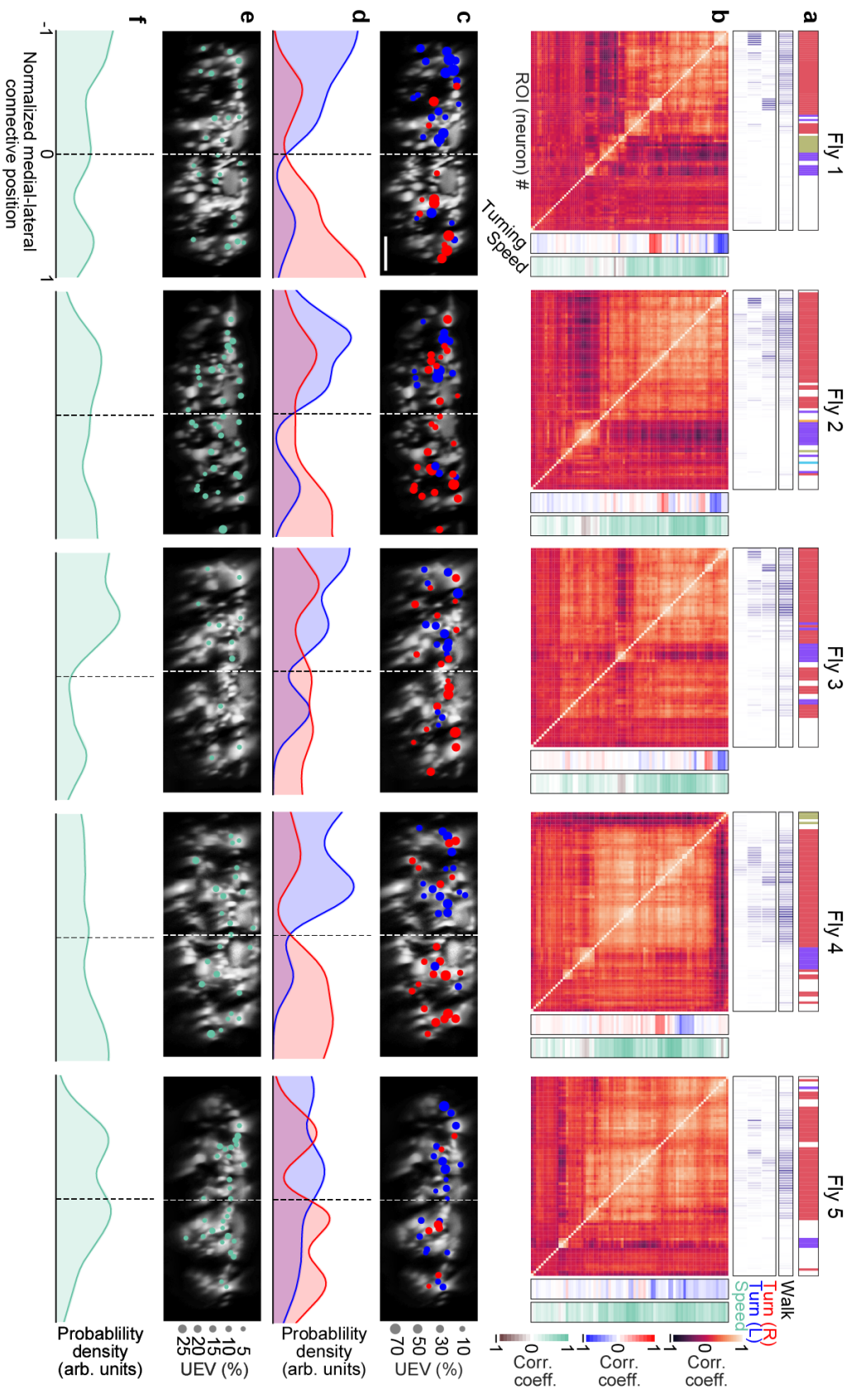

Figure S4.7: See Figure Legend on next page.

Figure S4.7: **Turning and speed encoding in descending neuron populations across individual animals. (a)** Unique explained variance of each DN from individual animals for left and right turning or walking speed. Walking R^2 values are taken from **Figure S4.3b**. The model for turning and speed was obtained using behavior regressors as well as neural activity. To compute the UEV, activity for a given neuron was temporally shuffled. The behavior whose variance is best explained by a given neuron is indicated (top). Neurons are ordered as in panel **b**. **(b)** Pearson's correlation coefficient matrix comparing neural activity across DNs is ordered by hierarchical clustering. Shown as well are the correlations of each DN's activity with right, left, and forward walking (right). Neurons are ordered as in panel **a**. **(c)** Locations of turn encoding DNs (UEV > 5%), color-coded by preferred direction (left, blue; right, red). Circle radii scale with UEV. Dashed white line indicates the approximate midline of the cervical connective. Scale bars are 10 µm for all subpanels and panel **e**. **(d)** Kernel density estimates of the distributions of turn encoding DNs across individuals (opaque lines). Probability densities are normalized by the number of DNs along the connective's medial-lateral axis. **(e)** Locations of speed encoding DNs (UEV > 2%). Circle radii scale with UEV. Dashed white line indicates the approximate midline of the cervical connective. **(f)** Kernel density estimates of the distributions of speed encoding DNs across individuals (opaque lines). Probability densities are normalized by the number of DNs along the connective medial-lateral axis.

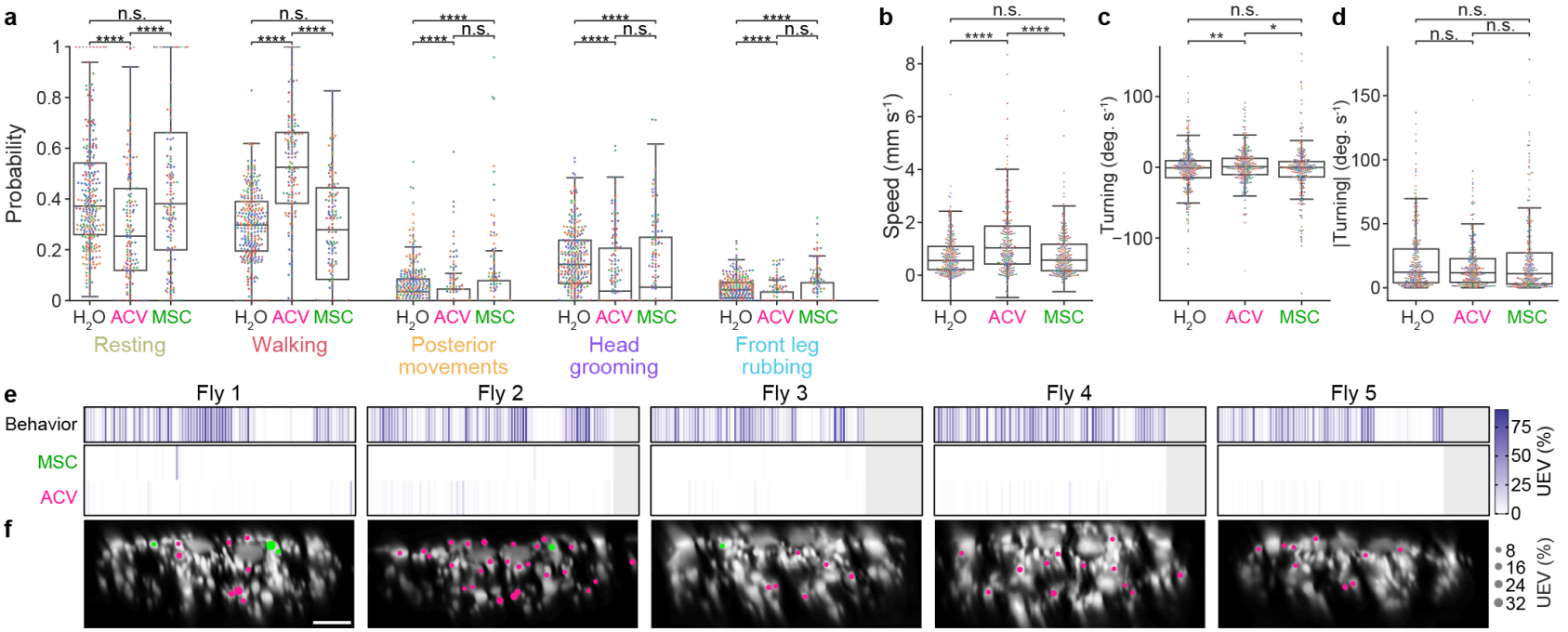

Figure S4.8: **Odor-modulated behaviors and encoding in descending neuron populations across individuals.** **(a)** The probabilities that classified behaviors occur during periods of stimulation with humidified air, ACV, or MSC odors. Stars indicate the significance level for a two-sided Mann-Whitney U test ('*': $p < 0.05$, '**': $p < 0.01$, '****': $p < 10^{-4}$). **(b-d)** Swarm plots showing a subset of 250 randomly sampled points indicating **(b)** walking speed, **(c)** turning angular velocity, and **(d)** absolute value of turning angular velocity during periods of stimulation with humidified air, ACV, or MSC odors. **(e)** Matrices showing the cross-validated ridge regression UEV of models that contain behavior and odor regressors for five individual animals. The first row ('Behavior') shows the composite R^2 for all behavior regressors and shuffled odor regressors. The second and third rows show UEVs for the odor regressors MSC and ACV, respectively. **(f)** Locations of odor encoding neurons (UEV > 5%) across five individual animals. Scale bar is 10 µm and applies to all images. Color indicates odor. Radii scale with the amount of variance explained.

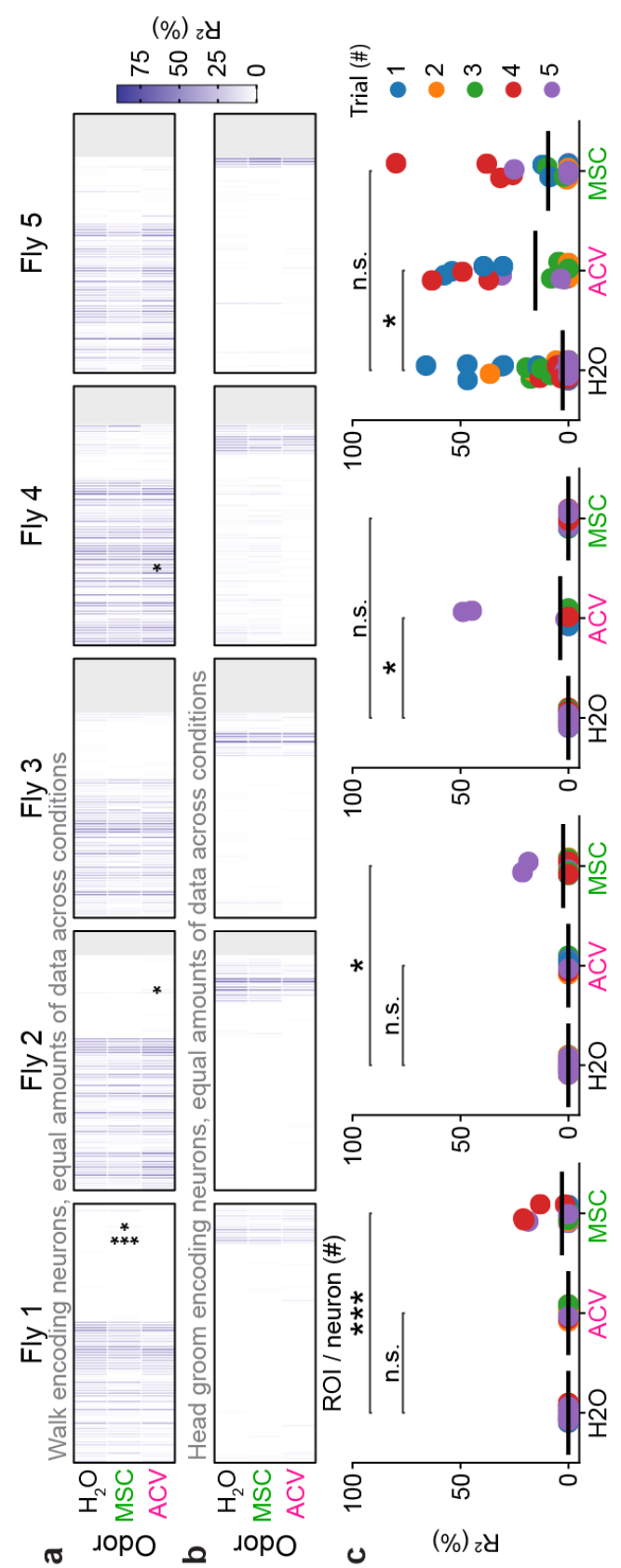

Figure S4.9: **Largely identical descending neuron populations are recruited during walking and head grooming irrespective of odor context.** **(a-b)** Amount of **(a)** walking or **(b)** head grooming variance explained by each DN using only frames during presentation of humidified air, apple cider vinegar (ACV), or methyl salicylate (MSC). Data during humidified air presentation were split into groups to match the amount of data available during ACV and MSC presentation. Indicated are cases where a two-sided Mann-Whitney U test comparing the cross-validation folds between humidified air and each of the odors yielded significant differences after Bonferroni correction for multiple comparisons (*: $p < 0.05$; ***: $p < 0.001$). **(c)** Individual ROI data points resulting in significant differences across conditions by a two-sided Mann-Whitney U test (shown from left to right: fly 1, ROI 82; fly 1, ROI 87; fly 2, ROI 74; fly 4, ROI 28). Each point is the result of a single cross-validation fold from a single trial.

4.11.3 Supplementary Videos

Video 1: Representative recording and processing of descending neuron population activity and animal behavior.
Fly behavior as seen by camera 5 **(top left)**. Odor stimulus presentation is indicated **(top right)**. Fictive walking trajectory of the fly calculated using FicTrac **(middle left)**. Trajectory turns gray for points after 2 s. 3D pose of the fly calculated using DeepFly3D **(bottom left)**. Text indicates the current behavior class (top right). Raw two-photon microscope image after center-of-mass alignment ((**top right)**. $\Delta F/F$ image after motion correction and denoising of the green channel **(middle right)**. Linear discriminant analysis-based low-dimensional representation of the neural data **(bottom right)**. Each dimension is a linear combination of neurons. The dimensions are chosen such that frames associated with different behaviors are maximally separated.
https://www.dropbox.com/s/vx2zd2tt7vg0bym/Video1-Summary.mov?dl=0

Video 2: Behavior-triggered average $\Delta F/F$ during walking epochs.
Averaged $\Delta F/F$ images aligned with respect to behavior onset for all walking epochs. Red circles (top left in each imaging panel) indicate the onset of behavior. When the red circle becomes cyan, less then seven behavior epochs remain and the final image with more than eight epochs is shown. Shown as well is an example behavior epoch **(top left)** indicating the time with respect to the onset of behavior and synchronized with $\Delta F/F$ panels.
https://www.dropbox.com/s/253ondwagchipfk/Video2_Walking.mov?dl=0

Video 3: Behavior-triggered average $\Delta F/F$ during resting epochs.
Averaged $\Delta F/F$ images aligned with respect to behavior onset for all resting epochs. Red circles (top left in each imaging panel) indicate the onset of behavior. When the red circle becomes cyan, less then seven behavior epochs remain and the final image with more than eight epochs is shown. Shown as well is an example behavior epoch **(top left)** indicating the time with respect to the onset of behavior and synchronized with $\Delta F/F$ panels.
https://www.dropbox.com/s/gklmy23f8pj20m8/Video3_Resting.mov?dl=0

Video 4: Behavior-triggered average $\Delta F/F$ during head grooming epochs.
Averaged $\Delta F/F$ images aligned with respect to behavior onset for all head grooming epochs. Red circles (top left in each imaging panel) indicate the onset of behavior. When the red circle becomes cyan, less then seven behavior epochs remain and the final image with more than eight epochs is shown. Shown as well is an example behavior epoch **(top left)** indicating the time with respect to the onset of behavior and synchronized with $\Delta F/F$ panels.
https://www.dropbox.com/s/quo2889b6im1z8w/Video4_HeadGroom.mov?dl=0

Video 5: Asymmetric activity in a pair of DNx01s is associated with asymmetric leg-antennal collisions during antennal grooming.
Three head grooming epochs in the same animal having leg contact with **(top)** primarily the left antenna, **(middle)** both antennae, or **(bottom)** primarily the right antenna. Shown are corresponding **(right)** neural activity $\Delta F/F$ images (note putative DNx01s in dashed white circles), **(middle)** behavior videos (camera 2), and **(left)** leg-antennal collisions (green) during kinematic replay of 3D poses in the NeuroMechFly physics simulation. Video playback is 0.25x real-time.
https://www.dropbox.com/s/5l72w86njxjhoxx/Video5-AsymmetricGrooming.mov?dl=0

5 Conclusions and future perspectives

Here, we have developed an approach that allows imaging neural activity in the thoracic cervical connective and the T1 neuromere of the ventral nerve cord in tethered but otherwise freely behaving *Drosophila*. This approach proved to be particularly useful for the investigation of ascending and descending neurons as they pass through the cervical connective. The reason for this is two fold: (i) imaging the connective allows the exclusion of non-descending/ascending interneurons targeted by the genetic driver line, (ii) the axons running in parallel through the connective can be imaged using 2D x-z scanning, which allows for faster frame rates and facilitates post-hoc motion correction, while individual neurons can be clearly separated without tracing of neurites.
We then used this approach to image the neural activity of a collection of sparse ascending neuron driver lines and found that ANs encode multi-joint behaviors, including walking, eye grooming, general front leg movements, and turning. The majority of ANs we study project to one of two brain regions, the anterior ventrolateral protocerebrum (AVLP) and the Gnathal Ganglia (GNG). Neurons projecting to the AVLP predominantly encoded self-motion in the form of walking and resting, while GNG projecting neurons mostly encoded more discrete actions such as grooming.
Our recordings from the population of supraesophageal descending neurons revealed neurons encoding walking, head grooming, and resting. The largest fraction of DNs encodes walking with some neurons encoding turning and speed suggesting that behaviors might be driven by a core set of DNs and modulated by additional DNs. Besides behavior, DNs also encoded sensory information in the form of odors and antennal touches. The two odors we presented were encoded by distinct sub-populations that had no relation to behavior. We did not observe any context dependent encoding of behaviors.

The fact that we found encoding of behavior rather than joint angles for both ANs as well as DNs suggests, within the limitations discussed below, that DNs and ANs might encode behavior in a similar fashion. Since both populations of neurons have to encode fly behavior in an efficient way, it makes sense that they might use similar encoding strategies. Using the same encoding strategy might also be beneficial when AN activity is compared to the efference

copy of descending commands. The efference copy could directly be compared to the AN activity and no complex forward model is needed in the brain.

