Title: Predicting natural behavior from whole-brain neural dynamics

Authors: Monika Scholz[1], Ashley N Linder[2], Francesco Randi[1], Anuj K Sharma[1], Xinwei Yu[1], Joshua W Shaevitz[1,3], Andrew M Leifer[1,2*]

Affiliations:

[1] Department of Physics, Princeton University, Princeton, NJ, USA

[2] Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA

[3] Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA

*Correspondence to: leifer@princeton.edu

Abstract: We record calcium activity from the majority of head neurons in freely moving *C. elegans* to reveal where and how natural behavior is encoded in a compact brain. We find that a sparse subset of neurons distributed throughout the head encode locomotion. A linear combination of these neurons' activity predicts the animal's velocity and body curvature and is sufficient to infer its posture. This sparse linear model outperforms single neuron or PCA models at predicting behavior. Among neurons important for the prediction are well-known locomotory neurons, such as AVA, as well as neurons not traditionally associated with locomotion. We compare neural activity of the same animal during unrestrained movement and during immobilization and find large differences between brain-wide neural dynamics during real and fictive locomotion.

One Sentence Summary: *C. elegans* behavior is predicted from neural activity.

Main Text:

How do patterns of neural activity in the brain represent an animal's behavior? Much of what is known about the neural codes for animal behaviors such as head direction (*1*, *2*), spatial location (*3*, *4*) or arm movements (*5*) comes from studying small regions of the brain. It is unclear to what extent neurons in other regions of the brain may also carry behavior related signals, or how those signals may be encoded. Recent reports of neural signals correlated with facial expression in sensory brain areas like the visual cortex (*6*) suggest that behavioral information may be prevalent throughout the brain and may account for a larger fraction of neural dynamics than previously thought. Advances in calcium imaging now allow for brain-wide investigations of neural activity at single cell resolution in small transparent organisms such as zebrafish (*7*), worms (*8*) or hydra (*9*). Pioneering work in these systems has begun to investigate how brain-wide patterns of neural activity correlate with behavior, primarily by measuring fictive locomotion from immobilized animals (*7*, *10*). Here we investigate brain-wide neural coding of behavior in unrestrained and freely moving *Caenorhabditis elegans*. We find that the brain exhibits starkly different neural dynamics during immobilization compared to awake and unrestrained locomotion. We present a linear model or 'decoder' that predicts the animal's velocity and body curvature from the activity of a subset of neurons spatially distributed throughout its head. Amongst those neurons are many that have previously been identified in

coding for locomotory behavior, such as AVA, SMB and AIY, as well as additional neurons that have not traditionally been implicated in locomotion (ALA, AIM, RIS, RMG, and RIP). Finally, we use our decoder to predict the animal's full posture from neural activity. Our work provides a revised picture of how behavior is encoded in brain-wide neural dynamics in the *C. elegans* model system.

Previous investigations of whole-brain neural activity of behavior in *C. elegans* have used animals that were immobilized with a paralytic, tetramisole (*10*, *11*). We wondered the extent to which neural dynamics of paralyzed or immobilized animals resembled those of freely moving animals. We used a whole-brain imaging system for freely moving animals ((*12*); see also (*13*)) to compare neural dynamics of the same animal during movement and immobilization. We recorded neural activity as an animal crawled freely in a microfluidic arena and then delivered a paralytic, part way through the same recording (Fig. 1A, additional trial in Fig. S1). Consistent with prior reports (*10*), neural dynamics during immobilization showed slow, antiphasic activity in two groups of neurons. In contrast, when the same animal was freely moving, the neural activity appeared to vary more rapidly and less coherently.

We recorded neural dynamics from 18 additional animals that were immobilized or freely moving for the entirety of the recording (including two that were reanalyzed from (*14*), Fig. S2, and Table S1). We calculated the autocorrelation of calcium activity for each condition, averaged across neurons (Fig. 1D). We found pronounced peaks in the correlogram of the immobilized recordings indicating slow (s), periodic activity of many neurons during immobilization. In contrast, such pronounced peaks were absent in the correlogram for activity during movement, indicating that during movement animals exhibited less periodic activity, especially on longer timescales. Here immobilized animals included those either paralyzed with tetramisole, or immobilized using nanobeads, both of which exhibited similar dynamics (see Supplementary Text and Fig. S3). We therefore concluded that neural dynamics during immobilization fluctuate more slowly and are more periodic than during movement.

To investigate the structure and dimensionality of neural dynamics during behavior we used principal components analysis (PCA, Fig. 1C). PCA extracts a basis set, called the principal components (PCs). A truncated set of PCs can be used to represent neural activity in a lower-dimensional space while retaining the features accounting for the most variance. In agreement with recent reports (*10*, *15*), we observed neural dynamics during immobilization that cycle repeatedly through a stereotyped trajectory in neural state space defined by the first three PCs (Fig. 1F, blue spiral-like trajectory). Interestingly, neural dynamics of the same animal during movement followed no such stereotyped structure (Fig. 1F, red hairball), suggesting a major difference between the neural dynamics of fictive and real movement.

We further wondered how well the observed neural dynamics were approximated by a low dimensional representation. The first three principal components explained 74% of the variance of neural activity for the immobile portion of our recording but only 59% for the moving portion (Fig. 1E). A statistical analysis of 18 recordings (moving and immobile) revealed that the number of significant principal components was larger for recordings during movement than during immobilization (Fig. S4, Supplementary Text). Similar analysis of calcium-insensitive GFP control animals suggest that these differences cannot be trivially explained by motion artefact (Supplementary Text, Fig. S4). That neural dynamics of movement are less well approximated by a three-dimensional representation than those during immobilization further suggests that there are important differences between neural dynamics of movement and

immobilization. We therefore proceeded to study neural activity exclusively during unrestrained movement.

To explore how behavior is encoded in the brain, we simultaneously measured locomotion and posture, and whole-brain neural activity. We extracted the animal's velocity and its body curvature (a feature directly related to turning behavior) using a PCA-based posture analysis (*16*). We imaged freely moving wild type (WT) animals on either an agar plate or in a microfluidic arena (Fig. 2). To control for potential imaging-induced behavior effects, we also imaged *lite-1* mutants that lack the animal's endogenous blue-light aversive response (*17*). Consistent with prior reports (*10–12*, *18*, *19*), we observed that the activity of individual neurons was correlated with velocity or body curvature (Fig. S5). We therefore developed models to *predict* such behavior from measured neural activity.

A previous report from immobilized animals suggests that neural activity projected along only the first three PCs should be sufficient to encode behavior (*10*). We therefore used linear regression to test how well linear combinations of projections of activity along the first three PCs predict a freely-moving animal's velocity and body curvature (PCA model, Fig. 2A-E, G). The regression model was trained on part of the recording and tested on the remaining portion (Fig. 2C). Surprisingly, this PCA model performed poorly at predicting velocity and body curvature (median R^2 of 0.13 and 0.01, respectively). For predicting velocity its performance was comparable to that of the best single neuron, and for predicting body curvature it was dramatically worse (Fig. 2G). Therefore, we conclude that for freely moving animals, the first three orthogonal components of neural activity do not trivially encode behavior.

We wondered whether a differently weighted collection of neurons could better predict behavior. We used sparse linear regression to find the linear combination of neurons from each recording whose collective activity best predicted turning and body curvature. `Elastic net' regularization was used to impose sparsity while retaining neurons with redundant but relevant signals. (see Fig. 2F, methods). We performed fitting and time-series cross-validation on part of the recording and tested the performance of the model on the remaining portion (Fig. 2C, methods). The sparse linear model (SLM) outperformed both the best single neuron and the PCA model at predicting velocity and body curvature, (median R^2 of 0.37 and 0.41, respectively). When applied to calcium insensitive GFP control animals, a similar SLM performed poorly, suggesting that the model derives its predictive power from calcium signals and not motion artefacts (Fig. 2G). Model performance was not restricted to a particular environment (plate or arena) or to a genetic background (WT or *lite-1*).

The SLM assigned non-zero weights to 19±5 and 15±2 neurons (mean ± s.e.m., across animals) for predicting velocity and body curvature, respectively, with little overlap, accounting for roughly 20% of the measured neurons (Fig. 3A, B). Neurons selected by the SLM were spread spatially throughout the recording region, with the highest density located in the nerve ring (Fig. 3C, D). Their location was roughly stereotyped across animals. We aligned neuron locations for each recording to a known cell atlas (*20*) and attempted to identify those neurons selected by the SLM by name. We were reassured to find many neurons that had previously been implicated in locomotory behavior, including neuron classes AVA, SMB, and AIY (Fig. 3C, D and Tab. S2) (*10*, *18*, *21*, *22*). Intriguingly, we also found neurons not traditionally known as locomotory, including ALA and RIS which are quiescence-promoting (*23*, *24*), AIM which is involved in foraging (*25*), RMG which is a multisensory integration hub (*26*) and RIP which is less well characterized (Fig. S6).

To infer the animal's posture from neural activity, we integrated information about the predicted behavioral variables— velocity and body curvature —over time. Given only neural activity and the animal's starting posture, we accurately predicted posture for up to 10 seconds, after which error accumulation caused the phase between the predicted and true postures to drift (Fig. 4A, B and Video S1). The SLM accurately predicts posture and posture dynamics during forward locomotion. However, during turns (when body curvature amplitude is large), the predicted curvature of the animal's posture is slightly too low (Fig. 4B). We found that adding an empirically derived nonlinearity to the body curvature decoder provided more accurate posture predictions during turns (Fig. 4C-F). This suggests that the nervous system may nonlinearly amplify curvature related neural signals, either in the brain or at the neuromuscular junction. Taken together, we conclude that a linear model with minor non-linear corrections is sufficient to fully reconstruct the animal's behavior and posture from a sparse combination of neural activity.

Discussion:

Our results provide new insights into where and how behavior is coded in a simple brain. In *C. elegans* we find that a sparse subset of neurons spatially located throughout the head encode behavior. A linear combination of activity from these neurons alone is sufficient to predict the animal's velocity and body curvature, and to further infer its posture. Among this subset of neurons are many that have long been associated with locomotion, like AVA, SMB, and AIY (*10*, *18*, *21*, *22*), as well as others that have not traditionally been implicated in locomotion, like RMG and RIP. Together, this suggests that neural coding of behavior is spread throughout the brain and includes more neurons than previously thought.

