Supporting Online Material - Material and Methods

Preparation. Animals were adult *Hirudo medicinalis* (2-5 gm), the European medicinal leech, maintained in artificial pond water at 15°C. To reliably evoke swimming and crawling, all experiments were performed between the months of December – June. We dissected out the full nerve cord, including the head brain, all 21 segmental ganglia, and the tail brain. We removed the blood sinus around the ganglion to be imaged, between G7 and G10. We dissected away the connective tissue sheath from the ventral surface of the exposed ganglion. We prepared at least two dorsal posterior (DP) nerves, chosen from between G13 and G16, for extracellular recording and stimulation. A motor neuron in this nerve bursts during both the dorsal contractile phase of swimming (at ~1 Hz) and the contractile phase of crawling (at ~0.05 Hz), so the nerve serves as a monitor of which locomotory behavior is elicited (*13, 14*). The purpose of recording with two suction electrodes was to confirm the propagation of behaviors down the nerve cord. One of these nerves was stimulated with 10 ms electrical pulses delivered at 15Hz, with an amplitude of 2-3 V. The duration of the stimulus train was adjusted empirically to generate crawling and swimming responses; they were generally 300 +/- 100 ms long. In all experiments, we stabilized the imaged ganglion by pinning small latex strips across the adjacent nerve cord to minimize motion artifact (*25*). The preparation was maintained in a chamber filled with ~10 ml of room temperature leech saline, consisting of (in mM): 115 NaCl, 4KCl, 1.8$CaCl_2$, 1.5 $MgCl_2$, 10 dextrose, 4.6 Tris maleate, and 5.4 Tris base, pH 7.4.

Staining with FRET dyes. We first stained the ganglion to be imaged with the FRET donor, *N*-(6-chloro-7-hydroxycoumarin-3-carbonyl)-

dimyristoylphosphatidylethanolamine (CC2-DMPE), a coumarin-labeled phospholipid. (*24*). We made a 30 µM staining solution from 3 µl of 10 mM CC2-DMPE in DMSO, 1.5 ul of 20 mg/ml pluronic F-127 in DMSO, and 1 ml of saline. This solution was then serially diluted to a 10 µM final staining solution. We pinned out the nerve cord in a Sylgard-coated dish, placed a small plastic cylinder over the exposed ganglion, and sealed it with petroleum jelly to make a watertight chamber. We then replaced the saline in the chamber with the staining solution and stained the ganglion for 30 min. During the staining, we constantly recirculated the staining solution using a peristaltic pump. After 30 min, the staining solution was removed and the tissue was washed several times with fresh saline.

We next stained the whole nerve cord with the FRET acceptor, bis(1,3-diethyl-thiobarbiturate)-trimethine oxonol [DiSBAC$_2$(3)], an oxonol dye (*23*). We made a 15 µM staining solution from 9.6 µl of 15.67 mM DiSBAC$_2$(3) in DMSO and 10 ml saline. This solution was then bath-sonicated for at least 3 min. We replaced the saline with this solution and left it on for 30 min. At the end of 30 min., the oxonol solution quenched the coumarin emission by approximately 50%. We then removed the oxonol solution and replaced the bath with fresh saline.

Electrophysiology and cell fills. We recorded intracellularly from cells using 40-60 MΩ glass microelectrodes filled with 1M potassium acetate, using an AxoClamp 2A amplifier. For cell 208, we measured membrane resistances of approximately 50 MΩ, meaning that our current injections resulted in membrane potential changes of about +/-75 mV from rest. We recorded extracellularly using suction electrodes and a four-

channel differential amplifier. We digitized all electrical data at 1 kHz using a 16-bit analog-to-digital board and custom Matlab software.

We filled individual candidate decision-making neurons using Alexa 488 dextran (10,000 MW). Microelectrodes were back-filled using concentrations of 25 mg/ml Alexa 488. Cells were then impaled, and dye was iontophoretically injected. Dye was allowed to diffuse for ~ 1 hr, and tissue was fixed overnight, dehydrated, cleared, and mounted on a slide for viewing. Fills were imaged using a confocal microscope. The dye was excited by the 488 nm emission line of a KrAr laser. Candidate neurons were compared, based on morphology and electrophysiological properties, to previously identified leech neurons.