We observed that many of the neurons we recorded were active during walking. This is consistent with the fact that large numbers of brain neurons are active during locomotion in flies [159, 229, 230, 239], mice [156], and C. elegans [249]. All these studies use calcium recordings as a surrogate of neural activity. This affords the ability to record from many neurons simultaneously, but lacks temporal resolution. Furthermore, recording from neurons during behavior without any perturbation only establishes a correlation between neural activity and behavior but is unable to reveal causal links. This is also true for this work. However, one can make the general assumption that at least some of the DNs recorded are causally related to the behaviors. This notion is corroborated by numerous optogenetic activation studies of individual pairs of DNs [53, 74, 75, 77, 78, 88]. For ANs, on the other hand, it can generally be assumed that they report behavioral states rather than drive behaviors. The large numbers of brain interneurons correlated with behavior likely fall into these two categories as well. The interneurons driving behavior are probably upstream of DNs, while those reporting behavior are presumably downstream of ANs.

Whether the encoding of behavior is consistent across areas of the nervous system and species remains to be seen. Our DN results suggest that at least some of the encoding principles might be conserved between mammels and *Drosophila*. In mice, turning is controlled by a predominantly ipsilaterally projecting group of DNs [102]. We found that the *Drosophila* DNs correlated turning are also predominantly ipsilaterally located. Ragarding speed on the other hand, the results are less consistent. In mice, two groups of DNs have been identified to control speed and gait [215], while our results did not show obvious subpopulations of DNs controlling speed. Furthermore, mice have DNs which can halt locomotion [105]. These might correspond to the rest-encoding DNs we found.

5.1 Limitations

Perhaps the biggest limitation of this work is the temporal resolutions at which we can infer neural activity. The cause of this is two fold: (i) the calcium indicator dynamics are slow, effectively low-pass filtering the signal, (ii) the acquisition frame rate of the microscope is slow compared to the movements of the fly. This mostly affects the conclusion regarding the encoding of joint angles. From previous work we know that at least one ascending neuron encodes a joint angle [169]. In order to conclusively answer the question whether the neurons we recorded encode joint angles, higher resolution data would probably be required.

Another limitation regarding our data is the sampling of behaviors. Each individual and each trial shows behavior in different frequencies with the possibility that some behaviors are not present at all. It can be assumed that these distributions of behaviors are different from the

distribution of freely behaving flies. In particular, the fly was not able to fly, groom its wings, and could not rotate its head as it usually does during head grooming. Furthermore, we observed unethological pushing and abdominal bending behaviors related to the treadmill setup.

Not only the sampling of behaviors is subject to bias, but also the selection of neurons we recorded from. Since the collection of ANs we used is not comprehensive, there is a possibility that the generation of ascending split GAL4 lines was unintentionally biased toward ascending neurons targeting the AVLP and GNG.
A similar problem arises for our DN recordings. First, the flippase based approach is inherently stochastic. This became apparent as we saw that in some flies, only one of the giant fibers expressed fluorescent protein. Second, we are unable to capture as many DNs as could be expected based on the literature [59]. This suggests that our resolution might introduce a bias toward larger axons or those with particularly high levels of GCaMP.

Lastly, functional calcium imaging is most commonly done using cell bodies as the region of interest. In flies the choice of region of interest is less obvious since the cell bodies are positioned around the neuropil and action potentials are generated within the neuropil away from the cell body [250]. Here we record calcium signals in axons and it is not clear what the relationship of axonal calcium concentration is to other compartments. We make the implicit assumption that calcium concentration in the axon is a good indicator for neural activity.

5.2 Future work

5.2.1 Technical improvements

Some of the technical limitations described above can be overcome or mitigated trough newer technology or different techniques. Over the last couple of years, voltage indicators started to gain traction in the field [237, 251, 252]. It remains to be seen if their signal-to-noise ratio is sufficient for two-photon imaging in the cervical connective. If they do provide sufficient signal to noise ratios, their faster response time will help study neurons that encode features on the time scale of individual joint movements. At the same time, this would render the question of calcium origin obsolete. However, their faster dynamics also mandate a faster imaging speed. Approaches that can lead to an increase in imaging speed include Bessel beam excitation [253], acousto-optic deflectors [254], and multibeam imaging [255].

5.2.2 Scientific perspective

Ascending neurons

Here we have proposed that ANs projecting to the AVLP encode self-motion which might modulate optical sensory gain and that ANs projecting to the GNG might contribute to action selection. These two hypotheses remain to be tested.

Since it is already known that vertical-system visual neurons are gain modulated during flight [126] and horizontal-system visual neurons are modulated during walking [55, 158], one could start by looking at the postsynaptic partners of the ascending neurons projecting to the AVLP. Postsynaptic partners projecting to the visual systems would then lend themselves to further investigation.
A first set of experiments to tackle the question of the involvement of ANs projecting to the GNG in action selection could look quite similar. If any of the postsynaptic partners in the GNG are DNs, they might be inhibited by the ANs to rule out motor commands that are incompatible with the current behavior state. This, however, bares the question of why such inhibition happens in the brain and not in the VNC, rendering this outcome unlikely compared to a more complex action selection circuitry. Regardless of whether there are direct connections between ANs and DNs in the GNG, it will be interesting to explore the connectome for GNG interneurons connecting the two populations.

Descending neurons

As described in the introduction, DNs can be split into supraesophageal and subesophageal DNs. Here, we have only investigated supraesophageal DNs. As a logical progression, one could image the population of subesophageal DNs to see if they encode different aspects of behavior. This could be done using the Dfd and Scr lines described in [240]. It will be particularly interesting to see if the groups of parallel DNs originating in the GNG fundamentally differ from our observations in the supraesophageal DNs. For the DNg02 population Namiki et al. found a population code controlling wingbeat amplitude [95].

Optogenetic activation studies have provided much insight of how DNs might control behavior. However, it remains to be seen how relevant the results are when compared to the neural population activity during naturally occurring behaviors. The dense axonal innervation of the GNG by descending neurons suggests that one descending neuron might recruit or inhibit others as it passes through the GNG. This hypothesis could be tested through optogenetic activation of a candidate DN while imaging the GNG descending neurons. Particularly interesting will be the activation of command or command-like DNs and the associated transition between behaviors at different light intensities. One can imagine several outcomes of this experiment: (i) only the stimulated neurons are active and the fly performs the stimulated behavior, (ii) the stimulated neurons and additional new DNs are active, (iii) the stimulated neurons and the neurons related to the behavior at stimulation onset are active. Each one of these has different implications for how DNs might control behavior. Outcome (i) suggests that a single pair of neurons can indeed be sufficient to activate complex behavioral programs in the VNC, while (ii) suggests that additional DNs are presumably needed and are recruited by the stimulated neurons. Finally, outcome (iii) will show whether the VNC can resolve conflicting DN activity depending on the behavior response and the stimulation intensity. Guo et al. [78] provide evidence that such conflicting DN activity can be mediated by the VNC in some cases.

In general, perturbations, like the optogenetic activation described above, will be needed to

better characterize how descending neurons control behaviors. Instead of activation, one could also silence individual neurons using GtACR [256] or lesioning with a spatial light modulator [257]. Ideally, the neural encoding would be analyzed while the fly is still under the microscope. This way we can identify potential targets involved in a particular behavior. These target neurons could then be silenced one at a time to determine which or how many neurons are required for the behavior. If data analysis during imaging proves unfeasible, the same fly could be imaged on a later day [187] once the data are analyzed. Alternatively, neurons in the connective have to be identified across flies such that a hypothesis developed based on the data of one fly can later be tested in another fly.
Despite *Drosophila*'s stereotyped nervous system, it is difficult to uniquely match neurons within the connective due to the limited morphological information contained in the x-z cross section images. So far we are only able to empirically identify few neurons like the large ventral grooming neurons based on their position, size and encoding. The few neurons we can identify show an unusual degree of variability in position for a stereotyped nervous system. This variability is likely related to the lack of synapses constraining relative position in the cervical connective. In a first step toward identifying neurons across animals, we have developed a deep learning approach that can predict behavior from $\Delta F/F$ images regardless of fly identity (see Appendix A). More work on this problem will be required for us to be able to fully leverage the stereotypy of the *Drosophila* nervous system.

Besides neural perturbations, behavioral perturbations can be used to characterize the level of descending control exerted on the VNC circuitry. Such perturbations can be introduced through gaps [258] and obstacles [259] in the fly's way, forcing it to adapt its steps accordingly. These adaptions are likely a result of visual input and/or reflexes. While visual input can only change the behavior through descending commands, reflexes are likely modifying the behavior locally within the VNC. In order to discern the contributions, flies could be recorded with or without light and with sets of leg sensory neurons silenced. Once the behavioral adaptation is characterized, one can try to find its neural signature using the imaging approach employed in this work.

Lastly, the ethological role of DNx01 remains unclear in this work. We hypothesize that it might lead to bracing during strong gusts of wind. To test this hypothesis, DNx01 can be optogenetically activated in different behavioral contexts. If a bracing response in behavior contexts other than antennal grooming is found, it is likely gated in the VNC since we observed activity of DNx01 in the cervical connective during antennal grooming. This gating would probably be mediated by DNs encoding grooming.

A Overcoming the Domain Gap in Neural Action Representations

Disclaimer: This chapter is based on the following article with permissions of all co-authors.

Semih Günel, **Florian Aymanns**, Sina Honari, Pavan Ramdya[1], Pascal Fua[1] "Overcoming the Domain Gap in Neural Action Representations", *arXiv*, 2112.01176 , 2021. The article can be found here: https://doi.org/10.48550/arXiv.2112.01176

[1]These authors contribute equally.

My contribution:
I collected and pre-processed all of the fly data for this paper. I also contributed to writing the methods.

A.1 Abstract

Relating behavior to brain activity is a fundamental goal in neuroscience, with practical applications in building robust brain-machine interfaces. However, the domain gap between individuals is a major issue that prevents the training of general models that work on unlabeled subjects.

Since 3D pose data can now be reliably extracted from multi-view video sequences without manual intervention, we propose to use it to guide the encoding of neural action representations together with a set of neural and behavioral augmentations exploiting the properties of microscopy imaging. To reduce the domain gap, during training, we mix features of neural and behavioral data across animals that seem to be performing similar actions.

To demonstrate this, we test our methods on three very different multimodal datasets; one that features flies and their neural activity, one new dataset that contains human neural Electrocorticography (ECoG) data, and lastly the RGB video data of human activities from different viewpoints.

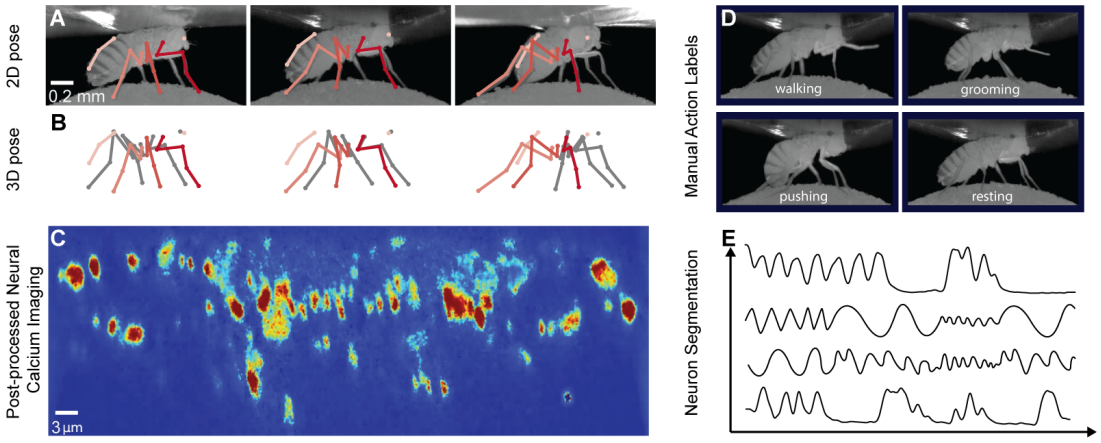

Figure A.1: **Our Motion Capture and Two-Photon (MC2P) Dataset.** A tethered fly (*Drosophila melanogaster*) is recorded using six multi-view infrared cameras and a two-photon microscope. The resulting dataset includes the following. **(A)** 2D poses extracted from different views (only three are shown), calculated on grayscale images. **(B)** 3D poses triangulated from the 2D views. **(C)** Synchronized, registered, and denoised single-channel fluorescence calcium imaging data using a two-photon microscope. Shown are color-coded activity patterns for populations of descending neurons from the brain. These carry action information (red is active, blue is inactive). **(D)** Annotations for eight animals of eight different behaviors, four of which are shown here. **(E)** Manual neural segmentation has been performed to extract neural activity traces for each neuron. We will release our MC2P publicly. Examples videos of selected actions and multi-modal data are in the Supplementary Material.

A.2 Introduction

Neural decoding of action, the accurate prediction of behavior from brain activity, is a fundamental challenge in neuroscience with important applications in the development of robust brain machine interfaces [260–263]. Recent technological advances have enabled simultaneous recordings of neural activity and behavioral data in experimental animals and humans [29, 171, 264–268]. Nevertheless, our understanding of the complex relationship between behavior and neural activity remains limited.

A major reason is that it is difficult to obtain many recordings from mammals and a few subjects are typically not enough to perform meaningful analyses [269]. This is less of a problem when studying the fly *Drosophila melanogaster*, for which long neural and behavioral datasets can be obtained for many individual animals **(Fig. A.1)**. Nevertheless, current supervised approaches for performing neural decoding [270, 271] still do not generalize well across animals because each nervous system is unique **(Fig. A.2)**. This creates a significant domain-gap that necessitates tedious and difficult manual labeling of actions. Furthermore, a different model must be trained for each individual animal, requiring more annotation and overwhelming the resources of most laboratories.

Another problem is that experimental neural imaging data often has unique temporal and

spatial properties. The slow decay time of fluorescence signals introduces temporal artifacts. Thus, neural imaging frames include information about an animal's previous behavioral state. This complicates decoding and requires specific handling that standard machine learning algorithms do not provide.

To address these challenges, we propose to learn neural action representations—embeddings of behavioral states within neural activity patterns—in an self-supervised fashion. To this end, we leverage the recent development of computer vision approaches for automated, markerless 3D pose estimation [37, 39] to provide the required supervisory signals without human intervention. We first show that using contrastive learning to generate latent vectors by maximizing the mutual information of simultaneously recorded neural and behavioral data modalities is not sufficient to overcome the domain gap between animals and to generalize to unlabeled animals at test time **(Fig. A.3B)**. To address this problem, we introduce two sets of techniques:

1. To close the domain gap between animals, we leverage 3D pose information. Specifically, we use pose data to find sequences of similar actions between a source and multiple target animals. Given these sequences, we mix and replace neural or behavioral data of the source animal with the ones composed of multiple target animals. To make this possible, we propose a new Mixup strategy which merges selected samples from multiple target animals, practically hiding the identity information. This allows us to train our decoder to ignore animal identity and close the domain gap.

2. To mitigate the slowly decaying calcium data impact from past actions on neural images, we add simulated randomized versions of this effect to our training neural images in the form of a temporally exponentially decaying random action. This trains our decoder to learn the necessary invariance and to ignore the real decay in neural calcium imaging data. Similarly, to make the neural encoders robust to imaging noise resulting from low image spatial resolution, we augment random sequences into sequences of neural data to replicate this noise.

The combination of these techniques allowed us to bridge the domain gap across animals in an unsupervised manner **(Fig. A.3D)**, making it possible to perform action recognition on unlabeled animals better than earlier techniques, including those requiring supervision [271–273]. To test the generalization capacity of neural decoding algorithms, we record and use MC2P dataset, which we will make publicly available [1]. It includes two-photon microscope recordings of multiple spontaneously behaving *Drosophila*, and associated behavioral data together with action labels.

Finally, to demonstrate that our technique generalizes beyond this one dataset, we tested it on two additional ones: One that features neural ECoG recordings and 2D pose data for

[1] During the review process, we attach our MC2P dataset in the supplementary material. The dataset will be made public upon acceptance.

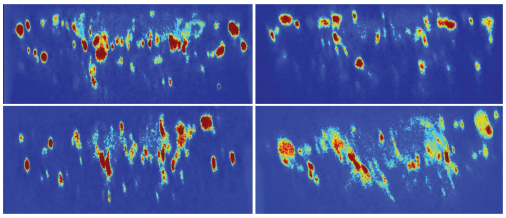

Figure A.2: **Domain gap between nervous systems.** Neural imaging data from four different animals in each corner. Images differ in terms of total brightness, the location of observed neurons, the number of visible neurons, and the shape and size of axons.

epileptic patients [274, 275] along with the well-known H36M dataset [276] in which we treat the multiple views as independent domains. Our method markedly improves across-subject action recognition in all datasets.

We hope our work will inspire the use and development of more general self-supervised neural feature extraction algorithms in neuroscience. These approaches promise to accelerate our understanding of how neural dynamics give rise to complex animal behaviors and can enable more robust neural decoding algorithms to be used in brain-machine interfaces.

A.3 Related Work

Neural Action Decoding:

The ability to infer behavioral intentions from neural data, or neural decoding of behavior, is essential for the development of effective brain-machine interfaces and for closed-loop experimentation [277, 278]. Neural decoders can be used to increase mobility of patients with disabilities [279, 280], or neuromuscular diseases [281], and can expand our understanding of how the nervous system works [282]. However, most neural decoding methods require manual annotations of training data that are both tedious to acquire and error prone [271, 283].

Existing self-supervised neural decoding methods [273, 284–286] cannot be used on unlabeled subjects without action labels. A potential solution would be to use domain adaptation techniques to treat each new subject as a new domain. However, existing domain adaptation studies of neural decoding [287, 288] have focused on gradual domain shifts associated with slow changes in sensor measurements rather than the challenge of generalizing across individual subjects. In contrast to these methods, our approach is self-supervised and can generalize to unlabeled subjects at test time, without requiring action labels for new individuals.

Action Recognition:

Contrastive learning has been extensively used on human motion sequences to perform action recognition using 3D pose data [289–291] and video-based action understanding [292, 293].

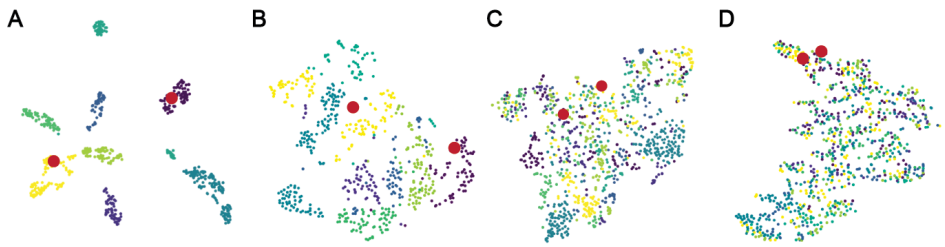

Figure A.3: **t-SNE plots of the neural data.** Each color denotes a different fly. The two red dots are embeddings of the same action label in two different animals. **(A)** Raw neural data. **(B)** SimCLR [297] representation, **(C)** Domain adaptation using a two-layer MLP discriminator and a Gradient Reversal Layer. **(D)** Ours. The identify of the animals is discarded and the semantic structure is preserved, as evidenced by the fact that the two red dots are very close.