What role might these neurons have in behavior? Some, like AVA, have been shown to modulate reversal frequency and generate high-level motor commands directly (*19*). Other neurons, however, are probably either reporting motor commands generated elsewhere as an efference copy (*27*), or are detecting locomotion independently as in proprioception (*28*).

Our results also demonstrate dramatic differences between neural dynamics of fictive and real locomotion that inform our picture of the brain. Previous studies from immobilized *C. elegans* had reported that the brain's neural dynamics are dominated by low dimensional representations of behavior contained in the first three PCs of neural activity (*10*). That result implies that motor outputs constitute a large fraction of the brain's neural dynamics. In contrast, our measurements in freely moving animals indicate that the first three PCs are poor predictors of behavior. The neural coding of behavior we find in moving animals is linear and low dimensional, but it does not so dominate brain wide neural dynamics as to appear in the first few principal components.

If the first few PCs of brain wide activity are not coding for behavior, what might be their role instead? In addition to generating behavior, the nervous system must also sense an animal's environment and maintain internal states. In mice, an internal arousal state (as measured by pupil diameter) accounts for the majority of the variance of membrane potential in an arbitrary brain region such as auditory cortex (*29*). We speculate that in *C. elegans* the first few PCs of neural activity may be involved in either sensory coding or maintaining internal brain states such as arousal, both of which complement behavior and are critical for survival.

References and Notes:

1. J. S. Taube, R. U. Muller, J. B. Ranck, Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.* (1990), doi:10.1212/01.wnl.0000299117.48935.2e.
2. S. S. Kim, H. Rouault, S. Druckmann, V. Jayaraman, Ring attractor dynamics in the Drosophila central brain. *Science (80-.).* (2017), doi:10.1126/science.aal4835.
3. J. O'Keefe, J. Dostrovsky, The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* (1971), doi:10.1016/0006-8993(71)90358-1.
4. T. Hafting, M. Fyhn, S. Molden, M. B. Moser, E. I. Moser, Microstructure of a spatial map in the entorhinal cortex. *Nature* (2005), doi:10.1038/nature03721.
5. E. V Evarts, Relation of pyramidal tract activity to force exerted during voluntary movement. *J. Neurophysiol.* (1968), doi:10.1152/jn.1968.31.1.14.
6. C. Stringer *et al.*, Spontaneous behaviors drive multidimensional, brain-wide population activity. *bioRxiv*, 306019 (2018).
7. P. J. Keller, M. B. Ahrens, Visualizing whole-brain activity and development at the single-cell level using light-sheet microscopy. *Neuron*. **85**, 462–483 (2015).
8. T. Schrödel, R. Prevedel, K. Aumayr, M. Zimmer, A. Vaziri, Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light. *Nat. Methods*. **10**, 1013 (2013).
9. C. Dupre, R. Yuste, Non-overlapping Neural Networks in Hydra vulgaris. *Curr. Biol.* (2017), doi:10.1016/j.cub.2017.02.049.
10. S. Kato *et al.*, Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell*. **163**, 656–669 (2015).
11. A. Gordus, N. Pokala, S. Levy, S. W. Flavell, C. I. Bargmann, Feedback from network states generates variability in a probabilistic olfactory circuit. *Cell*. **161**, 215–227 (2015).
12. J. P. Nguyen *et al.*, Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **113**, E1074–E1081 (2016).
13. V. Venkatachalam *et al.*, Pan-neuronal imaging in roaming *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **113**, E1082–E1088 (2016).
14. X. Chen, F. Randi, A. M. Leifer, W. Bialek, Searching for collective behavior in a small brain. *Submitt. to ArXiV* (2018).
15. A. L. A. Nichols, T. Eichler, R. Latham, M. Zimmer, A global brain state underlies *C. elegans* sleep behavior. *Science (80-.).* (2017), doi:10.1126/science.aam6851.
16. G. J. Stephens, B. Johnson-Kerner, W. Bialek, W. S. Ryu, Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS Comput. Biol.* **4**, e1000028 (2008).
17. S. L. Edwards *et al.*, A novel molecular solution for ultraviolet light detection in *Caenorhabditis elegans*. *PLoS Biol.* (2008), doi:10.1371/journal.pbio.0060198.
18. J. Ben Arous, Y. Tanizawa, I. Rabinowitch, D. Chatenay, W. R. Schafer, Automated

imaging of neuronal activity in freely behaving *Caenorhabditis elegans*. *J. Neurosci. Methods*. **187**, 229–234 (2010).

19. B. J. Piggott, J. Liu, Z. Feng, S. A. Wescott, X. Z. S. Xu, The neural circuits and synaptic mechanisms underlying motor initiation in *C. elegans*. *Cell*. **147**, 922–933 (2011).

20. S. Varier, M. Kaiser, Neural development features: Spatio-temporal development of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* (2011), doi:10.1371/journal.pcbi.1001044.

21. J. M. Gray, J. J. Hill, C. I. Bargmann, A circuit for navigation in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **102**, 3184–3191 (2005).

22. M. Chalfie *et al.*, The neural circuit for touch sensitivity in *Caenorhabditis elegans*. *J. Neurosci.* **5**, 956–964 (1985).

23. C. Van Buskirk, P. W. Sternberg, Epidermal growth factor signaling induces behavioral quiescence in *Caenorhabditis elegans*. *Nat. Neurosci.* (2007), doi:10.1038/nn1981.

24. M. Turek, I. Lewandrowski, H. Bringmann, An AP2 transcription factor is required for a sleep-active neuron to induce sleep-like quiescence in *C. elegans*. *Curr. Biol.* (2013), doi:10.1016/j.cub.2013.09.028.

25. G. Jafari, Y. Xie, A. Kullyev, B. Liang, J. Y. Sze, Regulation of Extrasynaptic 5-HT by Serotonin Reuptake Transporter Function in 5-HT-Absorbing Neurons Underscores Adaptation Behavior in *Caenorhabditis elegans*. *J. Neurosci.* (2011), doi:10.1523/JNEUROSCI.1692-11.2011.

26. E. Z. MacOsko *et al.*, A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*. *Nature* (2009), doi:10.1038/nature07886.

27. T. B. Crapse, M. A. Sommer, Corollary discharge across the animal kingdom. *Nat. Rev. Neurosci.* (2008), , doi:10.1038/nrn2457.

28. Q. Wen *et al.*, Proprioceptive Coupling within Motor Neurons Drives *C. elegans* Forward Locomotion. *Neuron* (2012), doi:10.1016/j.neuron.2012.08.039.

29. M. J. McGinley, S. V. David, D. A. McCormick, Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron* (2015), doi:10.1016/j.neuron.2015.05.038.

Acknowledgments: We thank J Pillow, D Tank and C Murphy for helpful discussions. **Funding:** This work was supported by grants from the Simons Foundation (SCGB #324285, and SCGB #543003, AML). This work was supported in part by the National Science Foundation, through the Center for the Physics of Biological Function (PHY-1734030). ANL is supported by a National Institutes of Health institutional training grant NIH T32 MH065214 through the Princeton Neuroscience Institute. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). **Author contributions:** Conceptualization: AML, JWS; Methodology: MS, ANL, FR, XY, AKS; Software: MS; Formal analysis: MS, ANL, FR; Investigation: MS, ANL, FR; Resources: FR, XY, AKS; Writing

(original draft): MS, AML; Writing (editing): MS, AML, JWS; Visualization: MS; Supervision: AML; Project administration: AML; Funding acquisition: AML, JWS; **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data and the code used in this paper are available online. Strains are being deposited at the CGC.

Supplementary Materials:

Materials and Methods

Figures S1-S7

Table S1-S2

Movies S1-S2

References (*1-27*)

Data: DOI 10.17605/OSF.IO/MW5ZS

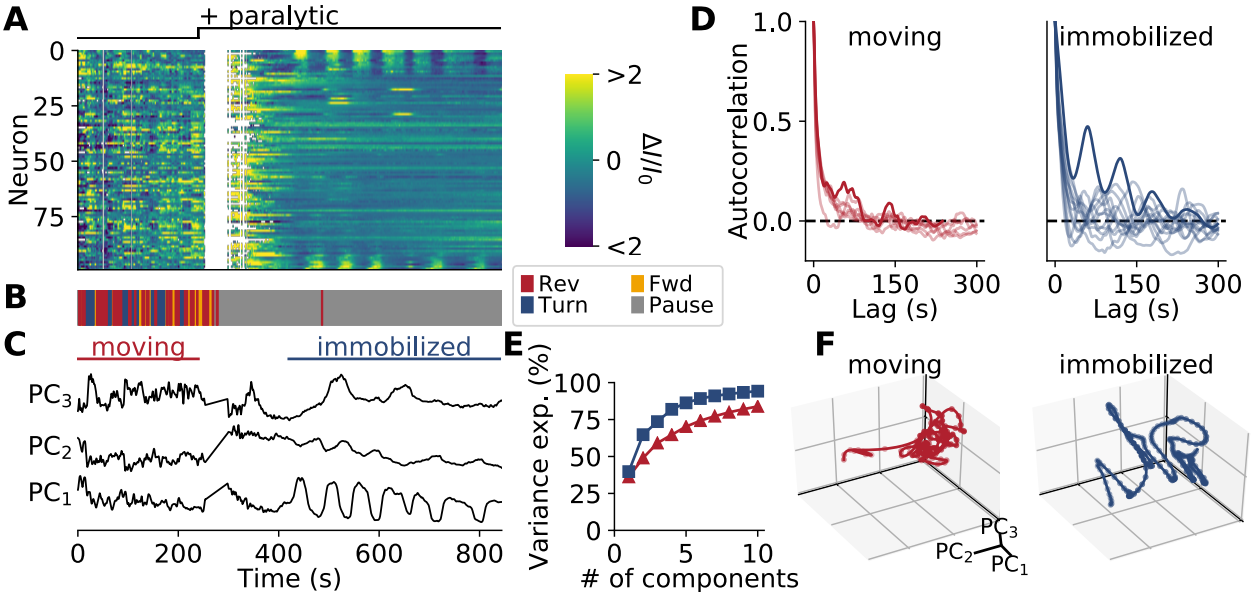

Fig. 1. Neural dynamics slow down when animals are immobilized

(A) Calcium activity of a transiently immobilized *C. elegans*. Calcium activity is recorded from an animal as it moves unrestrained and then is immobilized with tetramisole, a paralytic drug. Calcium activity is the fractional change above baseline of the motion-corrected intensity of GCaMP6s ($\Delta I/I_0$). **(B)** Ethogram shows animal's behavior. **(C)** Neural activity is shown projected onto its first three principal components, as determined by the immobilized portion of the recording. **(D)** Autocorrelation of neural activity, averaged across neurons, is shown for many recordings of either moving or immobilized animals. Thick line shows the recording from (A), split into two parts. Immobilized recordings include animals paralyzed by tetramisole or treated with nanobeads. Neural activity is more periodic and fluctuates more slowly in immobilized animals than in moving animals. **(E)** Cumulative variance explained by the first ten principal components for the moving (red) and immobile part (blue) of the recording in (A). **(F)** Neural state space trajectory for the same recording.

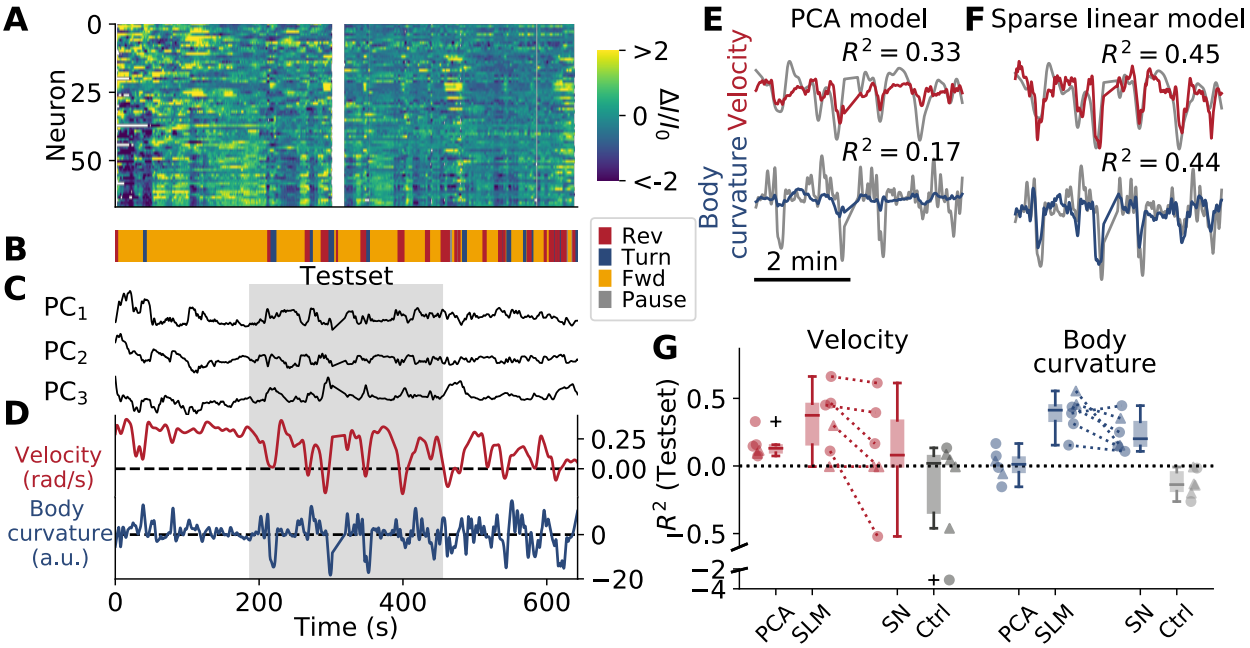

Fig. 2. Behavior is decoded from a population of neurons

(A) Neural activity of a freely moving animal. **(B)** Ethogram. **(C)** Neural activity projected onto its first three principal components (PCs). **(D)** Velocity and body curvature capture describe the animal's locomotion. **(E)** Velocity and body curvature are predicted by a PCA-based linear model 'PCA' that uses projections of neural activity onto three PCs as shown in (C). Prediction is colored. Observed behavior is gray. The model is trained on first and last portion of the recording and tested on the middle portion, labeled 'testset' in (A-D), and shown here. **(F)** Velocity and body curvature is predicted by a sparse linear model (SLM). **(G)** Model performance for multiple animals is shown and compared against the most predictive single neuron (SN), and a similar sparse linear model applied to calcium insensitive GFP control animals, (Ctrl). Box plot shows median. Circles denote wild type animals, triangles denote *lite-1* background animals.

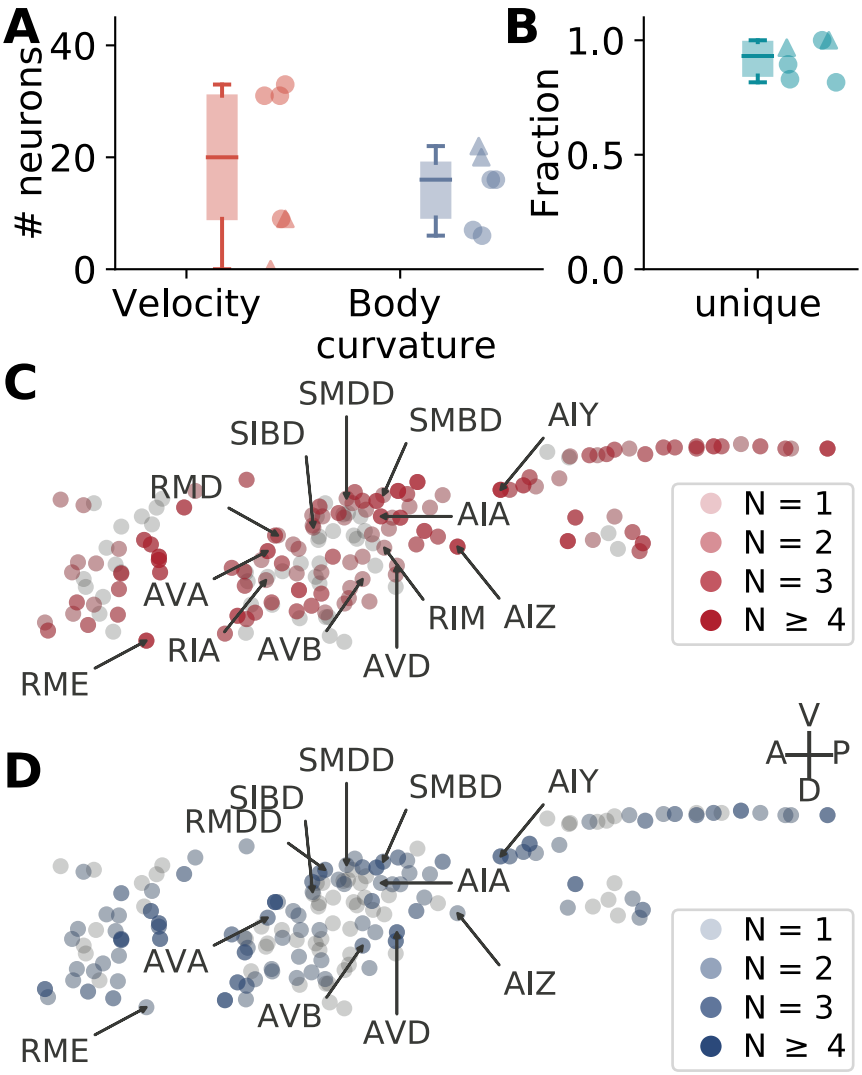

Fig. 3 Neural code for locomotion is spatially distributed throughout the head.

(A) Number of non-zero weighted neurons selected by the sparse linear model for predicting velocity or body curvature. (B) Fraction of those neurons that uniquely occur in only the velocity- or curvature-predicting subsets, but not both. (C) Putative identities of selected velocity and (D) body curvature neurons from 6 recordings are estimated by projecting onto a reference atlas. Shading indicates the number of recordings in which that neuron was present. Labeled neurons are those previously described to have a role in locomotion. For additional neurons see (Fig. S6 and Table S2).