However, a barrier to using these tools in neuroscience is that the statistics of our neural data—the locations and sizes of cells—and behavioral data—body part lengths and limb ranges of motion—can be very different from animal to animal, creating a large domain gap.

In theory, there are multimodal domain adaptation methods for action recognition that could deal with this gap [294–296]. However, they assume supervision in the form of labeled source data. In most laboratory settings, where large amounts of data are collected and resources are limited, this is an impractical solution.

Representation Learning:

Most efforts to derive a low dimensional representation of neural activity have used recurrent models [298–300], variational autoencoders [265, 301], and dynamical systems [302, 303]. To capture low-dimensional behavioral information, recent methods have enabled markerless predictions of 2D [40, 304–307] and 3D poses in animals [37, 38, 41, 308]. Video and pose data have previously been used to segment and cluster temporally related behavioral information [309–313].

By contrast, there have been relatively few approaches developed to extract behavioral representations from neural imaging data [271, 272, 314]. Most have focused on identifying simple relationships between these two modalities using simple supervised methods, such as correlation analysis, generalized linear models [136, 155, 156], or regressive methods [272]. We present a joint modeling of motion capture and neural modalities to fully extract behavioral information from neural data using a self-supervised learning technique.

Mixup Training:

Mix-up regularization was first proposed as a way to learn continuous latent spaces and to improve generalization for supervised learning [315–317]. Several previous studies have used a Mixup strategy to generate new positive pairs in contrastive learning [318, 319]. Mixup

has rarely been used for domain adaptation. Recent examples include temporal background mixing [320], prediction smoothing across domains [321], and training better discriminators on uni-modal datasets [322]. Our Mixup strategy can be regarded as a multi-modal extension of the previous approaches [317, 320] where per-frame feature-level stochastic Mixup between domains was performed to explore shared space and to hide identity information. Unlike these approaches, we explicitly condition the sampling procedure on the input data. We demonstrate that this approach helps learn domain-invariant neural features.

A.4 Approach

Our ultimate goal is to interpret neural data so that, given a neural image, one can generate latent representations that are useful for downstream tasks. This is challenging due to the wide domain-gap in neural representations between different animals **(Fig. A.2)**. Hence, we aim to leverage self-supervised learning techniques to derive rich features that, once trained, could be used on downstream tasks including action recognition to predict the behaviors of unlabeled animals.

Our data is composed of two-photon microscopy neural images synchronized with 3D behavioral data, where we do not know where each action starts and ends. We leveraged contrastive learning to generate latent vectors from both modalities such that their mutual information would be maximized and therefore describe the same underlying action. However, this is insufficient to address the domain-gap between animals **(Fig. A.3B)**. To do so, we implement an across-domain mixing strategy: We replace the original pose or neural data of an animal with mix of another set of animals for which there is a high degree of 3D pose similarity at each given instance in time. Unlike behavioral data, neural data has unique properties. Neural calcium data contains information about previous actions because it decays slowly across time and it involves limited spatial resolution. To teach our model the invariance of these artifacts of neural data, we propose two data augmentation techniques: (i) Neural Calcium augmentation - given a sequence of neural data, we apply an exponentially decaying neural snapshot to the sequence, which imitates the decaying impact of previous actions, (ii) Neural Noise augmentation - to make the model more robust to noise, we applied an augmentation which merges a sequence of neural data with another randomly sampled neural sequence using a coefficient.

Together, these augmentations enable a self-supervised approach to (i) bridge the domain gap between animals allowing testing on unlabeled ones, and (ii) imitate the temporal and spatial properties of neural data, diversifying it and making it more robust to noise. In the following section, we describe steps in more detail.

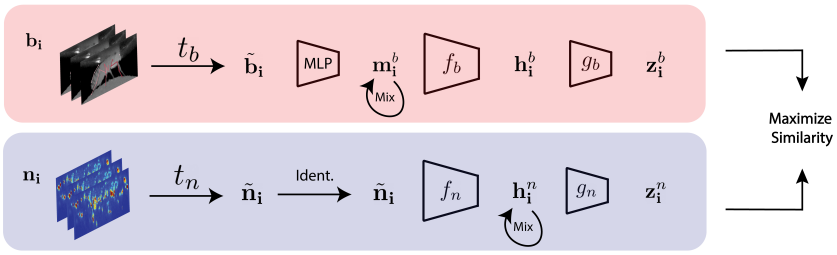

Figure A.4: **Our approach to learning an effective representation of behaviors.** First, we sample a synchronized set of behavioral and neural frames, $(\mathbf{b_i}, \mathbf{n_i})$. Then, we augment these data using randomly sampled augmentation functions t_b and t_n. Encoders f_b and f_n then generate intermediate representations $\mathbf{h}^b$ and $\mathbf{h}^n$, which are then projected into $\mathbf{z}_b$ and $\mathbf{z}_t$ by two separate projection heads g_b and g_n. For the behavioral modality, we first apply a frame-wise MLP before f_b. We then apply mixing on $\mathbf{m}_\mathbf{i}^b$. For the neural modality, we apply mixing on $\mathbf{h}_\mathbf{i}^n$ without an MLP, since mixing cannot be done at frame level. We maximize the similarity between the two projections using an InfoNCE loss. At test time, the red branch and $\mathbf{h}_\mathbf{i}^n$ is used for neural decoding.

A.4.1 Problem Definition

We assume a paired set of data $\mathscr{D}_s = \{(\mathbf{b}_\mathbf{i}^s, \mathbf{n}_\mathbf{i}^s)\}_{i=1}^{n_s}$, where $\mathbf{b}_\mathbf{i}^s$ and $\mathbf{n}_\mathbf{i}^s$ represent behavioral and neural information respectively, with n_s being the number of samples for animal $s \in \mathscr{S}$. We quantify behavioral information $\mathbf{b}_\mathbf{i}^s$ as a set of 3D poses $\mathbf{b}_k^s$ for each frame $k \in \mathbf{i}$ taken of animal s, and neural information $\mathbf{n}_\mathbf{i}^s$ as a set of two-photon microscope images $\mathbf{n}_k^s$, for all frames $k \in \mathbf{i}$ capturing the activity of neurons. The data is captured such that the two modalities are always synchronized (paired) without human intervention, and therefore describe the same set of events. Our goal is to learn an unsupervised parameterized image encoder function f_n, that maps a set of neural images $\mathbf{n}_\mathbf{i}^s$ to a low-dimensional representation. We aim for our learned representation to be representative of the underlying action label, while being agnostic to both modality and the identity. We assume that we are not given action labels during self-supervised training. Also note that we do not know at which point in the captured data an action starts and ends. We just have a series of unknown actions performed by different animals.

A.4.2 Contrastive Representation Learning

For each input pair $(\mathbf{b}_\mathbf{i}^s, \mathbf{n}_\mathbf{i}^s)$, we first draw a random augmented version $(\tilde{\mathbf{b}}_\mathbf{i}^s, \tilde{\mathbf{n}}_\mathbf{i}^s)$ with a sampled transformation function $t_n \sim \mathscr{T}_n$ and $t_b \sim \mathscr{T}_b$, where $\mathscr{T}_n$ and $\mathscr{T}_b$ represent a family of stochastic augmentation functions for behavioral and neural data, respectively, which are described in the following sections. Next, the encoder functions f_b and f_n transform the input data into low-dimensional vectors $\mathbf{h}_b$ and $\mathbf{h}_n$, followed by non-linear projection functions g_b and g_n, which further transform data into the vectors $\mathbf{z}_b$ and $\mathbf{z}_n$. For the behavioral modality, in order to facilitate mixing, we first transform augmented input data $\tilde{\mathbf{b}}_\mathbf{i}^s$ into $\mathbf{m}_\mathbf{i}^s$ using a shallow frame-wise MLP, as shown in **Fig A.4**. For the neural modality, we instead directly apply mixing

using $\mathbf{h}_n$, since frame-level mixing is not possible. We give the details of the mixing strategy in the next sections. During training, we sample a minibatch of N input pairs $\left(\mathbf{b}_{\mathbf{i}}^s, \mathbf{n}_{\mathbf{i}}^s\right)$, and train with the loss function

$$\mathscr{L}_{NCE}^{b\to n} = -\sum_{i=1}^{N} \log \frac{\exp\left(\left\langle \mathbf{z}_b^i, \mathbf{z}_n^i \right\rangle / \tau\right)}{\sum_{k=1}^{N} \exp\left(\left\langle \mathbf{z}_b^i, \mathbf{z}_n^k \right\rangle / \tau\right)} \tag{A.1}$$

where $\left\langle \mathbf{z}_b^i, \mathbf{z}_n^i \right\rangle$ is the cosine similarity between behavioral and neural modalities and $\tau \in \mathbb{R}^+$ is the temperature parameter. Intuitively, the loss function measures classification accuracy of a N-class classifier that tries to predict $\mathbf{z}_n^i$ given the true pair $\mathbf{z}_b^i$. To symmetrize the loss function with respect to the negative samples, we also define

$$\mathscr{L}_{NCE}^{n\to b} = -\sum_{i=1}^{N} \log \frac{\exp\left(\left\langle \mathbf{z}_b^i, \mathbf{z}_n^i \right\rangle / \tau\right)}{\sum_{k=1}^{N} \exp\left(\left\langle \mathbf{z}_b^k, \mathbf{z}_n^i \right\rangle / \tau\right)}. \tag{A.2}$$

We take the combined loss function to be $\mathscr{L}_{NCE} = \mathscr{L}_{NCE}^{b\to n} + \mathscr{L}_{NCE}^{n\to b}$, as in [323, 324]. The loss function maximizes the mutual information between two modalities [325]. Although standard contrastive learning bridges the gap between different modalities, it does not bridge the gap between different animals **(Fig. A.3B)**. This is a fundamental challenge that we address in this work through augmentations described in the following section, which are part of the neural and behavioral family of augmentations $\mathscr{T}_n$ and $\mathscr{T}_b$.

Mixup Strategy:

Given a set of consecutive 3D poses $\mathbf{b}_{\mathbf{i}}^s$ and their features $\mathbf{m}_{\mathbf{i}}^s$ calculated by a shallow MLP from augmented $\tilde{\mathbf{b}}_{\mathbf{i}}^s$, for each $k \in \mathbf{i}$, we stochastically replace $\mathbf{m}_k^s$ with a mix of its two pose neighbors, sampled from two animals, in the set of domains $\mathscr{D}_{\mathscr{S}}$, where $\mathscr{S}$ is the set of all animals. To get one of the neighbors, we first uniformly sample a domain $\hat{s} \in \mathscr{S}$ and define a probability distribution $\mathbf{P}_{\mathbf{b}_k^s}^{\hat{s}}$ over the domain $\mathscr{D}_{\hat{s}}$ with respect to single 3D pose $\mathbf{b}_k^s$,

$$\mathbf{P}_{\mathbf{b}_k^s}^{\hat{s}}(\mathbf{b}_l^{\hat{s}}) = \frac{\exp(-\|\mathbf{b}_l^{\hat{s}} - \mathbf{b}_k^s\|_2)}{\sum_{\mathbf{b}_m^{\hat{s}} \in \mathscr{D}_{\hat{s}}} \exp(-\|\mathbf{b}_m^{\hat{s}} - \mathbf{b}_k^s\|_2)}. \tag{A.3}$$

We then sample from the above distribution and pass it through the shallow MLP, which yields $\mathbf{m}_l^{\hat{s}} \sim \mathbf{P}_{\mathbf{b}_k^s}^{\hat{s}}$. Notice that, although distribution is conditioned on the 3D pose $\mathbf{b}^s$, we sample back 3D pose features $\mathbf{m}^s$. In practice, we calculate the distribution $\mathbf{P}$ only over the first $\mathbf{N}$ nearest neighbors of $\mathbf{b}_k^s$, in order to sample from a distribution of the most similar poses. We empirically set $\mathbf{N}$ to 128. Given two samples $\mathbf{m}_l^{\hat{s}}$ and $\mathbf{m}_j^{\bar{s}}$ from the above distribution from independent domains, we then return the mixed version of

$$\tilde{\mathbf{m}}_k^s = \lambda \mathbf{m}_l^{\hat{s}} + (1-\lambda)\mathbf{m}_j^{\bar{s}}. \tag{A.4}$$

We sample the mixing coefficient λ from the Beta distribution $\lambda \sim \text{Beta}(\alpha, \beta)$. Our Mixup strategy removes the identity information in the behavioral data without perturbing it to the extent that semantic action information is lost. Since each behavioral sample $\mathbf{m}_{\mathbf{i}}^{s}$ is composed of a set of 3D pose features, and each 3D pose feature $\mathbf{m}_{k}^{s}, \forall k \in \mathbf{i}$ is replaced with a feature of a random domain, the transformed sample $\tilde{\mathbf{m}}_{\mathbf{i}}^{s}$ is now composed of multiple domains. This forces the behavioral encoding function f_b to leave identity information out, therefore generalizing across multiple domains **(Fig. A.5)**.

Our Mixup augmentation is similar to the synonym replacement augmentation used in natural language processing [326], where randomly selected words in a sentence are replaced by their synonyms, therefore changing the syntactic form of the sentence without altering the semantics. Instead, we randomly replace each 3D pose in a motion sequence. To the best of our knowledge, we are the first to use frame-wise mix strategy in the context of time-series analysis or for domain adaptation that is conditioned on the input.

To keep mixing symmetric, we also mix the neural modality. To mix a set of neural features $\mathbf{h}_{\mathbf{i}}^{s}$, we take its behavioral pair $\mathbf{b}_{\mathbf{i}}^{s}$, and search for similar sets of poses in other domains, with the assumption that similar sets of poses describe the same action. Therefore, once similar behavioral data is found, their neural data can be mixed. Note that, unlike behavior mixing, we do not calculate the distribution on individual 3D pose $\mathbf{b}_{k}^{s}$, but instead on the whole set of behavioral data $\mathbf{b}_{\mathbf{i}}^{s}$, because similarity in a single pose does not necessarily imply similar actions and similar neural data. More formally, given the behavioral-neural pair $\left(\mathbf{b}_{\mathbf{i}}^{s}, \mathbf{n}_{\mathbf{i}}^{s}\right)$, we mix the neural modality features $\mathbf{h}_{\mathbf{i}}^{s}$ by sampling two new neural features $\mathbf{h}_{\mathbf{j}}^{\bar{s}}$ and $\mathbf{h}_{\mathbf{l}}^{\hat{s}}$ from distinct animals $\hat{s}$ and $\bar{s}$, using the probability distribution

$$\mathbf{P}_{\mathbf{n}_{\mathbf{i}}^{s}}^{\hat{s}}(\mathbf{b}_{\mathbf{j}}^{\hat{s}}) = \frac{\exp(-\|\mathbf{b}_{\mathbf{j}}^{\hat{s}} - \mathbf{b}_{\mathbf{i}}^{s}\|_2)}{\sum_{\mathbf{b}_{\mathbf{m}}^{\hat{s}} \in \mathscr{D}_{\hat{s}}} \exp(-\|\mathbf{b}_{\mathbf{m}}^{\hat{s}} - \mathbf{b}_{\mathbf{i}}^{s}\|_2)}, \tag{A.5}$$

and then we return

$$\tilde{\mathbf{h}}_{k}^{s} = \lambda \mathbf{h}_{l}^{\hat{s}} + (1-\lambda)\mathbf{h}_{j}^{\bar{s}}\,. \tag{A.6}$$

Similarly, first we sample the mixing coefficient λ from the Beta distribution $\lambda \sim \text{Beta}(\alpha, \beta)$. This yields new mixed neural feature $\tilde{\mathbf{h}}_{\mathbf{i}}^{s}$, where the augmented neural data comes from two different animals in $\mathscr{S}$.

Neural Calcium Augmentation:

Our neural data was obtained using two-photon microscopy and fluorescence calcium imaging. The resulting images are only a function of the underlying neural activity, and have temporal properties that differ from the true neural activity. For example, calcium signals from a neuron change much more slowly than the neuron's actual firing rate. Consequently, a single neural image $\mathbf{n}_t$ includes decaying information concerning neural activity from the recent past, and thus carries information about previous behaviors. This makes it harder to

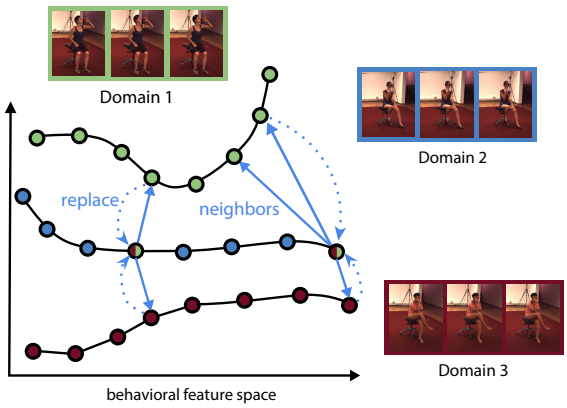

Figure A.5: **Our Mixup strategy.** Each 3D pose is processed by a pose-wise MLP to generate 3D pose features. Then, each 3D pose feature in the motion sequence of Domain 2 is randomly replaced with mix of two of its neighbors, from the set of domains $\hat{s} \in \mathscr{S}$, which includes Domains 1 and 3. The Mixup augmentation hides identity information, while keeping pose changes in the sequence minimal.

decode the current behavioral state.

We aimed to prevent this overlap of ongoing and previous actions. Specifically, we wanted to teach our network to be invariant with respect to past behavioral information by augmenting the set of possible past actions. To do this, we generated new data $\tilde{\mathbf{n}}_{\mathbf{i}}^{s}$, that included previous neural activity $\mathbf{n}_{k}^{s}$. To mimic calcium indicator decay dynamics, given a neural data sample $\mathbf{n}_{\mathbf{i}}^{s}$ of multiple frames, we sample a new neural frame $\mathbf{n}_{k}^{s}$ from the same domain, where $k \notin \mathbf{i}$. We then convolve $\mathbf{n}_{k}^{s}$ with the temporally decaying calcium convolutional kernel $\mathscr{K}$, therefore creating a set of images from a single frame $\mathbf{n}_{k}^{s}$, which we then add back to the original data sample $\mathbf{n}_{\mathbf{i}}^{s}$. This results in $\tilde{\mathbf{n}}_{\mathbf{i}}^{s} = \mathbf{n}_{\mathbf{i}}^{s} + \mathscr{K} * \mathbf{n}_{k}^{s}$ where $*$ denotes the convolutional operation. In the Supplementary Material, we explain calcium dynamics and our calculation of the kernel $\mathscr{K}$ in more detail.