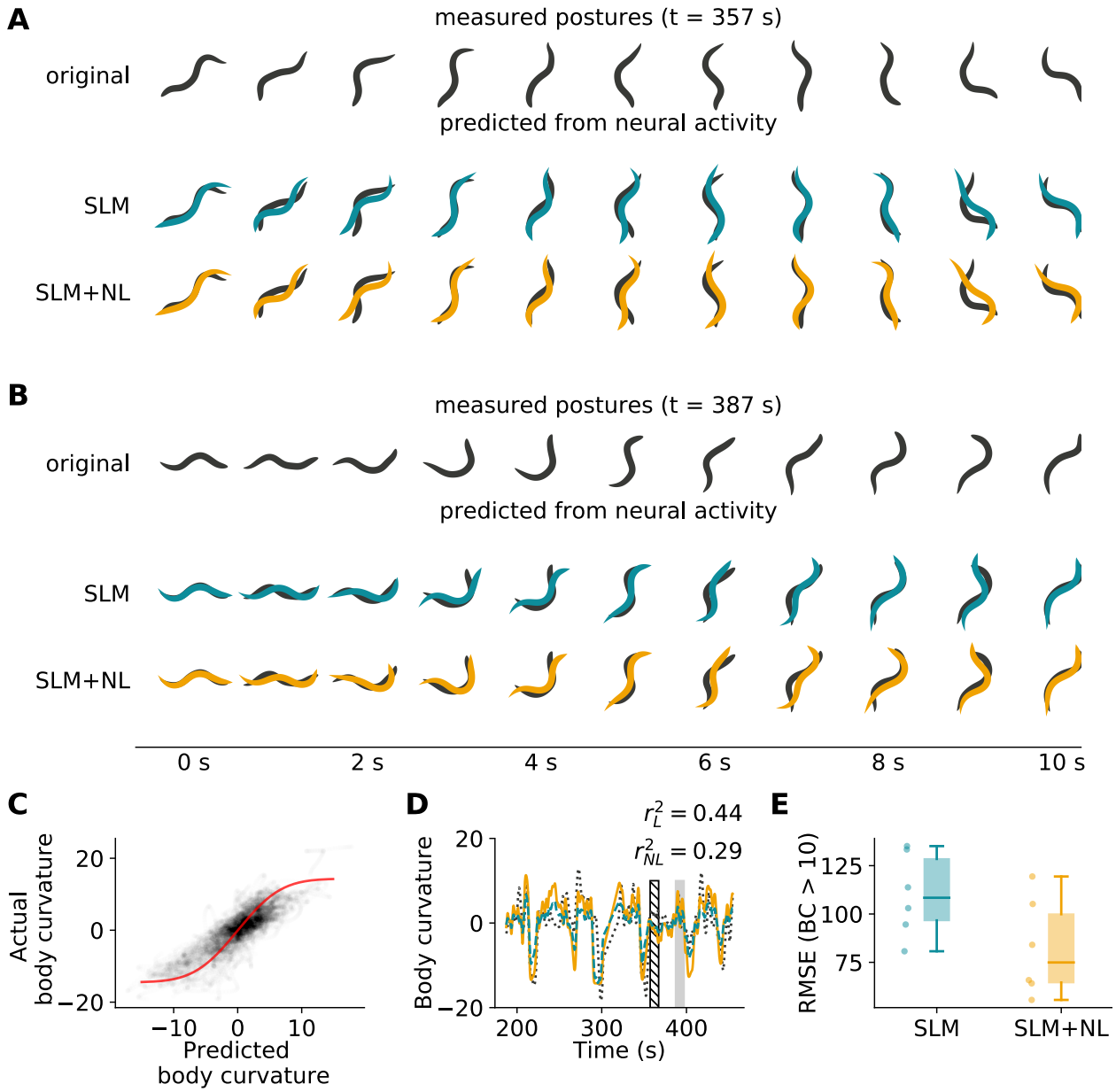

Fig. 4 Posture is inferred from neural activity

(A) Observed and predicted posture of an animal during neural recording. From top to bottom rows show: measured posture (original), the posture predicted using the sparse linear model (SLM), and the posture predicted from the SLM with a nonlinear correction (SLM+NL). The original posture is underlayed in gray for comparison. (B) Posture for a different portion of the same recording. (C) Scatter plot of the predicted (SLM) and true body curvature from all recordings. A nonlinear function (red) was fit to the point cloud. (D) Turn prediction before (SLM, blue) and after (SLM+NL, yellow) applying a nonlinear correction to enhance the turning amplitude. The ground truth is shown as dashed line. The shaded bars indicate the segments

shown in (A) and (B). (E) RMSE between the true posture and the posture predicted by the linear or non-linear corrected model during turns (body curvature amplitude larger than 10).

5

Supplementary Materials for

Predicting natural behavior from whole-brain neural dynamics

Monika Scholz, Ashley N Linder, Francesco Randi, Anuj K Sharma, Xinwei Yu, Joshua W Shaevitz, Andrew M Leifer

Correspondence to: leifer@princeton.edu

This PDF file includes:

Materials and Methods
Supplementary Text
References 1-20
Figs. S1 to S7
Table S1 to S2
Captions for Movies S1 to S2

Other Supplementary Materials for this manuscript include the following:

Movies S1 to S2
Data available at doi:10.17605/OSF.IO/MW5ZS

Materials and Methods

Strains

The following strains were used in this work: Strain AML32 (wtfIs5[P*rab-3*::NLS::GCaMP6s; P*rab-3*::NLS::tagRFP]) expresses the calcium indicator GCaMP6s and a fluorophore tag-RFP in the nuclei of all neurons (*1*). Strain AML18 (wtfIs3[P*rab-3*::NLS::GFP, P*rab-3*::NLS::tagRFP]) expresses the fluorophores GFP and tag-RFP in the nuclei of all neurons (*2*). Strain AML70 (*lite-1(ce314)*X; wtfIs5[P*rab-3*::NLS::GCaMP6s; P*rab-3*::NLS::tagRFP] also expresses GCaMP6s and tag-RFP in the nuclei, but does so in a mutant background that is defective for the animal's endogenous aversive blue-light response. AML70 was generated by backcrossing AML32 to KG1180(*lite-1(ce314)*). Similarly, strain AML175 (*lite-1(ce314)*X; wtfIs3[P*rab-3*::NLS::GFP; P*rab-3*::NLS::tagRFP] expresses GFP and tag-RFP in the nuclei in the light-insensitive mutant background. Strain AML175 was generated by backcrossing AML18 to KG1180(*lite-1(ce314)*). The presence of the *lite-1* mutation was confirmed by sequencing.

Strains were maintained on standard NGM media with *E. coli* OP50 as food source. Animals were imaged between young adult and day 1 adult stage.

Whole brain imaging of freely behaving animals

Whole brain imaging of freely moving animals was performed as described previously (*1*, *2*). Briefly, animals were placed on an imaging plate (a modified NGM media lacking cholesterol and agarose replacing agar) and covered with mineral oil. A coverslip was placed on top of the plate with 100 um plastic spacers between the coverglass and plate surface. The coverslip was fixed to the agarose plate with valap. Animals were imaged on the whole brain imager (*2*), which simultaneously records four video streams to image the calcium activity of the brain of freely moving animals. We record a 10x magnification darkfield image of the body posture, a 10x fluorescence image of the fluorescent brain for tracking, and a 40x image each of tagRFP, and of GCaMP6s (or GFP). The 10x images are recorded at 50 frames/s, and the neural dynamics are recorded with 200 slices/s, with a resulting acquisition rate of 6 volumes/s.

Whole brain imaging animals immobilized via nanobeads

Animals were washed in M9 and picked onto 10% agarose pads with 1 ul of 0.1 um polystyrene beads for immobilization (*3*). A coverslip was placed on top of the agarose and sealed with valap. The animals were recorded and analyzed as described above.

Whole brain imaging of animals paralyzed via tetramisole

For transient immobilization experiments as in Fig 1, animals were immobilized with the paralytic tetramisole during recording. A single animal was loaded into a customized 'artificial dirt' microfluidic chip arena, inspired by (*4*), that enabled the animal to crawl freely in an aqueous environment while precise quantities of liquid were delivered to the arena. A computer-controlled automated liquid delivery system (OB1 and MUX, Elveflow, Paris, France) was used to flow in the buffer and paralytic into the arena. M9 buffer was flowed in for the 4 minutes after which the buffer was switched to 5 uM tetramisole in M9, paralyzing the animal. The buffer switching sequence was pre-programmed and automated using the software provided by Elveflow. The animals were recorded continuously on the whole brain imager. Recording was manually synced to the start of the buffer sequence.

Segmenting and measuring neural and behavior signals from whole brain recordings

The resulting videos of freely moving, immobilized or paralyzed animals were analyzed with the NeRVE Clustering pipeline (*1*). In brief, the movies are synchronized temporally using light flashes. The animals' posture is extracted using an active contour fit on the frames of the 10x behavior recordings. The fit finds 100 (x,y) coordinates for each frame of the behavior recording describing the centerline of the animal. This centerline is computationally mapped onto the high magnification recordings to straighten the point set of detected neuronal nuclei. The neural dynamics are extracted by segmenting the neuronal nuclei in the red channel and straightening the image according to the body posture. Using repeated clustering, neurons are assigned identities over time. The GCaMP signal is extracted using the neural positions over time that resulted from tracking. Occasionally, badly tracked neurons were removed post-analysis during human quality checks. The pipeline returns datasets containing RFP and GCaMP6s (or GFP) fluorescence values for each successfully tracked neuron over time, and the centerline coordinates describing the posture of the animal over time. These are subsequently processed to extract neural activity or behavior features.

Extracting neural activity using independent component analysis

To measure neural activity, we sought to extract changes in fluorescence due to calcium activity and reject artifacts due to animal motion or noise. We pursued a strategy of extracting signals unique to the calcium indicator GCaMP6s and rejecting background signals common to both GCaMP6s and the calcium insensitive fluorophore RFP. We performed independent component analysis (ICA) on the red RFP and green GCaMP6s channels for each neuron independently. ICA has previously been used in neuroscience contexts, for example to identify spikes in intracranial recordings (*5*) or to automatically define regions of interest from large-scale calcium recordings (*6*). Here we use ICA to separate calcium signals from motion artifact. For each neuron, the fluorescence intensity from the red F_{RFP} and green F_{GCaMP} channels are smoothed with a gaussian filter of width 1 s. We then separate signal from noise using the FastICA algorithm based on the implementation described in (*7*) as provided by the python scikit-learn package (*8*). ICA extracts two components, one of which relates to the neural activity signal and the other describes any background that is present in both channels, and thus likely noise. To identify which independent component corresponds to the signal, we pick the component that is most correlated with the green channel, F_{GCaMP}. ICA is insensitive to the sign of the component, and it can converge to components that are inverted. We therefore assign the correct sign to the component by checking if the correlation with the green channel is positive or negative. We then normalize the ICA values by first subtracting the 20th percentile (I_0) and then also dividing by that value. In this work we therefore report calcium signals as $\Delta I/I_0 = \frac{I-I_0}{I_0}$, where $I = \mathrm{ICA}_{\mathrm{Signal}}(F_{GCaMP}, F_{RFP})$.

Compared to a ratiometric approach ($\Delta R/R_0$, $R = F_{GCaMP} / F_{RFP}$) that we had used previously (*2*), we find that the ICA approach improves the quality of extracted signals, and slightly improves the resulting performance of our models. While a ratiometric approach removes much of the noise created by motion, it handles certain types of artefacts poorly and has the disadvantage that independent channel noise is amplified by forming a ratio between the two variables. The ICA approach ultimately performs better, as illustrated in simulations (Fig. S7).

Measuring and representing behavior

To measure the animal's velocity and body curvature we used a PCA-based 'eigenworm' postural analysis (*9*) , as follows. Centerlines of the worms were extracted from the whole-brain imaging recordings using an active contour fit to a brightfield image of the worm as in (*1*). The resulting centerlines were projected onto a 4-dimensional basis set of eigenworms (*9*). To create the eigenworms, centerlines from 135,958 frames of behaving worms imaged in the whole brain imager were extracted. The centerline coordinates were transformed into a set of relative angles between segments and the mean angle was subtracted from each centerline. PCA of the relative angles results in a set of eigenvectors that describe the postural variation of the centerlines. The first four eigenworms explain 96% of the observed variance in the centerlines.