Neural Noise Augmentation:

Two-photon microscopy images often include multiple neural signals combined within a single pixel. This is due to the the fact that multiple axons can be present in a small tissue volume that is below the spatial resolution of the microscope. To mimic this noise-adding effect, given a neural image $\mathbf{n}_{\mathbf{i}}^{s}$, we randomly sample a set of frames $\mathbf{n}_{\mathbf{k}}^{\hat{s}}$, from a random domain $\hat{s}$. We then return the blend of these two videos, $\tilde{\mathbf{n}}_{\mathbf{i}}^{s} = \mathbf{n}_{\mathbf{i}}^{s} + \alpha \mathbf{n}_{\mathbf{k}}^{\hat{s}}$, to mix and hide the behavioral information. Unlike the CutMix [327] augmentations used for supervised training, we apply the augmentation in an unsupervised setup to make the model more robust to noise. We sample a random α for the entire set of samples in $\mathbf{n}_{\mathbf{i}}^{s}$.

A.5 Experiments

We test our method on three datasets. In this section, we describe these datasets, the set of baselines against which we compare our model, and finally the quantitative comparison of all models.

A.5.1 Datasets

We ran most of our experiments on a large dataset of fly neural and behavioral recordings that we acquired and describe below. To demonstrate our method's ability to generalize, we also adapted it to run on another multimodal dataset that features neural ECoG recordings and markerless motion capture[274, 275], as well as the well known H36M human motion dataset [276].

MC2P:

Since there was no available neural-behavioral dataset with a rich variety of spontaneous behaviors from multiple individuals, we acquired our own dataset that we name *Motion Capture and Two-photon Dataset (MC2P)*. We will release this dataset publicly. MC2P features data acquired from tethered behaving adult flies, *Drosophila melanogaster* **(Fig. A.1)**, It includes:

1. Infrared video sequences of the fly acquired using six synchronized and calibrated infrared cameras forming a ring with the animal at its center. The images are 480×960 pixels in size and recorded at 100 fps.
2. Neural activity imaging obtained from the axons of descending neurons that pass from the brain to fly's ventral nerve cord (motor system) and drive actions. The neural images are 480×736 pixels in size and recorded at 16 fps using a two-photon microscope [171] that measures the calcium influx which is a proxy for the neuron's actual firing rate.

We recorded 40 animals over 364 trials, resulting in 20.7 hours of recordings with 7,480,000 behavioral images and 1,197,025 neural images. We provide additional details and examples in the Supplementary Material. We give an example video of synchronized behavioral and neural modalities in **Supplementary Videos 1-2**.

To obtain quantitative behavioral data from video sequences, we extracted 3D poses expressed in terms of the 3D coordinates of 38 keypoints [37]. We provide an example of detected poses and motion capture in **Supplementary Videos 3-4**. For validation purposes, we manually annotated a subset of frames using eight behavioral labels: *forward walking*, *pushing*, *hindleg grooming*, *abdominal grooming*, *rest*, *foreleg grooming*, *antenna grooming*, and *eye grooming*. We provide an example of behavioral annotations in **Supplementary Video 5**.

Figure A.6: **Domain Gap in the H3.6M dataset.** Similar to the domain gap across nervous systems, RGB images show a significant domain gap when the camera angle changes across individuals. We guide action recognition across cameras in RGB images using 3D poses and behavioral mixing.

ECoG dataset [274, 275]:

This dataset was recorded from epilepsy patients over a period of 7-9 days. Each patient had 90 electrodes implanted under their skull. The data comprises human neural Electrocorticography (ECoG) recordings and markerless motion capture of upper-body 2D poses. The dataset is labeled to indicate periods of voluntary spontaneous motions, or rest. As for two-photon images in flies, ECoG recordings show a significant domain gap across individual subjects. We applied our multi-modal contrastive learning approach on ECoG and 2D pose data along with mixing-augmentation. Then, we applied an across-subject benchmark in which we do action recognition on a new subject without known action labels.

H3.6M [276]:

H3.6M is a multi-view motion capture dataset that is not inherently multimodal. However, to test our approach in a very different context than the other two cases, we treated the videos acquired by different camera angles as belonging to separate domains. Since videos are tied to 3D poses, we used these two modalities and applied mixing augmentation together with multimodal contrastive learning to reduce the domain gap across individuals. Then, we evaluated the learned representations by performing action recognition on a camera angle that we do not have action labels for. This simulates our across-subject benchmark used in the MC2P dataset. For each experiment we selected three actions, which can be classified without examining large window sizes. We give additional details in the Supplementary Material.

A.5.2 Baselines

We evaluated our method using two supervised baselines, Neural Linear and Neural MLP. These directly predict action labels from neural data without any self-supervised pretraining using cross-entropy loss. We also compared our approach to three regression methods that attempt to regress behavioral data from neural data, which is a common neural decoding technique. These include a recent neural decoding algorithm, BehaveNet [272], as well as to two other regression baselines with recurrent and convolutional approaches: Regression (Recurrent) and Regression (Convolution). In addition, we compare our approach to recent self-supervised representation learning methods, including SeqCLR [285] and SimCLR [297]. We also combine convolutional regression-based method (Reg. (Conv)) or the self-supervised

learning algorithm SimCLR with the common domain adaptation techniques Gradient Reversal Layer (GRL)[328], or Mean Maximum Discrepancy [329]. This yields four domain adaptation models. Finally, we apply a recent multi-modal domain adaptation network for action recognition, MM-SADA[294] on MC2P dataset. For all of these methods, we used the same backbone architecture. We describe the backbone architecture in more detail in the Supplementary Material. We describe the baselines in more detail in following:

Supervised:

A feedforward network trained with manually annotated action labels using cross-entropy loss, having neural data as input. We discarded datapoints that did not have associated behavioral labels. For the MLP baseline, we trained a simple three layer MLP with a hidden layer size of 128 neurons with ReLU activation and without batch normalization.

Regression (Convolutional):

A fully-convolutional feedforward network trained with MSE loss for behavioral reconstruction task, given the set of neural images. To keep the architectures consistent with the other methods, the average pooling is followed by a projection layer, which is used as the final representation of this model.

Regression (Recurrent):

This is similar to the one above but the last projection network was replaced with a two-layer GRU module. The GRU module takes as an input the fixed representation of neural images. At each time step, the GRU module predicts a single 3D pose with a total of eight steps to predict the eight poses associated with an input neural image. This model is trained with an MSE loss. We take the input of the GRU module as the final representation of neural encoder.

BehaveNet [272]:

This uses a discrete autoregressive hidden Markov model (ARHMM) to decompose 3D motion information into discrete “behavioral syllables." As in the regression baseline, the neural information is used to predict the posterior probability of observing each discrete syllable. Unlike the original method, we used 3D poses instead of RGB videos as targets. We skipped compressing the behavioral data using a convolutional autoencoder because, unlike RGB videos, 3D poses are already low-dimensional.

SimCLR [297]:

We trained the original SimCLR module without the calcium imaging data and mixing augmentations. As in our approach, we took the features before the projection layer as the final representation.

Gradient Reversal Layer (GRL) [328]:

Together with the contrastive loss, we trained a two-layer MLP domain discriminator per modality, D_b and D_n, which estimates the domain of the neural and behavioral representations. Discriminators were trained by minimizing

$$\mathscr{L}_D = \sum_{x\in\{\mathbf{b},\mathbf{n}\}} -d\log\left(D_m\left(f_m(x)\right)\right) \tag{A.7}$$

where d is the one-hot identity vector. Gradient Reversal layer is inserted before the projection layer. Given the reversed gradients, the neural and behavioral encoders f_n and f_b learn to fool the discriminator and outputs invariant representations across domains, hence acting as a domain adaptation module. We kept the hyperparameters of the discriminator the same as in previous work [294]. We froze the weights of the discriminator for the first 10 epochs, and trained only the $\mathscr{L}_{NCE}$. We trained the network using both loss functions, $\mathscr{L}_{NCE} + \lambda_D \mathscr{L}_D$, for the remainder of training. We set the hyperparameters λ_D to 10 empirically.

Maximum Mean Discrepancy (MMD) [329]:

We replaced adversarial loss in GRL baseline with a statistical test that minimizes the distributional discrepancy from different domains.

MM-SADA [294]:

A recent multi-modal domain adaptation model for action recognition that minimizes cross-entropy loss on target labels, adverserial loss for domain adaptation, and contrastive losses to maximize consistency between multiple modalities. As we do not assume any action labels during the contrastive training phase, we removed the cross-entropy loss.

SeqCLR [285]:

This approach learns a uni-modal self-supervised contrastive model. Hence, we only apply it to the neural imaging data, without using the behavioral modality. As this method was previously applied on datasets with Electroencephalography (ECoG) imaging technique, we removed ECoG specific augmentations.

		Single-Subject ↑		Across-Subject ↑		Identity Recog. ↓	
Tasks →							
Percentage of Data →		**0.5**	**1.0**	**0.5**	**1.0**	**0.5**	**1.0**
Random Guess		16.6	16.6	16.6	16.6	12.5	12.5
Neural (Linear)	**Sup.**	29.3	32.5	18.4	18.4	100.0	100.0
Neural (MLP)	**Sup.**	–	–	18.4	18.4	100.0	100.0
SeqCLR [285]	**Self-Supervised**	39.5	42.1	21.9	28.4	93.0	96.5
Ours (Neural Only)	**Self-Supervised**	42.0	44.8	21.3	30.6	94.1	96.8
SimCLR[297]	**Self-Supervised**	54.3	57.6	46.9	50.6	69.9	80.3
Regression (Recurrent)	**Self-Supervised**	53.6	59.7	49.4	51.2	89.5	91.8
Regression (Convolution)	**Self-Supervised**	52.6	59.6	50.6	55.8	88.7	92.5
BehaveNet [272]	**Self-Supervised**	54.6	60.2	50.5	56.8	80.2	83.4
Ours	**Self-Supervised**	**57.6**	**63.1**	**54.8**	**61.5**	**13.2**	**13.6**
SimCLR [297] + MMD [329]	**Dom. Ada.**	53.6	57.8	50.1	53.1	18.4	21.2
SimCLR [297] + GRL [328]	**Dom. Ada.**	53.5	56.3	49.9	52.3	16.7	19.1
Reg. (Conv.) + MMD [329]	**Dom. Ada.**	54.5	60.7	52.6	55.4	18.2	19.5
Reg. (Conv.) + GRL [328]	**Dom. Ada.**	55.5	60.2	51.8	55.7	17.2	17.3
MM-SADA [294]	**Dom. Ada.**	53.1	56.2	49.2	52.1	13.8	15.2

Table A.1: **Action Recognition Accuracy on MC2P Dataset.** Single- and Across-Subject action recognition results on the MC2P dataset. Neural MLP results for the single-subject task are removed because single subject often do not have enough labels for every action. Smaller numbers are better for Identity Recognition. Our method performs better than previous neural decoding methods and other self-supervised learning methods in all benchmarks, while at the same time closing the domain gap between animals, as shown by the identity recognition task.

Maximum Mean Discrepancy (MMD):

We replaced adversarial loss in GRL baseline with a statistical test to minimize the distributional discrepancy from different domains [329]. Similar to previous work, we applied MMD only on the representations before the projection layer independently on both modalities [294, 330]. Similar to the GLR baseline, we first trained 10 epochs only using the contrastive loss, and trained using the combined losses $\mathscr{L}_{NCE} + \lambda_{MMD}\mathscr{L}_{MMD}$ for the remainder. We set the hyperparameters λ_{MMD} as 1 empirically. For the domain adaptation methods GRL and MMD, we reformulated the denominator of the contrastive loss function. Given a domain function dom which gives the domain of the data sample, we replaced one side of L_{NCE} in Eq. A.1 with,

$$\log \frac{\exp\left(\langle \mathbf{z}_b^i, \mathbf{z}_n^i \rangle / \tau\right)}{\sum_{k=1}^{N} \mathbf{1}_{[dom(i)=dom(k)]} \exp\left(\langle \mathbf{z}_b^i, \mathbf{z}_n^k \rangle / \tau\right)}, \tag{A.8}$$

where selective negative sampling prevents the formation of trivial negative pairs across domains, therefore making it easier to merge multiple domains. Negative pairs formed during contrastive learning try to push away inter-domain pairs, whereas domain adaptation meth-

Dataset	Tasks → % of Data →	A.S. 0.5	A.S. 1.0	I.R. 1.0
H3.6M Walking, Sitting, Posing	Random Gu.	33.0	33.0	33.0
	Supervised	46.6	48.3	100.0
	SimCLR (RGB)	33.2	33.5	99.5
	SimCLR	53.3	55.7	99.2
	Regression (Conv.)	65.2	68.8	68.4
	Ours	**72.4**	**73.6**	**42.3**
H3.6M Walking, Directions, Eating	Random Gu.	33.3	33.3	33.3
	Supervised	31.2	30.9	100.0
	SimCLR (RGB)	34.6	34.4	100.0
	SimCLR	52.3	53.2	94.8
	Regression (Conv.)	44.8	48.7	62.1
	Ours	**63.2**	**68.3**	**44.8**
ECoG Moving, Rest	Random Gu.	50.0	50.0	33.3
	Supervised	54.2	53.8	100.0
	SimCLR (ECoG)	52.3	55.1	98.0
	SimCLR	64.6	72.1	81.1
	Regression (Conv.)	64.1	71.8	74.3
	Ours	**75.8**	**81.9**	**53.0**

Table A.2: **Action Recognition Accuracy on H36M and ECoG dataset.** Across-subject (A.S.) and identity recognition (I.R.) results on H3.6M dataset[276] using RGB and 3D pose, and on ECoG Move vs Rest [274] using neural ECoG recordings and 2D pose. For Ours, we remove calcium imaging specific augmentations and only use mixing strategy. Mixing strategy closes the domain gap for the contrastive learning and strongly improves across-subject action recognition on both datasets.

ods try to merge multiple domains to close the domain gap. We found that the training of contrastive and domain adaptation losses together could be quite unstable, unless the above changes were made to the contrastive loss function.

A.5.3 Benchmarks

Since our goal is to create useful representations of neural images in an self-supervised way, we focused on single- and across-subject action recognition. Specifically, we trained our neural decoder f_n along with the others without using any action labels. Then, freezing the neural encoder parameters, we trained a linear model on the encoded features, which is an evaluation protocol widely used in the field [291, 293, 297, 331]. We used either half or all action labels. We mention the specifics of the train-test split in the Supplementary Material.

Single-Subject Action Recognition:

For each subject, we trained and tested a simple linear classifier *independently* on the learned representations to predict action labels. We assume that we are given action labels on the subject we are testing. In **Table A.1** we report aggregated results.

Across-Subject Action Recognition:

We trained linear classifiers on N-1 subjects simultaneously and tested on the left-out one. Therefore, we assume we do not have action labels for the target subject. We repeated the experiment for each individual and report the mean accuracy in **Table A.1** and **Table A.2**.

Identity Recognition:

As a sanity check, we attempted to classify subject identity among the individuals given the learned representations. We again used a linear classifier to test the domain invariance of the learned representations. In the case that the learned representations are domain (subject) invariant, we expect that the linear classifier will not be able to detect the domain of the representations, resulting in a lower identity recognition accuracy. Identity recognition results are rerported in **Table A.1** and **Table A.2**.

A.5.4 Results

Single-Subject Action Recognition on M2CP:

For the Single-Subject baseline, joint modeling of common latent space out-performed supervised models by a large margin, even when the linear classifier was trained on the action labels of the tested animal. Our mixing and neural augmentations resulted in an accuracy boost when compared with a simple contrastive learning method, SimCLR[297]. Although regression-based methods can extract behavioral information from the neural data, they do not produce discriminative features. When combined with the proposed set of augmentations, our method performs better than previous neural decoding models because it extracts richer features thanks to a better self-supervised pretraining step. Domain adaptation techniques do not result in a significant difference in the single-subject baseline; the domain gap in a single animal is smaller than between animals.

Across-Subject Action Recognition on M2CP:

We show that supervised models do not generalize across animals, because each nervous system is unique. Before using the proposed augmentations, the contrastive method SimCLR performed worse than convolutional and recurrent regression-based methods including the current state-of-art BehaveNet [272]. This was due to large domain gap between animals in

Method	Single Subj.↑	Across Subj.↑	Identity Recog. ↓
w/ Mixing Strategy	▲ + 2.7	▲ + 7.9	▼ -63.0
+ w/ Calcium Augmentation	▲ + 2.1	▲ + 2.7	■ +1.2
+ w/ N. Noise Augmentation	▲ + 1.1	▲ + 1.2	■ -0.8

Table A.3: **Ablation on Effects of Different Augmentations.** Showing the effect of different augmentations on single-subject, across-subject and identity recognition benchmarks on MC2P Dataset, when compared to simple contrastive model.

the latent embeddings **(Fig. A.3B)**. Although the domain adaptation methods MMD (Maximum Mean Discrepancy) and GRL (Gradient Reversal Layer) close the domain gap when used with contrastive learning, they do not position semantically similar points near one another **(Fig. A.3C)**. As a result, domain adaptation-based methods do not result in significant improvements in the across-subject action recognition task. Although regression-based methods suffer less from the domain gap problem, they do not produce representations that are as discriminative as contrastive learning-based methods. Our proposed set of augmentations and strategies close the domain gap, while improving the action recognition baseline for self-supervised methods, for both single-subject and across-subject tasks **(Fig. A.3D)**.

Action Recognition on ECoG Motion vs Rest:

As shown at the bottom of **Table A.2**, our approach significantly lowers the identity information in ECoG embeddings, while significantly increasing across-subject action recognition accuracy compared to the regression and multi-modal SimCLR baselines. Low supervised accuracy confirms a strong domain gap across individuals. Note that uni-modal contrastive modeling of ECoG recordings (SimCLR (ECoG)) does not yield strong across-subject action classification accuracy because uni-modal modeling cannot deal with the large domain gap in the learned representations.

Human Action Recognition on H3.6M:

We observe in **Table A.2** that, similar to the previous datasets, the low performance of the supervised baseline and the uni-modal modeling of RGB images (SimCLR (RGB)) are due to the domain-gap in the across-subject benchmark. This observation is confirmed by the high identity recognition of these models. Our mixing strategy strongly improves compared to the regression and multi-modal contrastive (SimCLR) baselines. Similar to the previous datasets, uni-modal contrastive training cannot generalize across subjects, due to the large domain gap.

Method	Single Subj.↑	Across Subj.↑	Identity Recog. ↓
CaImAn	56.8	56.1	17.0
Ours	**63.1**	**61.5**	**13.6**

Table A.4: **Ablation on Neural Preprocessing.** Comparing standart neural processing library *CaImAn* with our augmentations on MC2P Dataset.

Method	Single Subj.↑	Across Subj.↑	Identity Recog. ↓
No Mix.	59.1	51.2	81.7
N. Mix.	60.4	58.9	36.8
B. Mix.	61.0	60.2	25.2
N. + B. Mix.	**63.1**	**61.5**	**13.6**

Table A.5: **Ablation on Mixing of Different Modalities on MC2P Dataset.** Showing the effect of mixing different modalities on MC2P dataset.

A.5.5 Ablation Study

We compare the individual contributions of different augmentations proposed in our method. We report these results in **Table A.3**. We observe that all augmentations contribute to single- and across-subject benchmarks. Our mixing augmentation strongly affects the across-subject benchmark, while at the same time greatly decreasing the domain gap, as quantified by the identity recognition result. Other augmentations have minimal effects on the domain gap, as they only slightly affect the identity recognition benchmark.

We compare our neural augmentations to standart neural preprocessing approaches commonly used in for neuroscience. To compare, we use state-of-art neural preprocessing library *CaImAn* [332]. *CaImAn* requires the tuning of 25 parameters for spike inference. Running a single-set of parameters took 5h of processing. As shown in **Tab. A.4**, this algorithm produced worse results, likely due to errors in ROI detection and spike inference. Because our method can be run on the raw data, without requiring ROI detection and spike inference, it removes the burden of an extensive hyperparameter search and unnecessarily long computational times. Thus, we believe that our augmentations and model is more general and useful for the community. Lastly, we performed ablation experiment on mixing individual modalities and report the results in **Tab. A.5**. Mixing both modalities results in the best scores. However, when mixed alone, the behavioral modality performs superior as it more effectively hides subject identity information, since it is mixed in pose level instead of window level.

A.6 Conclusion

We have introduced an self-supervised neural action representation framework for neural imaging and behavioral videography data. We extended previous methods by incorporating a

new mixing based domain adaptation technique which we have shown to be useful on three very different multimodal datasets, together with a set of domain-specific neural augmentations. Two of these datasets are publicly available. We created the third dataset, which we call MC2P, by recording video and neural data for *Drosophila* and will release it publicly to speed-up the development of self-supervised methods in neuroscience. We hope our work will help the development of effective brain machine interface and neural decoding algorithms. In future work, we plan to disentangle remaining long-term non-behavioral information that has a global effect on neural data, such as hunger or thirst, and test our method on different neural recording modalities. As a potential negative impact, we assume that once neural data is taken without consent, our method can be used to extract private information.

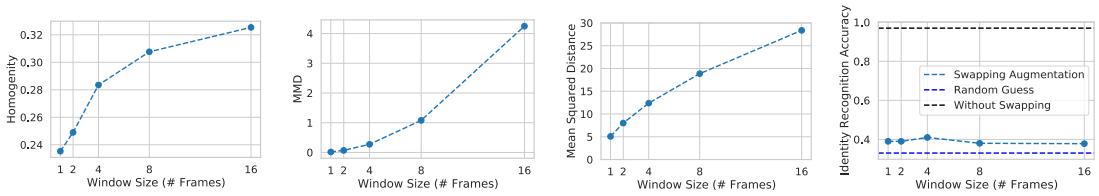

Figure SA.1: **Changing window size for the mixing of behavioral modality on the MC2P dataset.** Statistics of the behavioral modality as a function of changing the window size. Decreasing the window size increases clustering homogeneity and Mean Maximum Discrepancy (MMD) when applied to the raw data, therefore suggesting higher quality mixing in individual poses instead of sequences of poses. mixing augmentation with a smaller window size lowers the degree of perturbation, quantified by Mean Squared Distance. However, identity recognition accuracy does not change considerably when mixing is done with different window sizes.

A.7 Supplementary Materials

A.7.1 Human Actions

We apply multi-modal contrastive learning on windows of time series and RGB videos. We make the analogy that, similar to the neural data, RGB videos from different view angles show a domain gap although they are tied to the same 3D pose. Therefore, to test our method, we select three individuals with different camera angles where all actors perform the same three actions. We test domain adaptation using the Across-Subject benchmark, where we train our linear action classifier on labels of one individual and test it on the others. We repeat the same experiment three times and report the mean results. We show the results of Across-Subject and Identity Recognition in **Table A.2**.

For preprocessing, we remove global translation and rotation from 3D poses by subtracting the root joint and then rotating the skeletons to point in the same direction. We use resnet18 for the RGB encoder and a 4 layer convolutional network for the 3D pose encoder. We use S1, S5 and S7 and all their behaviors for training, except for the three behaviors which we used for testing. For each number, we report three-fold cross-validation results.

A.7.2 Dataset Details

Dataset Collection:

Here we provide a more detailed technical explanation of the experimental dataset. Transgenic female *Drosophila melanogaster* flies aged 2-4 days post-eclosion were selected for experiments. They were raised on a 12h:12h day, night light cycle and recorded in either the morning or late afternoon Zeitgeber time. Flies expressed both GCaMP6s and tdTomato in all brain neurons as delineated by otd-Gal4 expression,

(; $\frac{Otd-nls:FLPo(attP40)}{P20XUAS-IVS-GCaMP6sattP40}$; $\frac{R57C10-GAL4,tub>GAL80>}{Pw[+mC]=UAS-tdTom.S3}$.) The fluorescence of GCaMP6s proteins within the neuron increases when it binds to calcium. There is an increase in intracellular calcium when neurons become active and fire action potentials. Due to the relatively slow release (as opposed to binding) of calcium by GCaMP6s molecules, the signal decays exponentially. We also expressed the red fluorescent protein, tdTomato, in the same neurons as an anatomical fiduciary to be used for neural data registration. This compensates for image deformations and translations during animal movements. We recorded neural data using a two-photon microscope (ThorLabs, Germany; Bergamo2) by scanning the cervical connective. This neural tissue serves as a conduit between the brain and ventral nerve cord (VNC) [171]. The brain-only GCaMP6s expression pattern in combination with restrictions of recording to the cervical connective allowed us to record a large population of descending neuron axons while also being certain that none of the axons arose from ascending neurons in the VNC. Because descending neurons are expected to drive ongoing actions [74], this imaging approach has the added benefit of ensuring that the imaged cells should, in principle, relate to paired behavioral data.

Behavioral Pre-processing:

For the MC2P dataset, we register each 3D pose into a canonical coordinate system using Procrustes's analysis and normalize limb-lengths across subjects. Since each animal is tethered under the two-photon microscope, they do not change their rotation during the experiment. For each of the six legs, we set the body-coxa locations so that relative sizes of the animals do not reflected in the data. We then normalize the data using calculated mean and variance across animals.

Neural Pre-Processing:

For neural preprocessing, we developed our own, light-weight approach because conventional NMF methods (i) were designed for easy-to-identify, rodent neural cell bodies but do not generalize to tracking our axons, (ii) model action potentials, but do not take into account the distinct dynamical properties of graded potential neurons, and (iii) perform an optimization that, in our hands, is overly sensitive to initial conditions and computationally very expensive to run on our full dataset. For our approach, data were synchronized using a custom Python package [333]. We then estimated the motion of the neurons using images acquired on the red (tdTomato) PMT channel. The first image of the first trial was selected as a reference frame to which all other frames were registered. For image registration, we estimated the vector field describing the motion between two frames. To do this, we numerically solved the optimization problem in **Eq. A.9**, where w is the motion field, $\mathscr{I}_t$ is the image being transformed, $\mathscr{I}_r$ is the reference image, and Ω is the set of all pixel coordinates [171, 334].

$$\hat{w} = argmin_w \sum_{x\in\Omega} ||\mathscr{I}_t(x + w(x)) - \mathscr{I}_r(x)||_2^2 - \lambda \sum_{x\in\Omega} ||\nabla w(x)||_2^2 \tag{A.9}$$

A smoothness promoting parameter λ was empirically set to 800. We then applied $\hat{w}$ to the green PMT channel (GCaMP6s). To denoise the motion corrected green signal, we trained a DeepInterpolation network [223] for nine epochs for each animal and applied it to the rest of the frames. We only used the first 100 frames of each trial and used the first and last trials as validation data. The batch size was set to 20 and we used 30 frames before and after the current frame as input. In order to have a direct correlation between pixel intensity and neuronal activity we applied the following transformation to all neural images $\frac{F-F_0}{F_0} \times 100$, where F_0 is the baseline fluorescence in the absence of neural activity. To estimate F_0, we used the pixel-wise minimum of a moving average of 15 frames.

Neural Fluorescence Signal Decay:

The formal relationship between the neural image $\mathbf{n}_t$ and neural activity (underlying neural firings) $\mathbf{s}_t$ can be modeled as a first-order autoregressive process

$$\mathbf{n}_t = \gamma \mathbf{n}_{t-1} + \alpha \mathbf{s}_t,$$

where $\mathbf{s}_t$ is a binary variable indicating an event at time t (e.g. the neuron firing an action potential). The amplitudes γ and α determine the rate at which the signal decays and the initial response to an event, respectively. In general, $0 < \gamma < 1$, therefore resulting in an exponential decay of information pertaining to $\mathbf{s}_t$ to be inside of $\mathbf{n}_t$. A single neural image $\mathbf{n}_t$ includes decaying information from previous neural activity, and hence carries information from previous behaviors. For more detailed information on calcium dynamics, see [335, 336]. Assuming no neural firings, $\mathbf{s}_t = 0$, $\mathbf{n}_t$ is given by $\mathbf{n}_t = \gamma^t \mathbf{n}_0$. Therefore, we define the calcium kernel $\mathcal{K}$ as $\mathcal{K}_t = \gamma^t$.

Dataset Analysis:

We show the distribution of annotations across 7 animals and action duration distribution in **Supplementary Fig. SA.2**. Unlike scripted actions in human datasets, the animal behavior is spontaneous, therefore does not follow a uniform distribution. The average duration of behaviors can also change across behaviors. Walking is the most common behavior and lasts longer than other behaviors. We visualize the correlation between the neural and behavioral energy in **Supplementary Fig. SA.3**. We quantify the energy as the Euclidean distance between consecutive, vectorized 3D poses. Similarly, for the neural energy, we calculate the Euclidean distance between consecutive images. To be able to compare corresponding energies, we first synchronize neural and behavioral modalities. We then smooth the corresponding time series using Gaussian convolution with kernel size of 11 frames. We observe that there is a strong correlation between the modalities, suggesting large mutual information.

A.7.3 Method Details

Augmentations:

Aside from the augmentations mentioned before, for the neural image transformation family $\mathcal{T}_n$, we used a sequential application of Poisson noise and Gaussian blur and color jittering. In contrast with recent work on contrastive visual representation learning, we only applied brightness and contrast adjustments in color jittering because neural images have a single channel that measures calcium indicator fluorescence intensity. We did not apply any cropping augmentation, such as cutout, because action representation is often highly localized and non-redundant (e.g., grooming is associated with the activity of a small set of neurons and thus with only a small number of pixels). We applied the same augmentations to each frame in single sample of neural data.

For the behavior transformation family $\mathcal{T}_b$, we used a sequential application of scaling, shear, and random temporal and spatial dropping. We did not apply rotation and translation augmentations because the animals were tethered (i.e., restrained from moving freely), and their direction and absolute location were fixed throughout the experiment. We did not use time warping because neural and behavioral information are temporally linked (e.g., fast walking has different neural representations than slow walking).

Mixing Parameters:

We analyze the effects of mixing individual poses, instead of whole motion sequences, through our mixing augmentation in **Fig. SA.1**. We compare the distribution similarity across individuals when tested on single poses and windows of poses. We observe that the distribution similarity across individuals in behavioral modality is much larger in pose level when compared to the whole motion sequence, therefore making it easier to mix behavioral data in pose level. We quantify the distribution similarity using MMD (Mean Maximum Discrepancy) and Homogeneity metrics. Similarly, mixing individual poses decreases the overall change in the motion sequence, as quantified by the Mean Squared Distance. Yet, the degree to which identity information is hid does not strongly correlate with the window size of mixing. Therefore, overall, suggesting mixing in pose level is better than mixing whole motion sequences.

Implementation Details:

For all methods, we initialized the weights of the networks randomly unless otherwise specified. To keep the experiments consistent, we always paired 32 frames of neural data with 8 frames of behavioral data. For the neural data, we used a larger time window because the timescale during which dynamic changes occur are smaller. For the paired modalities, we considered data synchronized if their center frames had the same timestamp. We trained contrastive methods for 200 epochs and set the temperature value τ to 0.1. We set the output dimension of $\mathbf{z}_b$ and $\mathbf{z}_n$ to 128. We used a cosine training schedule with three epochs of warm-up. For

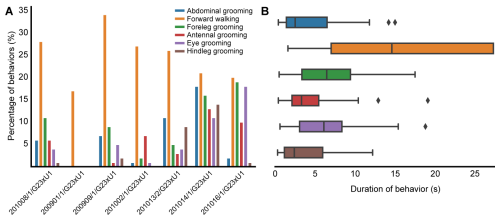

Figure SA.2: **Motion Capture and two-photon dataset statistics.** Visualizing **(A)** the number of annotations per animal and **(B)** the distribution of the durations of each behavior across animals. Unlike scripted human behaviors, animal behaviors occur spontaneously. The total number of behaviors and their durations do not follow a uniform distribution, therefore making it harder to model.

non-contrastive methods, we trained for 200 epochs with a learning rate of $1e-4$, and a weight decay of $1e-5$, using the Adam optimizer [204]. We ran all experiments using an Intel Core i9-7900X CPU, 32 GB of DDR4 RAM, and a GeForce GTX 1080. Training for a single SimCLR network for 200 epochs took 12 hours. To create train and test splits, we removed two trials from each animal and used them only for testing. We used the architecture shown in **Supplementary Table 1** for the neural image and behavioral pose encoder. Each layer except the final fully-connected layer was followed by Batch Normalization and a ReLU activation function [203]. For the self-attention mechanism in the behavioral encoder **(Supplementary Table 1)**, we implement Bahdanau attention [337]. Given the set of intermediate behavioral representations $S \in \mathbb{R}^{T \times D}$, we first calculated,

$$\mathbf{r} = W_2 \tanh\left(W_1 S^\top\right), \quad \mathbf{a}_i = -\log\left(\frac{\exp\left(\mathbf{r}_i\right)}{\sum_j \exp\left(\mathbf{r}_j\right)}\right)$$

where W_1 and W_2 are a set of matrices of shape $\mathbb{R}^{12 \times D}$ and $\mathbb{R}^{1 \times 12}$ respectively. $\mathbf{a}_i$ is the assigned score i-th pose in the sequence of motion. Then the final representation is given by $\sum_i^T \mathbf{a}_i S_i$.

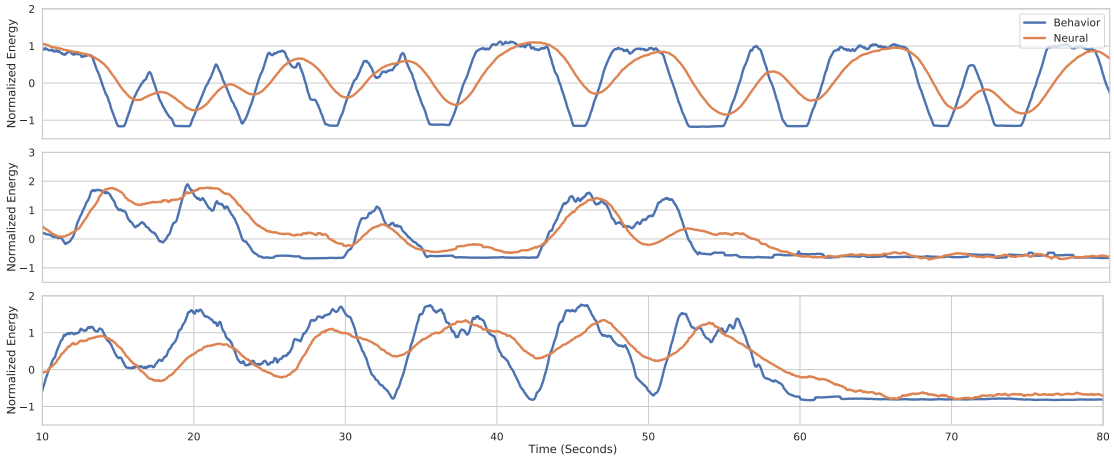

Figure SA.3: **Visualizing the temporal correlation between behavioral and neural energies on multiple animals.** The behavioral and neural energies are calculated as the normalized distances between consecutive frames. The multi-modal energies show a similar temporal pattern. The slower neural energy decay is due to the calcium dynamics.

A.7.4 Supplementary Videos

Motion Capture and Two-Photon (MC2P) Dataset:

The following videos are sample behavioral-neural recordings from two different flies. The videos show **(left)** raw behavioral RGB video together with **(right)** registered and denoised neural images at their original resolutions. The behavioral video is resampled and synchronized with the neural data. The colorbar indicates normalized relative intensity values. Calculation of $\Delta F/F$ is previously explained under Dataset Collection section.

Video 1: https://drive.google.com/file/d/1-xjiwn7qgiou3_nf0nlyu7549KKlSfUx

Video 2: https://drive.google.com/file/d/1DUOzSbbE8uPdPNvXChJWYb9bN9AyUfuO

Action Label Annotations:

Sample behavioral recordings from multiple animals using a single camera. Shown are eight different action labels: *forward walking, pushing, hindleg grooming, abdominal grooming, foreleg grooming, antennal grooming, eye grooming* and *resting*. Videos are temporally downsampled. Animals and labels are randomly sampled.

Video 3: https://drive.google.com/file/d/1cnwRRyDZ4crrVVxRBbx32Za-vlxSP7sy

Animal Motion Capture:

Sample behavioral recordings with 2D poses from six different camera views. Each color denotes a different limb. The videos are temporally down-sampled for easier view.

Video 4: https://drive.google.com/file/d/1uYcL7_Zl-N0mlG1VTrg67s2Cy71wml5S

Video 5: https://drive.google.com/file/d/1eMcP-Ec1c4yBQpC4CNv45py7gObmuUeA

First part of the Neural Encoder f_n

Layer	# filters	K	S	Output
input	1	-	-	$T \times 128 \times 128$
conv1	2	(3,3)	(1,1)	$T \times 128 \times 128$
mp2	-	(2,2)	(2,2)	$T \times 64 \times 64$
conv3	4	(3,3)	(1,1)	$T \times 64 \times 64$
mp4	-	(2,2)	(2,2)	$T \times 32 \times 32$
conv5	8	(3,3)	(1,1)	$T \times 32 \times 32$
mp6	-	(2,2)	(2,2)	$T \times 16 \times 16$
conv7	16	(3,3)	(1,1)	$T \times 16 \times 16$
mp8	-	(2,2)	(2,2)	$T \times 8 \times 8$
conv9	32	(3,3)	(1,1)	$T \times 8 \times 8$
mp10	-	(2,2)	(2,2)	$T \times 4 \times 4$
conv11	64	(3,3)	(1,1)	$T \times 4 \times 4$
mp12	-	(2,2)	(2,2)	$T \times 2 \times 2$
fc13	128	(1,1)	(1,1)	$T \times 1 \times 1$
fc14	128	(1,1)	(1,1)	$T \times 1 \times 1$

Second part of the Neural Encoder f_n

Layer	# filters	K	S	Output
input	60	-	-	$T \times 128$
conv1	64	(3)	(1)	$T \times 128$
conv2	80	(3)	(1)	$T \times 128$
mp2	-	(2)	(2)	$T/2 \times 128$
conv2	96	(3)	(1)	$T/2 \times 128$
conv2	112	(3)	(1)	$T/2 \times 128$
conv2	128	(3)	(1)	$T/2 \times 128$
attention6	-	(1)	(1)	1×128
fc7	128	(1)	(1)	1×128

Behavioral Encoder f_b

Layer	# filters	K	S	Output
input	60	-	-	$T \times 60$
conv1	64	(3)	(1)	$T \times 64$
conv2	80	(3)	(1)	$T \times 80$
mp2	-	(2)	(2)	$T/2 \times 80$
conv2	96	(3)	(1)	$T/2 \times 96$
conv2	112	(3)	(1)	$T/2 \times 112$
conv2	128	(3)	(1)	$T/2 \times 128$
attention6	-	(1)	(1)	1×128
fc7	128	(1)	(1)	1×128

Table SA.1: **Architecture details.** Shown are half of the neural encoder f_n and behavior encoder f_b functions. How these encoders are used is shown in **Fig. 3**. Neural encoder f_n is followed by 1D convolutions similar to the behavioral encoder f_b, by replacing the number of filters. Both encoders produce 128 dimensional output, while first half of the neural encoder do not downsample on the temporal axis. *mp* denotes a max-pooling layer. Batch Normalization and ReLU activation are added after every convolutional layer.