The projection of the centerlines onto the eigenworms results in a timeseries of coefficients, one per eigenworm. Two of the eigenworms describe the body bends that the worm creates during its sinusoidal locomotion. The phase of the body bends is $\Theta(t) = \tan\big(a_1(t)/a_2(t)\big)$, where a_1, a_2 are the coefficients for the first two eigenworms. The derivative of the phase $\dot{\Theta}$ is the phase velocity that describes the speed of bend propagation in the worm and is approximately the animal's center of mass velocity. Per our convention, $\dot{\Theta}$ is referred to as 'velocity' and a positive value of $\dot{\Theta}$ corresponds to the worm moving forward and a negative value of $\dot{\Theta}$ denotes a reversal. Velocity is obtained by filtering Θ with a gaussian derivative filter with width of 3 s. The third eigenworm coefficient $a_3(t)$ corresponds to the body curvature of the animal, which is related to turning. For example, when the animal makes a turn its velocity is positive, and its body curvature is large. The sign of $a_3(t)$ describes the bend direction. $a_3(t)$ is filtered with a gaussian filter of width 1 s.

Principal component analysis (PCA) of neural activity

PCA was performed on the mean-centered and standard deviation scaled data using packages present in the python module scikit-learn (*8*).

Predicting behavior from PC's of neural activity using linear regression

For the PCA model in Fig 2, we use linear regression on the first three neural principal components (PCs) to predict either velocity or turns. The model is trained (i.e., weights are found for each PC) on 60% of the neural activity data. The performance of the model is estimated on the remaining 40% of the data by calculating the coefficient of determination between model prediction and behavior.

Predicting behavior from neural activity using elastic net regularized linear models

For the sparse linear model (SLM), behavior was predicted from neural traces using a linear model with elastic net regularization which applies both l_1 and l_2 weight penalties (*10*).

$$\min_{\beta_0,\beta}\{|y - \beta_0 - X\beta|_2^2\};\ \text{subject to } (1-\alpha)|\beta|_1 + \alpha|\beta|_2^2 \le t, \text{where } \alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

Here, X are the model inputs, in our case the neural activities, y is the model output, in our case the behavior. β_0 and β denote the slope and offset of the linear model. The parameters $\alpha, \lambda_1, \lambda_2$ set the relative strengths of the l_1 and l_2 weight penalties.

The l_1 penalty allows weights to converge to zero, thus resulting in a sparse model. The l_2 penalty makes the fit robust to collinear inputs, by weighing them equally. In neural decoding, elastic net performs better than a simple LASSO regularization, since many of the neural traces are highly correlated.

To fit the sparse linear model, we hold out 40% of the data as test data. We use the remaining data to fit and cross-validate the model. To set the hyperparameter λ_1, we used 5-fold timeseries cross validation. For each fold, the training data is split such that the test data for cross validation is chronologically after the training data. This reduces information 'leakage' between test and training sets. We set the relative strength α of the l_1 and l_2 penalty at 0.95, favoring the l_1 term. After determining the optimal hyperparameter, we assess the prediction quality on the held-out test set (not used in cross-validation).

Assessing model performance in predicting behavior

The quality of the prediction of the PCA and the sparse linear models was assessed by calculating the coefficient of determination between the model and the test data, termed R^2.

$$R^2 \equiv 1 - \frac{SS_{res}}{SS_{tot}},$$

where SS_{res} is the sum of squares of the residuals between model prediction and test data, and SS_{tot} is the sum of squares of the test data. The coefficient of determination can be negative, since the model was trained on unseen data. A model prediction could in principle be anti-correlated with the test data and yield large negative values.

Posture dynamics reconstruction

Our three parameter description of animal posture has a direct correspondence to the predicted velocity and curvature. Because of this we can directly reconstruct an approximation of the animal's full posture from these predicted variables. To compose the posture, each eigenworm is weighted by its coefficients a_1, a_2, a_3 and summed. The SLM directly predicts body curvature, which correspond to a_3, but the undulatory coefficients need to be calculated from the predicted velocity.

We assume that the phase of the body undulations $\Theta(t)$ lies on a circle in the a_1-a_2 plane with radius r, which is a good approximation for moving worms (*9*). To reconstruct the coefficients of the eigenworms a_1, a_2, we can use the predicted velocity and the fact that the coefficients lie on a circle. By integrating the derivative of the phase angle $\dot{\Theta}$, which we called the velocity, we can obtain $\Theta(t)$. We calculate a_1, a_2 using

$$a_1^2 + a_2^2 = r^2$$
$$\tan\Theta = \frac{a_1}{a_2}$$

The reconstruction requires the input of an initial $\Theta(t)$ and the true radius $r(t = 0)$. The following 10 seconds are predicted solely from the SLM using the measured neural activity. In Supplementary Movies S1 and S2, where longer predictions are shown, we repeatedly reset the animal's predicted posture every 10 seconds using the true Θ and r.

Estimating neural identities

To estimate the identities of our recorded neurons with respect to *C. elegans* known neuro anatomical atlas (*11*) we attempted to align and register our recordings. The dataset in Fig 2a was chosen as a reference for all alignments. The 3D positions of the reference dataset were centered around (0,0,0). For each of the remaining datasets the neuronal positions were also centered and registered to the reference with a rigid point set registration followed by a non-rigid point set registration using the python implementation of the coherent point drift algorithm (*12*, *13*). Putative neuronal identities were established by projecting the neuronal positions into two dimensions (reducing the left-right axis) and aligning to a two-dimensional neuronal atlas (*14*). Neurons were assigned the ID of the closest atlas neuron. Neurons were ranked by how many times they were chosen in independent recordings and the 20 highest ranked candidates are shown in Table S2.

Nonlinear enhancement of turn amplitudes

We observed that body curvature predicted by the SLM were smaller in amplitude than the true body curvature. In the posture reconstruction this manifests as a shallow bend during turning, rather than the deep bends observed in experiments. We enhanced body curvature in Fig. 4 by applying an empirically derived non-linearity. The non-linearity was found by fitting all datasets with $\mathrm{turn}_{\mathrm{predicted}} = \mathrm{A}\,\mathrm{erf}\left(\mathrm{turn}_{\mathrm{predicted}}/\mathrm{s}\right) - \mathrm{m}$. To quantify the improvement in posture prediction provided by this non-linearity, calculate the root mean-square error of the predicted posture and the true posture.

Analysis Scripts

Analysis scripts in python are publicly available at https://github.com/monikascholz/PredictionCode.

Datasets

Calcium activity traces, animal posture, and other relevant datasets are publicly available at doi:10.17605/OSF.IO/MW5ZS.

Supplementary Text

Differences in moving and immobile neural dynamics cannot be explained by motion artefact

Recordings of freely behaving animals have the potential to contain motion artefacts and spurious correlations with behavior that would not be present in immobile animals. To test whether the neural dynamics and behavioral encoding we observe can be attributed to motion artefact, we conducted control experiments with animals expressing a calcium insensitive GFP instead of the calcium indicator GCaMP6s. Recordings of GFP control animals contain all of the motion artefacts but none of the calcium signals present in recordings of the GCaMP animals. The following evidence suggests that the behavioral encoding and neural dynamics we observe cannot be explained by motion artefact.

While recordings of freely moving GFP animals did show artefactual fluctuations, those fluctuations did not correlate well with velocity or body curvature. Crucially, GFP control animals performed extremely poorly at predicting behavior, and had essentially no predictive ability (see Fig. 2). This indicates that our predictions of behavior arise from neural signals and not motion artefact.

We further analyzed multiple recordings of freely moving GFP and GCaMP animals using PCA. We compared the number of principal components with significant eigenvalues in the GFP and GCaMP recordings. We find that GFP recordings have a smaller number of significant components than GCaMP recordings (3 and 5, respectively), indicating that there are additional, neural activity-related signals present in the GCaMP recordings that are absent in the GFP recordings (Fig. S3).

Similarities of neural dynamics during immobilization of tetramisole and nanobeads

We employ two common immobilization techniques, microbeads and paralytic drug. The neural dynamics of both appear qualitatively similar in that both show the characteristic slow antiphasic activity (Fig. S3). Others have also observed similarities in the neural dynamics of immobilization using confinement (as in a microfluidic device, presumably similar to nanobeads) compared to using paralytic (*15*). For all these reasons we chose to treat tetramisole and microbead recordings interchangeably as "immobilized." Details of which recordings use which immobilization technique are included in Table S1.

References

1. J. P. Nguyen, A. N. Linder, G. S. Plummer, J. W. Shaevitz, A. M. Leifer, Automatically tracking neurons in a moving and deforming brain. *PLoS Comput. Biol.* **13**, e1005517 (2017).
2. J. P. Nguyen *et al.*, Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. *Proc. Natl. Acad. Sci.* **113**, E1074–E1081 (2016).
3. E. Kim, L. Sun, C. V. Gabel, C. Fang-Yen, Long-Term Imaging of Caenorhabditis elegans Using Nanoparticle-Mediated Immobilization. *PLoS One* (2013), doi:10.1371/journal.pone.0053419.
4. S. R. Lockery *et al.*, Artificial Dirt: Microfluidic Substrates for Nematode Neurobiology and Behavior. *J. Neurophysiol.* (2008), doi:10.1152/jn.91327.2007.
5. K. Kobayashi, I. Merlet, J. Gotman, Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording. *Clin. Neurophysiol.* (2001), doi:10.1016/S1388-2457(01)00457-6.
6. E. A. Mukamel, A. Nimmerjahn, M. J. Schnitzer, Automated Analysis of Cellular Signals from Large-Scale Calcium Imaging Data. *Neuron* (2009), doi:10.1016/j.neuron.2009.08.009.
7. A. Hyvarinen, E. Oja, Independent component analysis: algorithms and applications. *Neural Networks* (2000), doi:10.1016/S0893-6080(00)00026-5.
8. F. Pedregosa *et al.*, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* (2012), doi:10.1007/s13398-014-0173-7.2.
9. G. J. Stephens, B. Johnson-Kerner, W. Bialek, W. S. Ryu, Dimensionality and dynamics in the behavior of C. elegans. *PLoS Comput. Biol.* **4**, e1000028 (2008).
10. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* (2005), doi:10.1111/j.1467-9868.2005.00503.x.
11. J. G. White, E. Southgate, J. N. Thomson, S. Brenner, The Structure of the Nervous System of the Nematode Caenorhabditis elegans. *Philos. Trans. R. Soc. B Biol. Sci.* (1986), doi:10.1098/rstb.1986.0056.