Bibliography

1. Kiehn, O. & Churchland, M. M. Editorial overview: Motor circuits and action. *Current Opinion in Neurobiology* **33,** v–vi (Aug. 2015).
2. Kandel, E. *et al. Principles of Neural Science, Fifth Edition* (McGraw-Hill Education, 2013).
3. Marshall, N. J. *et al.* Flexible neural control of motor units (May 2021).
4. Marder, E. & Bucher, D. Central pattern generators and the control of rhythmic movements. *Current Biology* **11,** R986–R996 (2001).
5. Rothwell, J., Day, B., Berardelli, A. & Marsden, C. Effects of motor cortex stimulation on spinal interneurones in intact man. *Experimental Brain Research* **54** (Mar. 1984).
6. Witvliet, D. *et al.* Connectomes across development reveal principles of brain maturation. *Nature* **596,** 257–261 (Aug. 2021).
7. Kiral, F. R. *et al.* Brain connectivity inversely scales with developmental temperature in Drosophila. *Cell Reports* **37,** 110145 (2021).
8. Werkhoven, Z. *et al.* The structure of behavioral variation within a genotype. *eLife* **10** (Oct. 2021).
9. Muller, H. J. GENETIC VARIABILITY, TWIN HYBRIDS AND CONSTANT HYBRIDS, IN A CASE OF BALANCED LETHAL FACTORS. *Genetics* **3,** 422–499 (Sept. 1918).
10. Greenspan, R. J. *Fly Pushing: The Theory and Practice of Drosophila Genetics.– 2nd ed.* (Cold Spring Harbor Laboratory Press, 2004).
11. Fischer, J. A., Giniger, E., Maniatis, T. & Ptashne, M. GAL4 activates transcription in *Drosophila*. *Nature* **332,** 853–856 (1988).
12. Brand, A. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118,** 401–415 (1993).
13. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22,** 451–461 (1999).
14. Luan, H., Peabody, N. C., Vinson, C. R. & White, B. H. Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. *Neuron* **52,** 425–436 (2006).

15. Pollack, H. MICRURGICAL STUDIES IN CELL PHYSIOLOGY. *Journal of General Physiology* **11,** 539–545 (May 1928).
16. Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein. *Nature Biotechnology* **19,** 137–141 (Feb. 2001).
17. Blinks, J. R., Wier, W., Hess, P. & Prendergast, F. G. Measurement of Ca2+ concentrations in living cells. *Progress in Biophysics and Molecular Biology* **40,** 1–114 (1982).
18. Miyawaki, A. *et al.* Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature* **388,** 882–887 (Aug. 1997).
19. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499,** 295–300 (2013).
20. Zhang, Y. *et al. jGCaMP8 Fast Genetically Encoded Calcium Indicators* 2020.
21. Catterall, W. A. Voltage-Gated Calcium Channels. *Cold Spring Harbor Perspectives in Biology* **3,** a003947–a003947 (July 2011).
22. Geppert, M. *et al.* Synaptotagmin I: A major Ca2+ sensor for transmitter release at a central synapse. *Cell* **79,** 717–727 (Nov. 1994).
23. Higley, M. J. & Sabatini, B. L. Calcium Signaling in Dendritic Spines. *Cold Spring Harbor Perspectives in Biology* **4,** a005686–a005686 (Feb. 2012).
24. Grienberger, C. & Konnerth, A. Imaging Calcium in Neurons. *Neuron* **73,** 862–885 (2012).
25. Pnevmatikakis, E. A. *et al.* Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron* **89,** 285–299 (2016).
26. Farouj, Y., Karahanoğlu, F. I. & Van De Ville, D. Deconvolution of Sustained Neural Activity From Large-Scale Calcium Imaging Data. *IEEE Transactions on Medical Imaging* **39,** 1094–1103 (2020).
27. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLOS Computational Biology* **13** (ed Vogelstein, J.) e1005423 (Mar. 2017).
28. Miller, C. T. *et al.* Natural behavior is the language of the brain. *Current Biology* **32,** R482–R493 (May 2022).
29. Seelig, J. D. *et al.* Two-photon calcium imaging from head-fixed Drosophila during optomotor walking behavior. *Nature Methods* **7,** 535–540 (June 2010).
30. Ghosh, K. K. *et al.* Miniaturized integration of a fluorescence microscope. *Nature Methods* **8,** 871–878 (Sept. 2011).
31. Cui, G. *et al.* Deep brain optical measurements of cell type–specific neural activity in behaving mice. *Nature Protocols* **9,** 1213–1228 (May 2014).
32. Lobato-Rios, V. *et al.* NeuroMechFly, a neuromechanical model of adult Drosophila melanogaster. *Nature Methods,* 1–8 (2022).
33. Carey, J. R. *et al.* Age-specific and lifetime behavior patterns in Drosophila melanogaster and the Mediterranean fruit fly, Ceratitis capitata. *Experimental Gerontology* **41,** 93–97 (2006).

34. Larson, A. M. Multiphoton microscopy. *Nature Photonics* **5,** 1–1 (Dec. 2010).
35. Aragon, M. J. *et al.* Multiphoton imaging of neural structure and activity in Drosophila through the intact cuticle. *eLife* **11** (Jan. 2022).
36. Moore, R. J. *et al.* FicTrac: A visual method for tracking spherical motion and generating fictive animal paths. *Journal of Neuroscience Methods* **225,** 106–119 (2014).
37. Günel, S. *et al.* DeepFly3D, a deep learning-based approach for 3D limb and appendage tracking in tethered, adult *Drosophila*. *eLife* **8** (eds O'Leary, T., Calabrese, R. L. & Shaevitz, J. W.) e48571 (2019).
38. Gosztolai, A. *et al.* LiftPose3D, a deep learning-based approach for transforming two-dimensional to three-dimensional poses in laboratory animals. *Nature Methods* **18,** 975–981 (Aug. 2021).
39. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* **21,** 1281–1289 (Aug. 2018).
40. Pereira, T. D. *et al.* SLEAP: A deep learning system for multi-animal pose tracking. *Nature Methods* **19,** 486–495 (Apr. 2022).
41. Karashchuk, P. *et al.* Anipose: A toolkit for robust markerless 3D pose estimation. *Cell Reports* **36,** 109730 (2021).
42. Whiteway, M. R. *et al.* Semi-supervised sequence modeling for improved behavioral segmentation. *bioRxiv* (2021).
43. Hutter, H. Formation of longitudinal axon pathways in Caenorhabditis elegans. *Seminars in Cell & Developmental Biology* **85.** Axon guidance: Signaling pathways old and new across a variety of developmental contexts, 60–70 (2019).
44. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. en. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314,** 1–340 (Nov. 1986).
45. Harris, R. M., Pfeiffer, B. D., Rubin, G. M. & Truman, J. W. Neuron hemilineages provide the functional ground plan for the *Drosophila* ventral nervous system. *eLife* **4,** e04493 (2015).
46. Egeth, M. Behavioral Responses to Light by Headless Anesthetized *Drosophila Melanogaster*. *Perception* **40,** 247–248 (Jan. 2011).
47. Booker, R. & Quinn, W. G. Conditioning of leg position in normal and mutant Drosophila. *Proceedings of the National Academy of Sciences* **78,** 3940–3944 (June 1981).
48. WHELAN, P. Control of locomotion in the decerebrate cat. *Progress in Neurobiology* **49,** 481–515 (Aug. 1996).
49. Cole, J. *Losing Touch* (Oxford University Press, June 2016).
50. Miall, R. C., Afanasyeva, D., Cole, J. D. & Mason, P. Perception of body shape and size without touch or proprioception: evidence from individuals with congenital and acquired neuropathy. *Experimental Brain Research* **239,** 1203–1221 (Feb. 2021).

51. Tsubouchi, A. *et al.* Topological and modality-specific representation of somatosensory information in the fly brain. *Science* **358,** 615–623 (2017).

52. Tuthill, J. C. & Wilson, R. I. Parallel transformation of tactile signals in central circuits of *Drosophila*. *Cell* **164,** 1046–1059 (2016).

53. Bidaye, S. S., Machacek, C., Wu, Y. & Dickson, B. J. Neuronal control of *Drosophila* walking direction. *Science* **344,** 97–101 (2014).

54. Mann, K., Gordon, M. & Scott, K. A pair of interneurons influences the choice between feeding and locomotion in *Drosophila*. *Neuron* **79,** 754–765 (2013).

55. Fujiwara, T., Brotas, M. & Chiappe, M. E. Walking strides direct rapid and flexible recruitment of visual circuits for course control in Drosophila. *Neuron* (May 2022).

56. Crapse, T. B. & Sommer, M. A. Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience* **9,** 587–600 (2008).

57. Poulet, J. F. A. & Hedwig, B. The Cellular Basis of a Corollary Discharge. *Science* **311,** 518–522 (Jan. 2006).

58. Poulet, J. & Hedwig, B. A Corollary Discharge Mechanism Modulates Central Auditory Processing in Singing Crickets. *Journal of Neurophysiology* **89,** 1528–1540 (Mar. 2003).

59. Namiki, S., Dickinson, M. H., Wong, A. M., Korff, W. & Card, G. M. The functional organization of descending sensory-motor pathways in Drosophila. *Elife* **7,** e34272 (2018).

60. Kupfermann, I. & Weiss, K. R. The command neuron concept. *Behavioral and Brain Sciences* **1,** 3–10 (Mar. 1978).

61. Yoshihara, M. & Yoshihara, M. 'Necessary and sufficient' in biology is not necessarily necessary – confusions and erroneous conclusions resulting from misapplied logic in the field of biology, especially neuroscience. *Journal of Neurogenetics* **32,** 53–64 (Apr. 2018).

62. Kien, J. & Altman, J. S. Descending interneurones from the brain and suboesophageal ganglia and their role in the control of locust behaviour. *Journal of Insect Physiology* **30,** 59–72 (Jan. 1984).

63. Kien, J. Neuronal activity during spontaneous walking—I. Starting and stopping. *Comparative Biochemistry and Physiology Part A: Physiology* **95,** 607–621 (1990).

64. Staudacher, E. Distribution and morphology of descending brain neurons in the cricket Gryllus bimaculatus. *Cell and Tissue Research* **294,** 187–202 (Sept. 1998).

65. Heinrich, R. Impact of descending brain neurons on the control of stridulation, walking, and flight in orthoptera. *Microscopy Research and Technique* **56,** 292–301 (Feb. 2002).

66. Zorović, M. & Hedwig, B. Processing of species-specific auditory patterns in the cricket brain by ascending, local, and descending neurons during standing and walking. *Journal of Neurophysiology* **105,** 2181–2194 (May 2011).

67. Zorović, M. & Hedwig, B. Descending brain neurons in the cricket Gryllus bimaculatus (de Geer): auditory responses and impact on walking. *Journal of Comparative Physiology A* **199,** 25–34 (Oct. 2012).

68. Namiki, S., Wada, S. & Kanzaki, R. Descending neurons from the lateral accessory lobe and posterior slope in the brain of the silkmoth Bombyx mori. *Scientific Reports* **8** (June 2018).

69. Kanzaki, R., Ikeda, A. & Shibuya, T. Morphological and physiological properties of pheromone-triggered flipflopping descending interneurons of the male silkworm moth, Bombyx mori. *Journal of Comparative Physiology A* **175** (July 1994).

70. Ache, J. M., Haupt, S. S. & Durr, V. A Direct Descending Pathway Informing Locomotor Networks about Tactile Sensor Movement. *Journal of Neuroscience* **35,** 4081–4091 (Mar. 2015).

71. Severina, I. Y., Isavnina, I. L. & Knyazev, A. N. Topographic anatomy of ascending and descending neurons of the supraesophageal, meso- and metathoracic ganglia in paleo- and neopterous insects. *Journal of Evolutionary Biochemistry and Physiology* **52,** 397–406 (Sept. 2016).

72. Gal, R. & Libersat, F. New vistas on the initiation and maintenance of insect motor behaviors revealed by specific lesions of the head ganglia. *Journal of Comparative Physiology A* **192,** 1003–1020 (May 2006).

73. Okada, R., Sakura, M. & Mizunami, M. Distribution of dendrites of descending neurons and its implications for the basic organization of the cockroach brain. *The Journal of Comparative Neurology* **458,** 158–174 (Feb. 2003).

74. Cande, J. *et al.* Optogenetic dissection of descending behavioral control in *Drosophila*. *Elife* **7,** e34275 (2018).

75. Bidaye, S. S. *et al.* Two brain pathways initiate distinct forward walking programs in *Drosophila*. *Neuron* **108,** 469–485.e8 (2020).

76. Ache, J. M. *et al.* Neural Basis for Looming Size and Velocity Encoding in the Drosophila Giant Fiber Escape Pathway. *Current Biology* **29,** 1073–1081.e4 (Mar. 2019).

77. Ache, J. M., Namiki, S., Lee, A., Branson, K. & Card, G. M. State-dependent decoupling of sensory and motor circuits underlies behavioral flexibility in Drosophila. *Nature Neuroscience* **22,** 1132–1139 (June 2019).

78. Guo, L., Zhang, N. & Simpson, J. H. Descending neurons coordinate anterior grooming behavior in Drosophila. *Current Biology* **32,** 823–833.e4 (Feb. 2022).

79. Hsu, C. T. & Bhandawat, V. Organization of descending neurons in Drosophila melanogaster. *Scientific Reports* **6** (Feb. 2016).

80. Strausfeld, N., Bassemir, U., Singh, R. & Bacon, J. Organizational principles of outputs from Dipteran brains. *Journal of Insect Physiology* **30,** 73–93 (Jan. 1984).

81. Coggshall, J. C., Boschek, C. B. & Buchner, S. M. Preliminary Investigations on a Pair of Giant Fibers in the Central Nervous System of Dipteran Flies. *Zeitschrift für Naturforschung C* **28,** 783–784b (1973).

82. Tastekin, I. *et al.* Role of the subesophageal zone in sensorimotor control of orientation in *Drosophila* larva. *Current Biology* **25,** 1448–1460 (2015).

83. Sterne, G. R., Otsuna, H., Dickson, B. J. & Scott, K. Classification and genetic targeting of cell types in the primary taste and premotor center of the adult *Drosophila* brain. *Elife* **10,** e71679 (2021).

84. Scheffer, L. K. *et al.* A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9** (eds Marder, E., Eisen, M. B., Pipkin, J. & Doe, C. Q.) e57443 (Sept. 2020).

85. Phelps, J. S. *et al.* Reconstruction of motor control circuits in adult *Drosophila* using automated transmission electron microscopy. *Cell* **184,** 759–774 (2021).

86. Du Beau, A. *et al.* Neurotransmitter phenotypes of descending systems in the rat lumbar spinal cord. *Neuroscience* **227,** 67–79 (2012).

87. Schnell, B., Ros, I. G. & Dickinson, M. H. A Descending Neuron Correlated with the Rapid Steering Maneuvers of Flying Drosophila. *Current Biology* **27,** 1200–1205 (Apr. 2017).

88. Von Philipsborn, A. C. *et al.* Neuronal Control of Drosophila Courtship Song. *Neuron* **69,** 509–522 (Feb. 2011).

89. Rayshubskiy, A. *et al.* Neural circuit mechanisms for steering control in walking *Drosophila. bioRxiv,* 2020.04.04.024703 (2020).

90. Von Reyn, C. R. *et al.* A spike-timing mechanism for action selection. *Nature Neuroscience* **17,** 962–970 (June 2014).

91. Rowell, C. H. F. The orthopteran descending movement detector (DMD) neurones: a characterisation and review. *Zeitschrift für Vergleichende Physiologie* **73,** 167–194 (1971).

92. Burrows, M. & Rowell, C. H. F. Connections between descending visual interneurons and metathoracic motoneurons in the locust. *Journal of Comparative Physiology* **85,** 221–234 (Sept. 1973).

93. Hampel, S., Franconville, R., Simpson, J. H. & Seeds, A. M. A neural command circuit for grooming movement control. *eLife* **4** (Sept. 2015).

94. Berendes, V. & Dürr, V. Active tactile exploration and tactually induced turning in tethered walking stick insects. *Journal of Experimental Biology* **225.** jeb243190 (Mar. 2022).

95. Namiki, S. *et al.* A population of descending neurons that regulates the flight motor of Drosophila. *Current Biology* **32,** 1189–1196.e6 (Mar. 2022).

96. Böhm, H. & Schildberger, K. BRAIN NEURONES INVOLVED IN THE CONTROL OF WALKING IN THE CRICKET GRYLLUS BIMACULATUS. *Journal of Experimental Biology* **166,** 113–130 (May 1992).

97. Staudacher, E. & Schildberger, K. Gating of sensory responses of descending brain neurones during walking in crickets. *Journal of Experimental Biology* **201,** 559–572 (Feb. 1998).

98. McKellar, C. E. *et al.* Threshold-Based Ordering of Sequential Actions during Drosophila Courtship. *Current Biology* **29,** 426–434.e6 (Feb. 2019).

99. Lemon, R. N. Descending Pathways in Motor Control. *Annual Review of Neuroscience* **31,** 195–218 (July 2008).

100. Kuypers, H. in *Progress in Brain Research* 178–202 (Elsevier, 1964).

101. Capelli, P., Pivetta, C., Soledad Esposito, M. & Arber, S. Locomotor speed control circuits in the caudal brainstem. *Nature* **551,** 373–377 (2017).

102. Cregg, J. M. *et al.* Brainstem neurons that command mammalian locomotor asymmetries. *Nature Neuroscience* **23,** 730–740 (May 2020).

103. Esposito, M. S., Capelli, P. & Arber, S. Brainstem nucleus MdV mediates skilled forelimb motor tasks. *Nature* **508,** 351–356 (Feb. 2014).

104. Hou, X. H. *et al.* Central Control Circuit for Context-Dependent Micturition. *Cell* **167,** 73–86.e12 (Sept. 2016).

105. Bouvier, J. *et al.* Descending command neurons in the brainstem that halt locomotion. *Cell* **163,** 1191–1203 (2015).