12. A. Myronenko *et al.*, Non-rigid point set registration: Coherent Point Drift. *Adv. Neural Inf. Process. Syst. 19* (2007), doi:10.1109/TPAMI.2010.46.
13. Siavash Khallaghi, PyCPD (2018), (available at http://github.com/siavashk/pycpd/).
14. S. Varier, M. Kaiser, Neural development features: Spatio-temporal development of the Caenorhabditis elegans neuronal network. *PLoS Comput. Biol.* (2011), doi:10.1371/journal.pcbi.1001044.
15. S. Kato *et al.*, Global brain dynamics embed the motor command sequence of Caenorhabditis elegans. *Cell*. **163**, 656–669 (2015).
16. X. Chen, F. Randi, A. M. Leifer, W. Bialek, Searching for collective behavior in a small brain. *Submitt. to ArXiV* (2018).
17. B. J. Piggott, J. Liu, Z. Feng, S. A. Wescott, X. Z. S. Xu, The neural circuits and synaptic mechanisms underlying motor initiation in C. elegans. *Cell*. **147**, 922–933 (2011).
18. J. M. Gray, J. J. Hill, C. I. Bargmann, A circuit for navigation in Caenorhabditis elegans. *Proc. Natl. Acad. Sci.* **102**, 3184–3191 (2005).
19. Z. Li, J. Liu, M. Zheng, X. Z. S. Xu, Encoding of both analog- and digital-like behavioral outputs by one C. Elegans interneuron. *Cell* (2014), doi:10.1016/j.cell.2014.09.056.
20. M. Chalfie *et al.*, The neural circuit for touch sensitivity in Caenorhabditis elegans. *J. Neurosci.* **5**, 956–964 (1985).

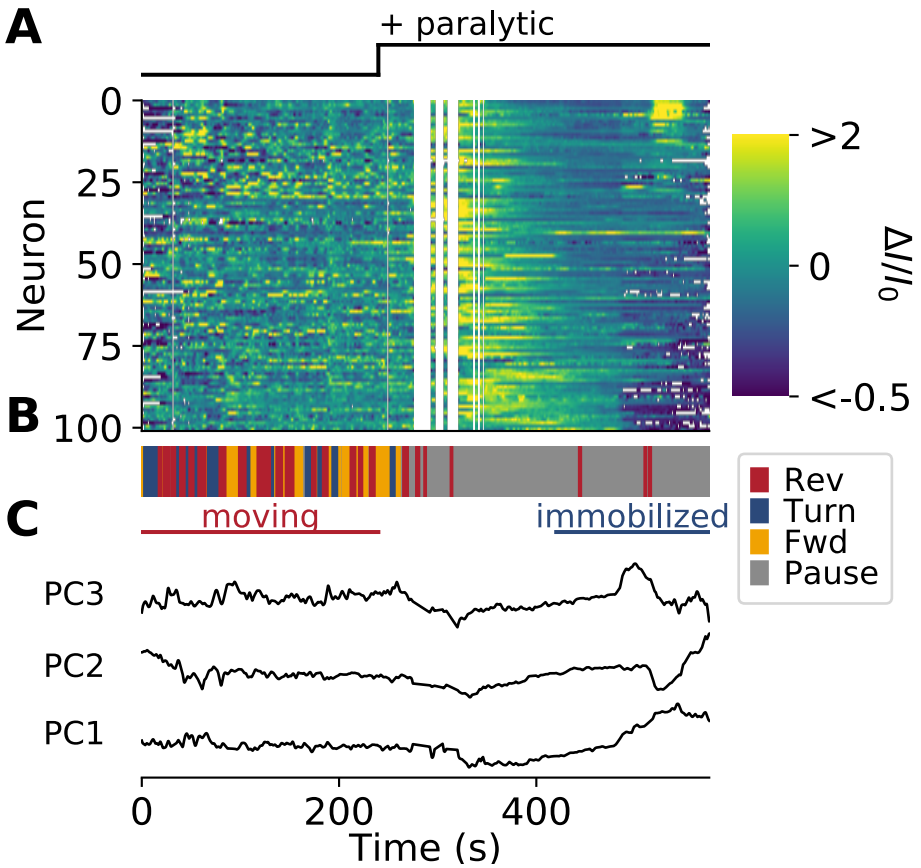

Fig. S1 Additional recording of a transiently paralyzed animal
(A) Calcium activity of a transiently immobilized animal. The animal moves unrestrained in a microfluidic chip for 4 minutes, and then is paralyzed. Calcium activity is the fractional change above baseline of the motion-corrected intensity of GCaMP6s ($\Delta I/I_0$). **(B)** Ethogram of the animal's behavior. **(C)** Neural activity of (A) projected onto the first three principal components (PCs). PCs are determined by the immobilized portion of the recording.

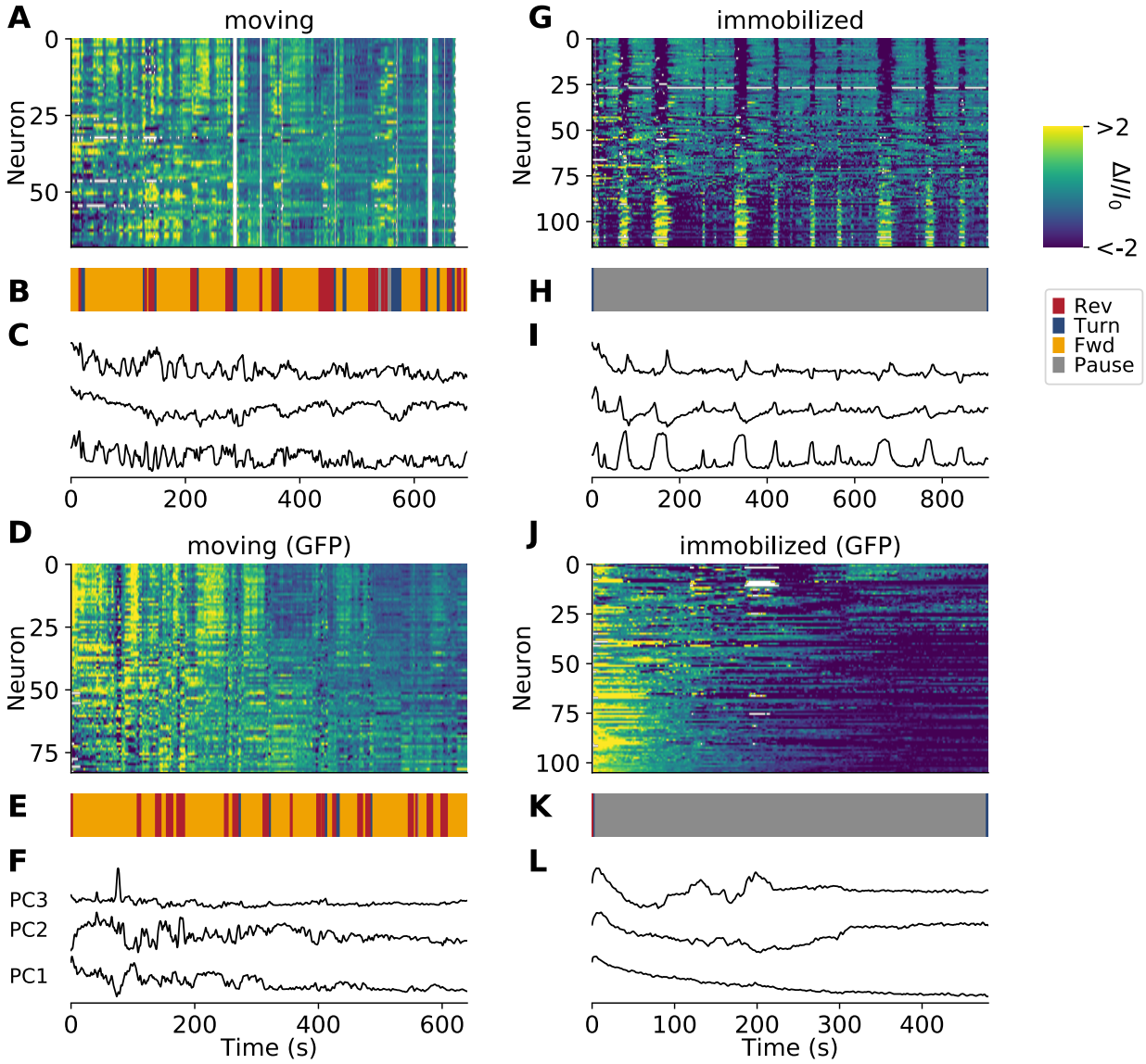

Fig. S2 Examples of recordings from freely moving and immobilized GCaMP6s and GFP-control animals

(A) Heatmap showing the activity of neurons over time for moving animal expressing GCaMP6s. **(B)** Ethogram corresponding to the recordings in (A). **(C)** Neural activity of (A) projected onto the first three principal components (PCs). PCs are determined from the full recording.

(D-F) The same as (A-C) for a freely moving animal expressing GFP.

(G-I) The same as (A-C) for an immobilized animal expressing GCaMP6s.

(J-L) The same as (A-C) for an immobilized animal expressing GFP.

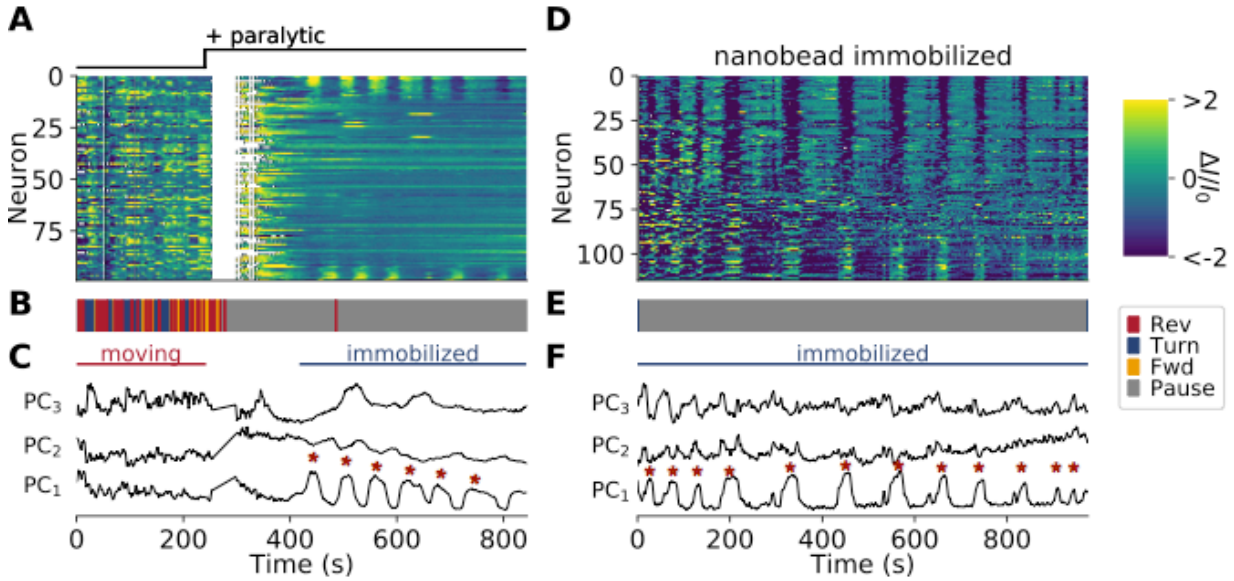

Fig. S3 Immobilization by paralytic and nanobeads result in similar neural dynamics
(A) Calcium activity of moving animal expressing GCaMP6s. Animal is immobilized with the paralytic tetramisole after four minutes of movement. Same recording as in Fig 1. **(B)** Ethogram. **(C)** Neural activity projected onto the first three principal components (PCs) as determined from the full recording. Asterisks denote peaks of slow antiphasic activity present in PC1 during immobilization. **(D-F)** Same as (A-C) for an animal immobilized using nanobeads. Slow antiphasic peaks are similarly present during immobilization via nanobeads.

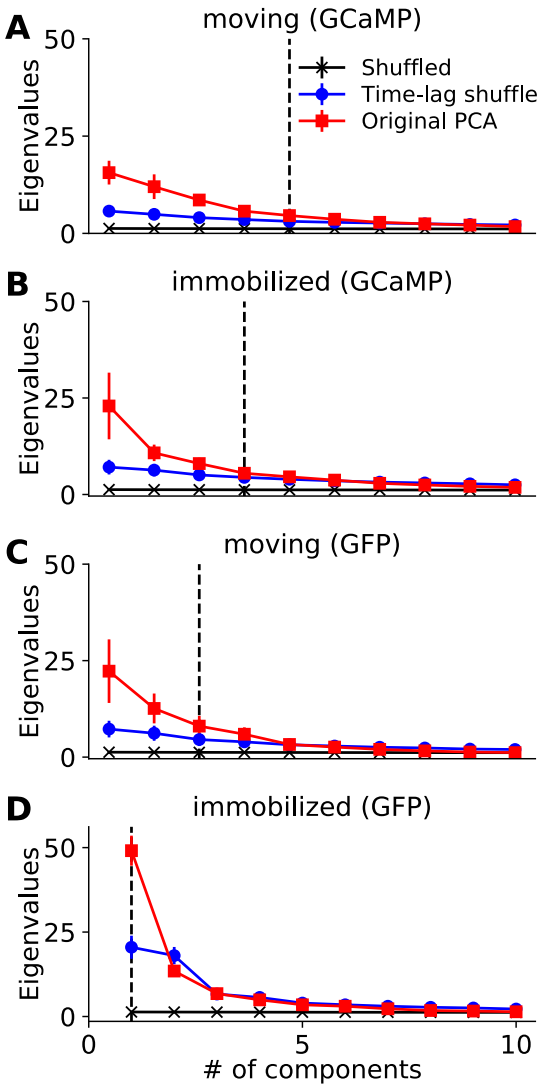

Fig. S4 Number of significant PCs in neural activity during movement and immobilization.

Eigenvalue spectra **from PCA of neural activity**, averaged across multiple recordings, is shown. The vertical dashed line shows an estimate of the number of significant principal components (PCs) in the recording. Specifically, it denotes the number of PCs beyond which the eigenvalues of the neural activity (red) are no longer significantly greater than a time-lag-shuffle in which each neuron's activity is circularly permuted a random amount in time (blue; Welch's t-test, $p<0.05$). The time-lag shuffle destroys correlations between neurons but preserves temporal correlations within each neuron. For comparison, spectra are also shown for a random shuffle (black) in which each neuron's activity is separately shuffled in time so as to also destroy temporal correlations. Eigenvalue spectra are compared for **(A)** immobilized animals expressing the calcium indicator GCaMP6s (N=12), **(B)** moving animals expressing the calcium indicator GCaMP6s (N=6) and for **(C)** immobile animals expressing the calcium insensitive GFP control (N=3) **(D)** moving animals expressing the calcium insensitive GFP control (N=6). Of the four

conditions, moving animals expressing calcium indicator have the highest estimated number of significant PC's.

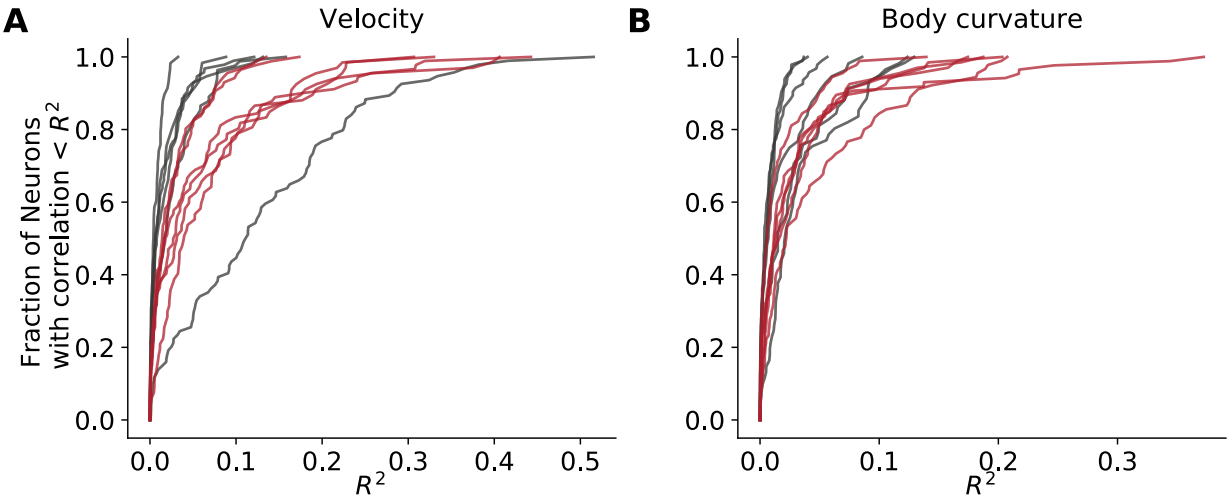

Fig. S5 Individual neurons correlate with behavior
(A) Cumulative distribution of neurons correlating with velocity for GCaMP6s (red, N=6) and GFP control recordings (gray, N=6) of freely moving animals. Each line is a separate recording.
(B) Cumulative distribution of neurons correlating with turns.

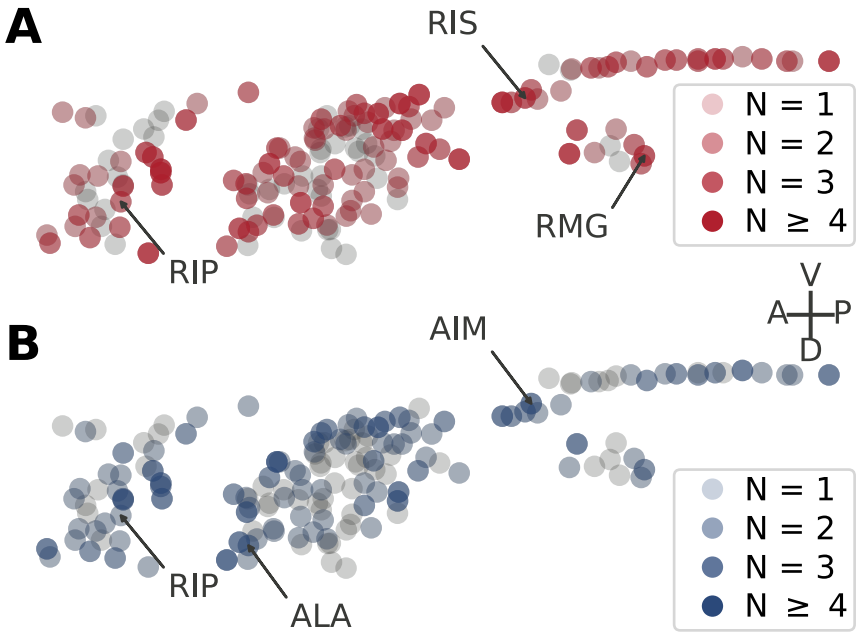

Fig. S6 Selection of unexpected neurons found to encode velocity or body curvature

(A) Identities of velocity and **(B)** body curvature neurons from 6 recordings are estimated by projecting onto a reference atlas. Shading indicates the number of recordings in which that neuron was present. The annotated neurons are not typically associated with locomotion but were nonetheless assigned non-zero weights in the sparse linear model. AIM, ALA, RIS and RMG neurons have known function in quiescence, foraging and as chemosensory hub. However, a direct connection to locomotion encoding was unknown. The role of RIP has been less well-characterized.