106. Arber, S. & Costa, R. M. Connecting neuronal circuits for movement. *Science* **360,** 1403–1404 (June 2018).

107. Arber, S. & Costa, R. M. Networking brainstem and basal ganglia circuits for movement. *Nature Reviews Neuroscience* **23,** 342–360 (Apr. 2022).

108. Moreno-Lopez, Y., Bichara, C., Delbecq, G., Isope, P. & Cordero-Erausquin, M. The corticospinal tract primarily modulates sensory inputs in the mouse lumbar cord. *eLife* **10** (Sept. 2021).

109. Svane, C., Forman, C. R., Nielsen, J. B. & Geertsen, S. S. Characterization of corticospinal activation of finger motor neurons during precision and power grip in humans. *Experimental Brain Research* **236,** 745–753 (Jan. 2018).

110. Bacon, S. & Smith, A. A monosynaptic pathway from an identified vasomotor centre in the medial prefrontal cortex to an autonomic area in the thoracic spinal cord. *Neuroscience* **54,** 719–728 (June 1993).

111. Norton, J. J. & Wolpaw, J. R. Acquisition, maintenance, and therapeutic use of a simple motor skill. *Current Opinion in Behavioral Sciences* **20,** 138–144 (Apr. 2018).

112. Martin, J. H., Kably, B. & Hacking, A. Activity-dependent development of cortical axon terminations in the spinal cord and brain stem. *Experimental Brain Research* **125,** 184–199 (Mar. 1999).

113. Bidaye, S. S., Bockemühl, T. & Büschges, A. Six-legged walking in insects: how CPGs, peripheral feedback, and descending signals generate coordinated and adaptive motor rhythms. *Journal of Neurophysiology* **119,** 459–475 (2018).

114. Court, R. *et al.* A systematic nomenclature for the *Drosophila* ventral nerve cord. *Neuron* **107,** 1071–1079.e2 (2020).

115. Wosnitza, A., Bockemühl, T., Dübbert, M., Scholz, H. & Büschges, A. Inter-leg coordination in the control of walking speed in *Drosophila*. *Journal of Experimental Biology* **216,** 480–491 (2013).

116. Mendes, C. S., Bartos, I., Akay, T., Márka, S. & Mann, R. S. Quantification of gait parameters in freely walking wild type and sensory deprived *Drosophila* melanogaster. *eLife* **2,** e00231 (2013).

117. Niven, J. E. Visuomotor control: *Drosophila* bridges the gap. *Current Biology* **20,** R309–R311 (2010).

118. Card, G. & Dickinson, M. H. Visually mediated motor planning in the escape response of *Drosophila*. *Current Biology* **18,** 1300–1307 (2008).

119. Pavlou, H. J. & Goodwin, S. F. Courtship behavior in *Drosophila melanogaster*: towards a 'courtship connectome'. *Current Opinion in Neurobiology* **23,** 76–83 (2013).

120. Zwarts, L., Versteven, M. & Callaerts, P. Genetics and neurobiology of aggression in *Drosophila*. *Fly* **6,** 35–48 (2012).

121. Seeds, A. M. *et al.* A suppression hierarchy among competing motor programs drives sequential grooming in *Drosophila*. *eLife* **3,** e02951 (2014).

122. Trimarchi, J. R. & Murphey, R. K. The *shaking-B*^{2} Mutation Disrupts Electrical Synapses in a Flight Circuit in Adult *Drosophila*. *The Journal of Neuroscience* **17,** 4700–4710 (June 1997).

123. Ikeda, K. & Kaplan, W. D. Neurophysiological genetics in *Drosophila melanogaster*. *American Zoologist* **14,** 1055–1066 (2015).

124. Hedwig, B. & Burrows, M. Presynaptic inhibition of sensory neurons during kicking movements in the locust. *Journal of Neurophysiology* **75,** 1221–1232 (1996).

125. Bässler, U. & Büschges, A. Pattern generation for stick insect walking movements—multisensory control of a locomotor program. *Brain Research Reviews* **27,** 65–88 (1998).

126. Maimon, G., Straw, A. D. & Dickinson, M. H. Active flight increases the gain of visual motion processing in *Drosophila*. *Nature Neuroscience* **13,** 393–399 (2010).

127. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnology* **22,** 1567–1572 (2004).

128. Jenett, A. *et al.* A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Reports* **2,** 991–1001 (2012).

129. Hedwig, B. & Poulet, J. F. A. Complex auditory behaviour emerges from simple reactive steering. *Nature* **430,** 781–785 (2004).

130. Enriquez, J. *et al.* Specification of Individual Adult Motor Neuron Morphologies by Combinatorial Transcription Factor Codes. *Neuron* **86,** 955–970 (2015).

131. Sen, R. *et al.* Moonwalker descending neurons mediate visually evoked retreat in *Drosophila*. *Current Biology* **27,** 766–771 (2017).

132. White, K. *et al.* Genetic control of programmed cell death in *Drosophila*. *Science* **264,** 677–683 (1994).

133. Ramdya, P., Schaffter, T., Floreano, D. & Benton, R. Fluorescence Behavioral Imaging (FBI) tracks identity in heterogeneous groups of *Drosophila*. *PLOS ONE* **7,** 1–8 (2012).

134. Reedy, M. C., Bullard, B. & Vigoreaux, J. O. Flightin is essential for thick filament assembly and sarcomere stability in *Drosophila* flight muscles. *Journal of Cell Biology* **151,** 1483–1500 (2000).

135. Lu, R. *et al.* Video-rate volumetric functional imaging of the brain at synaptic resolution. *Nature Neuroscience* **20,** 620–628 (2017).

136. Robie, A. A. *et al.* Mapping the Neural Substrates of Behavior. *Cell* **170,** 393–406.e28 (2017).

137. Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E. & Perrimon, N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nature Genetics* **40,** 476–483 (2008).

138. Nongthomba, U. & Ramachandra, N. B. A direct screen identifies new flight muscle mutants on the*Drosophila* second chromosome. *Genetics* **153,** 261–274 (1999).

139. Viswanathan, M. C., Blice-Baum, A. C., Schmidt, W., Foster, D. B. & Cammarato, A. Pseudo-acetylation of K326 and K328 of actin disrupts *Drosophila melanogaster* indirect flight muscle structure and performance. *Frontiers in Physiology* **6** (2015).

140. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9,** 676–682 (2012).

141. Weir, P. T. & Dickinson, M. H. Functional divisions for visual processing in the central brain of flying *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **112,** E5523–5532 (2015).

142. Fayyazuddin, A. & Dickinson, M. H. Haltere afferents provide Direct, electrotonic input to a steering motor neuron in the blowfly, *Calliphora*. *Journal of Neuroscience* **16,** 5225–5232 (1996).

143. Horn, B. K. & Schunck, B. G. Determining optical flow. *Artificial Intelligence* **17,** 185–203 (1981).

144. Brox, T., Bruhn, A., Papenberg, N. & Weickert, J. English. in *Computer Vision* 1st ed., 25–36 (Springer, Berlin, Heidelberg, 2004).

145. Revaud, J., Weinzaepfel, P., Harchaoui, Z. & Schmid, C. DeepMatching: Hierarchical deformable dense matching. *arXiv* (2015).

146. Pnevmatikakis, E. A. & Giovannucci, A. NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *Journal of Neuroscience Methods* **291,** 83–94 (2017).

147. Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundation and Trends® in Machine Learning* **3,** 1–122 (2011).

148. Unser, M., Soubies, E., Soulez, F., McCann, M. & Donati, L. *GlobalBioIm: A unifying computational framework for solving inverse problems* in *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)* (Optica Publishing Group, 2017), CTu1B.1.

149. Sun, D., Roth, S. & Black, M. J. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision* **106,** 115–137 (2014).

150. Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. *Opt. Lett.* **33,** 156–158 (2008).

151. Weir, P. T. *et al.* Anatomical reconstruction and functional imaging reveal an ordered array of skylight polarization detectors in *Drosophila*. *Journal of Neuroscience* **36,** 5397–5404 (2016).

152. Ikegaya, Y. *et al.* Synfire chains and cortical songs: temporal modules of cortical activity. *Science* **304,** 559–564 (2004).

153. Brooks, R. A. A robust layered control system for a mobile robot. *IEEE Journal on Robotics and Automation* **2,** 14–23 (1986).

154. Niell, C. M. & Stryker, M. P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65,** 472–479 (2010).

155. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience* **22,** 1677–1686 (2019).

156. Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364,** 255–255 (2019).

157. Chiappe, M. E., Seelig, J. D., Reiser, M. B. & Jayaraman, V. Walking modulates speed sensitivity in *Drosophila* motion vision. *Current Biology* **20,** 1470–1475 (2010).

158. Fujiwara, T., Cruz, T. L., Bohnslav, J. P. & Chiappe, M. E. A faithful internal representation of walking movements in the *Drosophila* visual system. *Nature Neuroscience* **20,** 72–81 (2016).

159. Aimon, S. *et al.* Fast near-whole–brain imaging in adult *Drosophila* during responses to stimuli and behavior. *PLOS Biology* **17,** e2006732 (2019).

160. Kim, A. J., Fitzgerald, J. K. & Maimon, G. Cellular evidence for efference copy in *Drosophila* visuomotor processing. *Nature Neuroscience* **18,** 1247–1255 (2015).

161. Zacarias, R., Namiki, S., Card, G. M., Vasconcelos, M. L. & Moita, M. A. Speed dependent descending control of freezing behavior in *Drosophila melanogaster*. *Nature Communications* **9,** 3697 (2018).

162. Patestas, M. & Gartner, L. P. in *A Textbook of Neuroanatomy* 1st ed., 137–170 (Wiley, 2006).

163. Poulet, J. F. & Hedwig, B. New insights into corollary discharges mediated by identified neural pathways. *Trends in Neurosciences* **30,** 14–21 (2007).

164. Buchanan, J. T. & Einum, J. F. The spinobulbar system in lamprey. *Brain Research Reviews* **57,** 37–45 (2008).

165. Stecina, K., Fedirchuk, B. & Hultborn, H. Information to cerebellum on spinal motor networks mediated by the dorsal spinocerebellar tract. *The Journal of Physiology* **591,** 5433–5443 (2013).

166. Homberg, U. Flight-correlated activity changes in neurons of the lateral accessory lobes in the brain of the locust *Schistocerca gregaria*. *Journal of Comparative Physiology A* **175,** 597–610 (1994).

167. Burrows, M. English. in *The Neurobiology of An Insect Brain* 1st ed., 541–544 (Oxford University Press, Oxford, 1996).

168. Chen, C. *et al.* Functional architecture of neural circuits for leg proprioception in *Drosophila*. *Current Biology* **31,** 1–13 (2021).

169. Agrawal, S. *et al.* Central processing of leg proprioception in *Drosophila*. *eLife* **9** (eds Calabrese, R. L., Marder, E. & Fujiwara, T.) e60299 (2020).

170. Nern, A., Pfeiffer, B. D. & Rubin, G. M. Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. *Proceedings of the National Academy of Sciences of the United States of America* **112,** E2967 LP –E2976 (2015).

171. Chen, C.-L. *et al.* Imaging neural activity in the ventral nerve cord of behaving adult *Drosophila*. *Nature Communications* **9,** 4390 (2018).

172. Panser, K. *et al.* Automatic segmentation of *Drosophila* neural compartments using GAL4 expression data reveals novel visual pathways. *Current Biology* **26,** 1943–1954 (2016).

173. Mohamed, A. A. M., Hansson, B. S. & Sachse, S. Third-order neurons in the lateral horn Enhance bilateral contrast of odor inputs through contralateral inhibition in *Drosophila*. *Frontiers in Physiology* **10,** 851 (2019).

174. Matsuo, E. *et al.* Organization of projection neurons and local neurons of the primary auditory center in the fruit fly *Drosophila melanogaster*. *Journal of Comparative Neurology* **524,** 1099–1164 (2016).

175. Lai, J. S.-Y., Lo, S.-J., Dickson, B. J. & Chiang, A.-S. Auditory circuit in the *Drosophila* brain. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 2607–2612 (2012).

176. Kamikouchi, A., Shimada, T. & Ito, K. Comprehensive classification of the auditory sensory projections in the brain of the fruit fly *Drosophila melanogaster*. *Journal of Comparative Neurology* **499,** 317–356 (2006).

177. Miyamoto, T. & Amrein, H. Suppression of male courtship by a *Drosophila* pheromone receptor. *Nature Neuroscience* **11,** 874–876 (2008).

178. Trueta, C. & De-Miguel, F. Extrasynaptic exocytosis and its mechanisms: a source of molecules mediating volume transmission in the nervous system. *Frontiers in Physiology* **3,** 319 (2012).

179. Edwards, C. J., Leary, C. J. & Rose, G. J. Counting on inhibition and rate-dependent excitation in the auditory system. *Journal of Neuroscience* **27,** 13384–13392 (2007).

180. Naud, R., Houtman, D., Rose, G. J. & Longtin, A. Counting on dis-inhibition: a circuit motif for interval counting and selectivity in the anuran auditory system. *Journal of Neurophysiology* **114,** 2804–2815 (2015).

181. Barak, O., Sussillo, D., Romo, R., Tsodyks, M. & Abbott, L. From fixed points to chaos: three models of delayed discrimination. *Progress in Neurobiology* **103,** 214–222 (2013).

182. Miller, P. Dynamical systems, attractors, and neural circuits. *F1000Research* **5,** F1000 Faculty Rev–992 (2016).

183. Van Alphen, B., Semenza, E. R., Yap, M., van Swinderen, B. & Allada, R. A deep sleep stage in *Drosophila* with a functional role in waste clearance. *Science Advances* **7,** eabc2999 (2021).

184. Bosco, G. & Poppele, R. Proprioception from a spinocerebellar perspective. *Physiological Reviews* **81,** 539–568 (2001).

185. Mamiya, A., Gurung, P. & Tuthill, J. C. Neural coding of leg proprioception in *Drosophila*. *Neuron* **100,** 636–650.e6 (2018).

186. Isakov, A. *et al.* Recovery of locomotion after injury in *Drosophila melanogaster* depends on proprioception. *The Journal of Experimental Biology* **219,** 1760–1771 (2016).

187. Hermans, L. *et al.* Long-term imaging of the ventral nerve cord in behaving adult *Drosophila*. *bioRxiv,* 2021.10.15.463778 (2021).

188. Dubuc, R. & Grillner, S. The role of spinal cord inputs in modulating the activity of reticulospinal neurons during fictive locomotion in the lamprey. *Brain Research* **483,** 196–200 (1989).

189. Rothwell, J. in *Control of Human Voluntary Movement* 217–251 (Springer, 1994).

190. Choi, S. *et al.* Parallel ascending spinal pathways for affective touch and pain. *Nature* **587,** 258–263 (2020).

191. Jefferis, G. S. *et al.* Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell* **128,** 1187–1203 (2007).

192. Bogovic, J. A. *et al.* An unbiased template of the *Drosophila* brain and ventral nerve cord. *PLOS ONE* **15,** e0236495 (2021).

193. Berman, G. J., Choi, D. M., Bialek, W. & Shaevitz, J. W. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface* **11,** 20140672 (2014).

194. Graving, J. M. *behavelet: a wavelet transform for mapping behavior* 2019.

195. Ke, G. *et al. LightGBM: a highly efficient gradient boosting decision tree* in *Advances in Neural Information Processing Systems* (eds Guyon, I. *et al.*) **30** (Curran Associates, Inc., 2017).

196. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* **16,** 321–357 (2002).

197. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33,** 1–22 (2010).

198. Razali, N. M. & Wah, Y. B. Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics* **2,** 21–23 (2011).

199. Ronneberger, O., P.Fischer & Brox, T. *U-Net: convolutional networks for biomedical image segmentation* in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* **9351** (Springer, 2015), 234–241.

200. Payer, C., Štern, D., Neff, T., Bischof, H. & Urschler, M. *Instance segmentation and tracking with cosine embeddings and recurrent hourglass networks* in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (eds Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C. & Fichtinger, G.) (Springer International Publishing, Cham, 2018), 3–11.

201. Çiçek, O., Abdulkadir, A., Lienkamp, S., Brox, T. & Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *arXiv,* 1606.06650 (2016).

202. Wang, P., Cuccolo, N. G., Tyagi, R., Hacihaliloglu, I. & Patel, V. M. *Automatic real-time CNN-based neonatal brain ventricles segmentation* in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (2018), 716–719.

203. Ioffe, S. & Szegedy, C. *Batch normalization: accelerating deep network training by reducing internal covariate shift* in (eds Bach, F. & Blei, D.) **37** (PMLR, Lille, France, 2015), 448–456.

204. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv,* 1412.6980 (2017).

205. Sørensen, T. J. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skar.* **5,** 1–34 (1948).

206. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* **26,** 297–302 (1945).

207. Buades, A., Coll, B. & Morel, J. M. *Denoising image sequences does not require motion estimation* in *IEEE Conference on Advanced Video and Signal Based Surveillance, 2005.* (2005), 70–74.

208. Bradski, G. The OpenCV library. *Dr. Dobb's Journal of Software Tools* **122-125,** 120 (2000).

209. Otsu, N. A Threshold selection method from gray-Level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9,** 62–66 (1979).

210. Ankerst, M., Breunig, M. M., Kriegel, H.-P. & Sander, J. *OPTICS: ordering points to identify the clustering structure* in *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data* (Association for Computing Machinery, Philadelphia, Pennsylvania, USA, 1999), 49–60.

211. Kuhn, H. W. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* **2,** 83–97 (1955).

212. Van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2,** e453 (2014).

213. Guizar, M. *Efficient subpixel image registration by cross-correlation* https : / / www . mathworks . com / matlabcentral / fileexchange / 18401 - efficient - subpixel - image - registration - by - cross - correlation, MATLAB Central File Exchange. Retrieved November 9, 2020. (2020).

214. Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annual review of neuroscience* **33,** 269–298 (2010).

215. Caggiano, V. *et al.* Midbrain circuits that set locomotor speed and gait selection. *Nature* **553,** 455–460 (2018).

216. Orger, M. B., Kampff, A. R., Severi, K. E., Bollmann, J. H. & Engert, F. Control of visually guided behavior by distinct populations of spinal projection neurons. *Nature neuroscience* **11,** 327–333 (2008).

217. Olsen, S. R. & Wilson, R. I. Cracking neural circuits in a tiny brain: new approaches for understanding the neural circuitry of Drosophila. *Trends in neurosciences* **31,** 512–520 (2008).

218. Zheng, Z. *et al.* A complete electron microscopy volume of the brain of adult Drosophila melanogaster. *Cell* **174,** 730–743 (2018).

219. Lima, S. Q. & Miesenböck, G. Remote control of behavior through genetically targeted photostimulation of neurons. *Cell* **121,** 141–152 (2005).

220. Israel, S., Rozenfeld, E., Weber, D., Huetteroth, W. & Parnas, M. Olfactory stimuli and moonwalker SEZ neurons can drive backward locomotion in Drosophila. *Current Biology* **32,** 1131–1149.e7 (2022).

221. Mann, K., Gallen, C. L. & Clandinin, T. R. Whole-brain calcium imaging reveals an intrinsic functional network in Drosophila. *Current Biology* **27,** 2389–2396 (2017).

222. Asahina, K. *et al.* Tachykinin-Expressing Neurons Control Male-Specific Aggressive Arousal in Drosophila. *Cell* **156,** 221–235 (Jan. 2014).

223. Lecoq, J. *et al.* Removing independent noise in systems neuroscience data using deepinterpolation. *Nature Methods* **18,** 1401–1408 (2021).

224. Semmelhack, J. L. & Wang, J. W. Select Drosophila glomeruli mediate innate olfactory attraction and aversion. *Nature* **459,** 218–223 (2009).

225. Ravbar, P., Zhang, N. & Simpson, J. H. Behavioral evidence for nested central pattern generator control of Drosophila grooming. *eLife* **10** (Dec. 2021).

226. Kato, S. *et al.* Global brain dynamics embed the motor command sequence of Caenorhabditis elegans. *Cell* **163,** 656–669 (2015).

227. Braitenberg, V. *Vehicles: Experiments in synthetic psychology* (MIT press, 1986).

228. Chen, C.-L. *et al.* Ascending neurons convey behavioral state to integrative sensory and action selection centers in the brain. *bioRxiv* (2022).

229. Schaffer, E. S. *et al.* Flygenvectors: the spatial and temporal structure of neural activity across the fly brain. *bioRxiv* (2021).

230. Brezovec, L. E., Berger, A. B., Druckmann, S. & Clandinin, T. R. Mapping the Neural Dynamics of Locomotion across the Drosophila Brain. *bioRxiv* (2022).

231. Tanaka, R. & Clark, D. A. Neural mechanisms to exploit positional geometry for collision avoidance. *Current Biology* (2022).

232. Triphan, T., Poeck, B., Neuser, K. & Strauss, R. Visual targeting of motor actions in climbing Drosophila. *Current biology* **20,** 663–668 (2010).

233. Coen, P. *et al.* Dynamic sensory cues shape song structure in Drosophila. *Nature* **507,** 233–237 (2014).

234. Nässel, D. R., Högmo, O. & Hallberg, E. Antennal receptors in the blowfly *Calliphora erythrocephala*. I. The gigantic central projection of the pedicellar campaniform sensillum. *Journal of Morphology* **180,** 159–169 (May 1984).

235. King, D. G. & Wyman, R. J. Anatomy of the giant fibre pathway inDrosophila. I. Three thoracic components of the pathway. *Journal of Neurocytology* **9,** 753–770 (Dec. 1980).

236. Azevedo, A. W. *et al.* A size principle for recruitment of Drosophila leg motor neurons. *eLife* **9** (June 2020).

237. Piatkevich, K. D. *et al.* A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nature Chemical Biology* **14,** 352–360 (Feb. 2018).

238. Voleti, V. *et al.* Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. *Nature methods* **16,** 1054–1062 (2019).

239. Aimon, S., Cheng, K. Y., Gjorgjieva, J. & Kadow, I. C. G. Walking elicits global brain activity in *Drosophila*. *bioRxiv* (Jan. 2022).

240. Simpson, J. H. Rationally subdividing the fly nervous system with versatile expression reagents. *Journal of Neurogenetics* **30,** 185–194 (Oct. 2016).

241. Krull, A., Buchholz, T.-O. & Jug, F. *Noise2void-learning denoising from single noisy images* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), 2129–2137.

242. Saalfeld, S., Cardona, A., Hartenstein, V. & Tomančák, P. CATMAID: collaborative annotation toolkit for massive amounts of image data. *Bioinformatics* **25,** 1984–1986 (Apr. 2009).

243. Schneider-Mizell, C. M. *et al.* Quantitative neuroanatomy for connectomics in *Drosophila*. *eLife* **5** (ed Calabrese, R. L.) e12059 (Mar. 2016).

244. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17,** 261–272 (2020).

245. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12,** 2825–2830 (2011).

246. Chartrand, R. Numerical Differentiation of Noisy, Nonsmooth Data. *ISRN Applied Mathematics* **2011,** 1–11 (May 2011).

247. Jr., J. H. W. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **58,** 236–244 (1963).

248. Jiao, W. *et al.* Intact Drosophila Whole Brain Cellular Quantitation reveals Sexual Dimorphism. *bioRxiv* (Nov. 2021).

249. Kaplan, H. S., Salazar Thula, O., Khoss, N. & Zimmer, M. Nested neuronal dynamics orchestrate a behavioral hierarchy across timescales. en. *Neuron* **105,** 562–576.e9 (Feb. 2020).

250. Gouwens, N. W. & Wilson, R. I. Signal Propagation in Drosophila Central Neurons. *Journal of Neuroscience* **29,** 6239–6249 (May 2009).

251. Xu, Y., Zou, P. & Cohen, A. E. Voltage imaging with genetically encoded indicators. *Current Opinion in Chemical Biology* **39.** Molecular Imaging Chemical Genetics and Epigenetics, 1–10 (2017).

252. Zhu, M. H., Jang, J., Milosevic, M. M. & Antic, S. D. Population imaging discrepancies between a genetically-encoded calcium indicator (GECI) versus a genetically-encoded voltage indicator (GEVI). *Scientific Reports* **11,** 5295 (2021).

253. Fan, J. L. *et al.* High-speed volumetric two-photon fluorescence imaging of neurovascular dynamics. *Nature Communications* **11** (Nov. 2020).

254. Rózsa, B., Szalay, G. & Katona, G. in *Advanced Patch-Clamp Analysis for Neuroscientists* 213–245 (Springer New York, 2016).

255. Niesner, R., Andresen, V., Neumann, J., Spiecker, H. & Gunzer, M. The Power of Single and Multibeam Two-Photon Microscopy for High-Resolution and High-Speed Deep Tissue and Intravital Imaging. *Biophysical Journal* **93,** 2519–2529 (Oct. 2007).

256. Mahn, M. *et al.* High-efficiency optogenetic silencing with soma-targeted anion-conducting channelrhodopsins. *Nature Communications* **9** (Oct. 2018).

257. Bovetti, S. & Fellin, T. Optical dissection of brain circuits with patterned illumination through the phase modulation of light. *Journal of Neuroscience Methods* **241,** 66–77 (Feb. 2015).

258. Pick, S. & Strauss, R. Goal-driven behavioral adaptations in gap-climbing *Drosophila*. *Current Biology* **15,** 1473–1478 (2005).

259. Warren, R. A. *et al.* A rapid whisker-based decision underlying skilled locomotion in mice. *eLife* **10** (Jan. 2021).

260. Ahmed, H., Wilbur, R. B., Bharadwaj, H. M. & Siskind, J. M. *Object Classification From Randomized EEG Trials* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021).

261. Spampinato, C. *et al. Deep Learning Human Mind for Automated Visual Classification* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017).

262. Palazzo, S., Kavasidis, I., Kastaniotis, D. & Dimitriadis, S. I. *Recent Advances at the Brain-Driven Computer Vision Workshop 2018* in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops* (2018).

263. Palazzo, S. *et al.* Decoding Brain Representations by Multimodal Learning of Neural Activity and Visual Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2021).

264. Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging Large-Scale Neural Activity with Cellular Resolution in Awake, Mobile Mice. *Neuron* **56,** 43–57 (2007).

265. Pandarinath, C. *et al.* Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods* **15,** 805–815 (2018).

266. Ecker, A. S. *et al.* Decorrelated Neuronal Firing in Cortical Microcircuits. *Science* **327,** 584–587 (2010).

267. Topalovic, U. *et al.* Wireless Programmable Recording and Stimulation of Deep Brain Activity in Freely Moving Humans. *Neuron* **108,** 322–334.e9 (2020).

268. Urai, A. E., Doiron, B., Leifer, A. M. & Churchland, A. K. Large-scale neural recordings call for new insights to link brain and behavior. *arXiv* (2021).

269. Pei, F. *et al. Neural Latents Benchmark '21: Evaluating latent variable models of neural population activity* 2021.

270. Nakagome, S., Luu, T. P., He, Y., Ravindran, A. S. & Contreras-Vidal, J. L. An empirical comparison of neural networks and machine learning algorithms for EEG gait decoding. *Nature Scientific Reports* **10,** 4372 (2020).

271. Glaser, J. I. *et al.* Machine Learning for Neural Decoding. *eNeuro* **7** (2020).

272. Batty, E. *et al. BehaveNet: nonlinear embedding and Bayesian neural decoding of behavioral videos* in *Advances in Neural Information Processing Systems (NeurIPS)* (2019).

273. Kostas, D., Aroca-Ouellette, S. & Rudzicz, F. BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data. *Frontiers in Human Neuroscience* **15,** 253 (2021).

274. Peterson, S. *ECoG and arm position during movement and rest* Sept. 2021.

275. Singh, S. H., Peterson, S. M., Rao, R. P. & Brunton, B. W. Mining naturalistic human behaviors in long-term video and neural recordings. *Journal of Neuroscience Methods* **358,** 109199 (2021).

276. Ionescu, C., Papava, D., Olaru, V. & Sminchisescu, C. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36,** 1325–1339 (2014).

277. Wen, S. *et al.* Capturing spike train temporal pattern with wavelet average coefficient for brain machine interface. *Scientific Reports* **11,** 19020 (2021).

278. Lau, C. K. S., Jelen, M. & Gordon, M. D. A closed-loop optogenetic screen for neurons controlling feeding in Drosophila. *G3 (Bethesda)* **11** (May 2021).

279. Collinger, J. L., Gaunt, R. A. & Schwartz, A. B. Progress towards restoring upper limb movement and sensation through intracortical brain-computer interfaces. *Current Opinion in Biomedical Engineering* **8.** Neural Engineering/ Novel Biomedical Technologies: Neuromodulation, 84–92 (2018).

280. Ganzer, P. D. *et al.* Restoring the Sense of Touch Using a Sensorimotor Demultiplexing Neural Interface. *Cell* **181,** 763–773.e12 (2020).

281. Utsumi, K. *et al.* Operation of a P300-based brain-computer interface in patients with Duchenne muscular dystrophy. *Scientific Reports* **8,** 1753 (2018).

282. Sani, O. G. *et al.* Mood variations decoded from multi-site intracranial human brain activity. *Nature Biotechnology* **36,** 954–961 (2018).

283. Lacourse, K., Yetton, B., Mednick, S. & Warby, S. C. Massive online data annotation, crowdsourcing to generate high quality sleep spindle annotations from EEG data. *Scientific Data* **7,** 190 (2020).

284. Wang, X., Farhadi, A., Rao, R. P. N. & Brunton, B. W. *AJILE Movement Prediction: Multimodal Deep Learning for Natural Human Neural Recordings and Video* in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)* (2018).

285. Mohsenvand, M. N., Izadi, M. R. & Maes, P. *Contrastive Representation Learning for Electroencephalogram Classification* in *Proceedings of the Machine Learning for Health NeurIPS Workshop* (2020).

286. Peterson, S. M., Rao, R. P. N. & Brunton, B. W. Learning neural decoders without labels using multiple data streams. *bioRxiv* (2021).

287. Li, W. *et al.* Multi-source domain adaptation for decoder calibration of intracortical brain-machine interface. *Journal of neural engineering* **17** (Oct. 2020).

288. Farshchian, A. *et al.* Adversarial Domain Adaptation for Stable Brain-Machine Interfaces. *arXiv* (2018).

289. Liu, Y., Yan, Q. & Alahi, A. Social NCE: Contrastive Learning of Socially-aware Motion Representations. *arXiv* (2020).

290. Su, K., Liu, X. & Shlizerman, E. *Predict & cluster: Unsupervised skeleton based action recognition* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020).

291. Lin, L., Song, S., Yang, W. & Liu, J. *MS2L: Multi-Task Self-Supervised Learning for Skeleton Based Action Recognition* in *Proceedings of the ACM International Conference on Multimedia* (2020).

292. Pan, T., Song, Y., Yang, T., Jiang, W. & Liu, W. *Videomoco: Contrastive video representation learning with temporally adversarial examples* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021).

293. Dave, I., Gupta, R., Rizve, M. N. & Shah, M. TCLR: Temporal Contrastive Learning for Video Representation. *arXiv* (2021).

294. Munro, J. & Damen, D. *Multi-modal Domain Adaptation for Fine-grained Action Recognition* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020).

295. Chen, M.-H. *et al. Temporal attentive alignment for large-scale video domain adaptation* in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2019).

296. Xu, Y. *et al.* Aligning Correlation Information for Domain Adaptation in Action Recognition. *arXiv* (2021).

297. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. *A Simple Framework for Contrastive Learning of Visual Representations* in *Proceedings of the International Conference on Machine Learning (ICML)* (2020).

298. Nassar, J., Linderman, S. W., Bugallo, M. & Park, I.-S. Tree-Structured Recurrent Switching Linear Dynamical Systems for Multi-Scale Modeling. *arXiv* (2019).

299. Linderman, S., Nichols, A., Blei, D., Zimmer, M. & Paninski, L. Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in C. elegans. *bioRxiv* (2019).

300. Linderman, S. *et al. Bayesian Learning and Inference in Recurrent Switching Linear Dynamical Systems* in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)* (2017).

301. Gao, Y., Archer, E., Paninski, L. & Cunningham, J. *Linear dynamical neural population models through nonlinear embeddings* in *Advances in Neural Information Processing Systems (NeurIPS)* (2016).

302. Abbaspourazad, H., Choudhury, M., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Multi-scale low-dimensional motor cortical state dynamics predict naturalistic reach-and-grasp behavior. *Nature Communications* **12,** 607 (2021).

303. Shenoy, K. V. & Kao, J. C. Measurement, manipulation and modeling of brain-wide neural population dynamics. *Nature Communications* **12,** 633 (2021).

304. Wu, A. *et al. Deep Graph Pose: a semi-supervised deep graphical model for improved animal pose tracking* in *Advances in Neural Information Processing Systems (NeurIPS)* (2020).

305. Bala, P. C. *et al.* Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio. *Nature Communications* **11,** 4560 (2020).

306. Graving, J. M. *et al.* DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* **8** (eds Baldwin, I. T., Shaevitz, J. W., Shaevitz, J. W. & Stephens, G.) e47994 (2019).

307. Li, S. *et al. Deformation-aware Unpaired Image Translation for Pose Estimation on Laboratory Animals* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020).

308. Pedersen, M., Haurum, J. B., Bengtson, S. H. & Moeslund, T. B. *3D-ZeF: A 3D Zebrafish Tracking Benchmark Dataset* in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2020).

309. Sun, J. J. *et al. Task programming: Learning data efficient behavior representations* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021).

310. Segalin, C. *et al.* The Mouse Action Recognition System (MARS): a software pipeline for automated analysis of social behaviors in mice. *bioRxiv* (2020).

311. Overman, K., Choi, D., Leung, K., Shaevitz, J. & Berman, G. Measuring the repertoire of age-related behavioral changes in Drosophila melanogaster. *bioRxiv* (2021).

312. Pereira, T. D., Shaevitz, J. W. & Murthy, M. Quantifying behavior to understand the brain. *Nature Neuroscience* **23,** 1537–1549 (2020).

313. Johnson, R. E. *et al.* Probabilistic Models of Larval Zebrafish Behavior Reveal Structure on Many Scales. *Current Biology* **30,** 70–82.e4 (2020).

314. Sani, O. G., Abbaspourazad, H., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nature Neuroscience* **24,** 140–149 (2021).

315. Berthelot, D. *et al.* in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (2019).

316. Verma, V. *et al. Manifold Mixup: Better Representations by Interpolating Hidden States* in *International Conference on Learning Representations, (ICLR)* (2019).

317. Zhang, H., Cissé, M., Dauphin, Y. N. & Lopez-Paz, D. *mixup: Beyond Empirical Risk Minimization* in *International Conference on Learning Representations, (ICLR)* (2018).

318. Shen, Z. *et al. Un-Mix: Rethinking Image Mixtures for Unsupervised Visual Representation Learning* in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)* (2022).

319. Lee, K. *et al. i-Mix: A Domain-Agnostic Strategy for Contrastive Representation Learning* in *International Conference on Learning Representations, (ICLR)* (2021).

320. Sahoo, A., Shah, R., Panda, R., Saenko, K. & Das, A. Contrast and Mix: Temporal Contrastive Video Domain Adaptation with Background Mixing. *arXiv* (2021).

321. Mao, X., Ma, Y., Yang, Z., Chen, Y. & Li, Q. Virtual Mixup Training for Unsupervised Domain Adaptation. *arXiv* (2019).

322. Sahoo, A., Panda, R., Feris, R. S., Saenko, K. & Das, A. Select, Label, and Mix: Learning Discriminative Invariant Feature Representations for Partial Domain Adaptation. *arXiv* (2020).

323. Zhang, Y., Jiang, H., Miura, Y., Manning, C. D. & Langlotz, C. P. Contrastive Learning of Medical Visual Representations from Paired Images and Text. *arXiv* (2020).

324. Yuan, X. *et al. Multimodal Contrastive Training for Visual Representation Learning* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021).

325. Van den Oord, A., Li, Y. & Vinyals, O. Representation Learning with Contrastive Predictive Coding. *arXiv* (2019).

326. Wei, J. & Zou, K. *EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks* in *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (2019).

327. Yun, S. *et al. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features* in *International Conference on Computer Vision (ICCV)* (2019).

328. Ganin, Y. & Lempitsky, V. *Unsupervised Domain Adaptation by Backpropagation* in *Proceedings of the International Conference on Machine Learning (ICML)* (2015).

329. Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. & Smola, A. J. *A Kernel Method for the Two-Sample-Problem* in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)* (2006).

330. Kang, G., Jiang, L., Wei, Y., Yang, Y. & Hauptmann, A. G. Contrastive Adaptation Network for Single-and Multi-Source Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2020).

331. He, K., Fan, H., Wu, Y., Xie, S. & Girshick, R. *Momentum Contrast for Unsupervised Visual Representation Learning* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2020).

332. Giovannucci, A. *et al.* CaImAn: An open source tool for scalable Calcium Imaging data Analysis. *eLife* **8,** e38173 (2019).

333. Aymanns, F. utils2p. *https://doi.org/10.5281/zenodo.5501119* (Sept. 2021).

334. Aymanns, F. ofco: optical flow motion correction. *https://doi.org/10.5281/zenodo.5518800* (Sept. 2021).

335. Pnevmatikakis, E. A., Merel, J., Pakman, A. & Paninski, L. Bayesian spike inference from calcium imaging data. *arXiv* (2013).

336. Rupprecht, P. *et al.* A database and deep learning toolbox for noise-optimized, generalized spike inference from calcium imaging. *Nature Neuroscience* **24,** 1324–1337 (2021).

337. Bahdanau, D., Cho, K. & Bengio, Y. *Neural machine translation by jointly learning to align and translate* in *Proceedings of the International Conference on Machine Learning (ICML)* (2015).