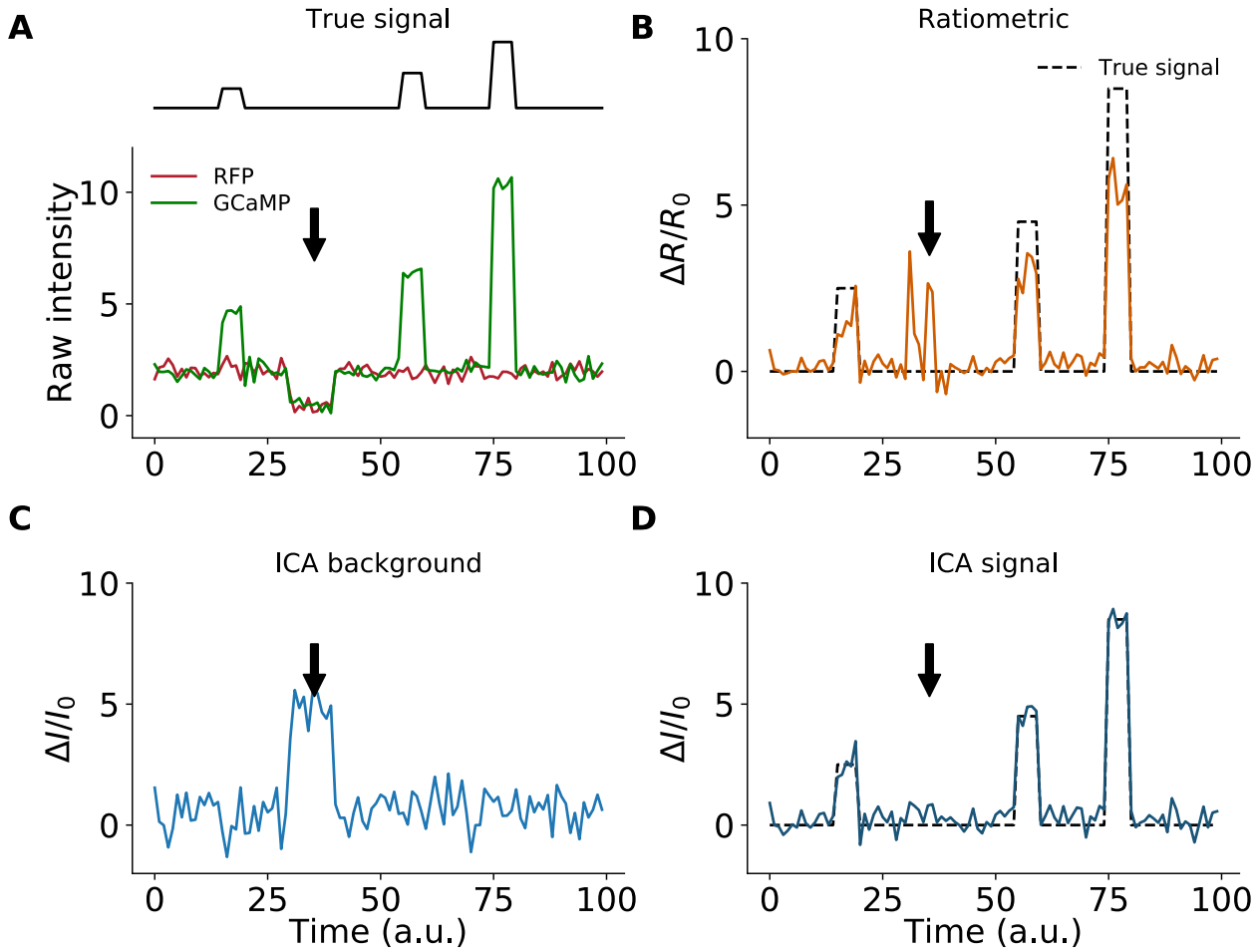

Fig. S7 Simulation of motion artefact using ratiometric and independent component analysis

(A) Simulated fluorescence signals for two color channels reflect a true underlying calcium signal (black, dashed). The GCaMP channel (green trace) contains signal, motion artefacts (arrow) and gaussian noise. The RFP channel (red trace) contains only background motion artefact and gaussian noise. **(B)** Signal recovered through ratiometric analysis. The amplitude of the recovered signal (blue) does not agree with the true signal (black, dashed) **(C)** Background obtained from independent component analysis (ICA) matches the motion artefact background inserted in (A). **(D)** Signal extracted using ICA (blue) more closely matches the actual signal (black, dashed), as compared to the ratiometric approach.

Table S1. Summary of recordings

Details of all recordings used in this paper are described below. Unique identifier can be used to locate the corresponding publicly accessible dataset at doi:10.17605/OSF.IO/MW5ZS. The two recordings indicated with "*" were previously reported in (*16*).

Condition	Strain	Unique Identifier	Fluorophore	Background	Arena	Duration (min)	Number of Neurons	Figure Panels
transiently paralyzed with tetramisole	AML70	BrainScanner20180329_152141	GCaMP6s	*lite-1*	microfluidic	8	101	Fig.S1
	AML32	BrainScanner20180511_134913	GCaMP6s	wt	microfluidic	13	99	Fig.1, S3
Immobilized with nanobeads	AML32*	BrainScanner20171017_170419	GCaMP6s	wt	agarose pad	13	85	Fig1D, S4
	AML32*	BrainScanner20171017_171956	GCaMP6s	wt	agarose pad	10	70	Fig1D, S4
	AML32	BrainScanner20171017_184114	GCaMP6s	wt	agarose pad	16	72	Fig1D, S4
	AML32	BrainScanner20171212_160415	GCaMP6s	wt	agarose pad	10	80	Fig1D, S4
	AML32	BrainScanner20180510_090158	GCaMP6s	wt	agarose pad	16	115	Fig1D, S4
	AML32	BrainScanner20180510_092218	GCaMP6s	wt	agarose pad	15	114	Fig1D, S4, S2
	AML70	BrainScanner20180221_150819	GCaMP6s	*lite-1*	agarose pad	11	118	Fig1D, S4
	AML70	BrainScanner20180221_152059	GCaMP6s	*lite-1*	agarose pad	10	71	Fig1D, S4
	AML70	BrainScanner20180221_154553	GCaMP6s	*lite-1*	agarose pad	12	83	Fig1D, S4
	AML70	BrainScanner20180323_140814	GCaMP6s	*lite-1*	agarose pad	11	74	Fig1D, S4
	AML70	BrainScanner20180323_143102	GCaMP6s	*lite-1*	agarose pad	6	81	Fig1D, S4
	AML70	BrainScanner20180323_144442	GCaMP6s	*lite-1*	agarose pad	7	84	Fig1D, S4
	AML18	BrainScanner20180518_091402	GFP	wt	agarose pad	11	111	Fig.S4

Condition	Strain	Unique Identifier	Fluorophore	Background	Arena	Duration (min)	Number of Neurons	Figure Panels
	AML18	BrainScanner20180518_093125	GFP	wt	agarose pad	8	108	Fig.S4
	AML18	BrainScanner20180518_094052	GFP	wt	agarose pad	8	105	Fig.S4, S2
freely moving	AML18	BrainScanner20160506_155051	GFP	wt	imaging plate	9	95	Fig.3, S4
	AML18	BrainScanner20160506_160928	GFP	wt	imaging plate	10	83	Fig.3, S4, S2
	AML175	BrainScanner20180223_141721	GFP	*lite-1*	imaging plate	6	62	Fig.3, S4, S5
	AML175	BrainScanner20180223_142554	GFP	*lite-1*	imaging plate	10	59	Fig.3, S4, S5
	AML175	BrainScanner20180330_160650	GFP	*lite-1*	imaging plate	9	57	Fig.3, S4, S5
	AML175	BrainScanner20180330_162137	GFP	*lite-1*	imaging plate	11	95	Fig.3, S4, S5
	AML70	BrainScanner20180327_152059	GCaMP6s	*lite-1*	microfluidic	12	87	Fig. 1D, 2E-G, 3, 4C,E, S4, S5, S6
	AML70	BrainScanner20180430_141614	GCaMP6s	*lite-1*	microfluidic	9	91	Fig. 1D, 2E-G, 3, 4C,E, S4, S5, S6
	AML32	BrainScanner20170424_105620	GCaMP6s	wt	imaging plate	11	68	Fig. 1D, 2E-G, 3, 4, S4,S5, S4, S6
	AML32	BrainScanner20170610_105634	GCaMP6s	wt	imaging plate	10	68	Fig. 1D, 2E-G, 3, 4C,E, S4, S5, S6
	AML32	BrainScanner20170613_134800	GCaMP6s	wt	imaging plate	10	67	Fig. 1D, 2, 3, 4, S4, S5, S6
	AML32	BrainScanner20180709_100433	GCaMP6s	wt	imaging plate	10	91	Fig. 1D, 2E-G, 3, 4C,E, S4, S5, S6

Table S2. Neurons most often found to be predictive of behavior across recordings
Identities of neurons that were reliably found to be predictive of velocity or body curvature. Neurons are ranked by how often they appear across recordings among the non-zero weighted neurons found by the sparse linear decoder model. The top ranked 20 neurons for each behavior are shown. Neurons already known to have signals relating to velocity or turning frequency (which is relevant for body curvature) are highlighted in red (*17–20*). Neurons with a previously described function unrelated to locomotion, shown in Fig S5, are highlighted in blue.

Velocity neurons	Number of recordings	Turn neurons	Number of recordings
AIYR	5	RMDR	5
AIZL	4	RID	4
BAGL/R	4/3	AIML	3
FLPR/L	4/3	AINR	3
RMED/R/L	4/4/3	AIYR	3
SIADR	4	AVDL	3
SMBDR/VR	4/3	BAGR/L	3/3
VD2	4	CEPDR	3
AIAL	3	DB1	3
AVAL	3	FLPL	3
AVL	3	IL2DR	3
AWBR	3	RIPR	2
CEPDR/VL	3/3	RME	2
OLQVL	3	SAA	2
RICL	3	SMBVR	2
RIGL	3	AVAL	2
RIPR	3	AVBL	2
RIS	3	AIMR	2
RMGL	3	AIZR	2
		ALA	2

Movie S1. Worm posture predicted from neural activity with a sparse linear model

https://vimeo.com/295461323

The video shows an animal's posture predicted from its neural activity using the Sparse Linear Model (right) alongside the animal's true posture (left) for comparison. The gray track shows the predicted and true center of mass location of the animal, respectively. The predicted posture is initialized with the animal's true posture at the beginning of the recording and updated every 10 seconds with information from the animal's true posture so as to correct for accumulating errors. The timestamp flashes red to indicate each update.

Movie S2. Worm posture predicted from neural activity with a PCA model

https://vimeo.com/295462311

The video shows an animal's posture predicted from its neural activity using the PCA mode (right) alongside the animal's true posture (left) for comparison. The PCA model performs more poorly than the Sparse Linear Model shown in Movie S1. The gray track shows the predicted and true center of mass location of the animal, respectively. The predicted posture is initialized with the animal's true posture at the beginning of the recording and updated every 10 seconds with information from the animal's true posture so as to correct for accumulating errors. The timestamp flashes red to indicate each update.