Optical recording. We acquired fluorescence images using an upright microscope. We used a 20x, 0.5 numerical aperture (NA), water-immersion objective. For epi-illumination we used a tungsten halogen lamp in a standard housing, powered by a low-ripple power supply. We filtered the excitation light at the coumarin excitation band with a 405 ± 15 nm bandpass filter. Emitted light was filtered with a custom designed wavelength switching device. This device split the emitted light into two optical paths with a 430 nm dichroic mirror. The light was filtered in each path at either the coumarin emission wavelength (460 ± 25 nm) or the oxonol emission wavelength (560LP). Optical shutters in each path were alternately opened or closed, allowing only one emission band to pass per frame. Following the shutters, the two paths were recombined (430 nm dichroic) and imaged onto a CCD camera. We used a cooled CCD camera operated in frame-transfer mode to acquire the optical data at a resolution of 256 x 256 pixels. Images were acquired at a frame rate of 40 Hz. Since every other frame captured one of the two

emission wavelengths, the effective frame rate was 20 Hz. Imaging data were acquired using custom software written in Matlab. We synchronized the optical and electrical recordings by feeding the frame timing signals emitted by the camera into the A-to-D board, along with all of the electrophysiology signals.

Analysis. In a typical experiment, we imaged 10 – 20 trials, performed the following analysis and then intracellularly impaled candidate decision-making neurons. After acquiring the data, we analyzed them using Matlab. We outlined the images of individual somata manually using a custom-made graphic user interface. All pixels within each cellular outline were then averaged in each frame, yielding two raw fluorescence signals for each cell, one for the coumarin emission wavelength ($F_{CC2}(t)$) and one for the oxonol emission wavelength ($F_{OX}(t)$). The noise in these single-cell signals was shot-noise dominated. The raw fluorescence signals were divided, $F_{CC2}(t) / F_{OX}(t)$, to yield a ratioed fluorescence signal $F_R(t)$. Ratioing improved the signal-to-noise ratio of the signals and corrected both bleaching and motion artifacts. Signals were then normalized by the average baseline fluorescence before a stimulus was delivered (the first second of each trial) and expressed as a percent change in fluorescence (dF/F) from this baseline. We measured sensitivities of $F_R(t)$ in the range of 15-20%/100 mV for 1 Hz square-wave voltage signals with a 10 mV amplitude, centered around a baseline voltage of -55 mV. Principal component analysis (PCA) and linear discriminant analysis (LDA) were performed using Matlab toolboxes and custom code. Prior to performing any analysis, each trial was classified manually as either a swim or crawl trial. We were aware of the possibility of behavioral misclassification, but the difference between the motor pattern

for swimming and crawling is very obvious. The periods of the motor bursts for the two behaviors do not overlap.

ANOVA. For the single neuron analysis, we performed a sliding window ANOVA using a 300 ms bin width. The earliest discrimination time for each neuron occurred when the p value first fell below a threshold ($p < 10^{-6}$). We chose a small p value threshold to increase our confidence in the discrimination time. We performed this same analysis on rectified, low-pass filtered (300 ms time constant) versions of the DP nerve recordings to determine the nerve discrimination time.

Principal Component Analysis. For the PCA, we organized the optical data from an experiment into an [MP] x N matrix, where M is the number of trials, P is the number of frames per trial, and N is the number of neurons imaged. Each column of this matrix was normalized by the standard deviation of the data in that column. To visualize the data in a 3-dimensional principal component space we deconcatenated each trial from the PCA score matrix and plotted each trial separately. See additional supplemental materials for more details.

Linear Discriminant Analysis. A LDA was performed at each time point in the principal component space using a sliding window of variable width and variable number of PC dimensions. We chose the linear discriminant that resulted in the earliest discrimination time (see Fig. S1 for the procedure). See additional supplemental materials for more details.

Supporting Online Material – Supplemental Figures

Figure S1

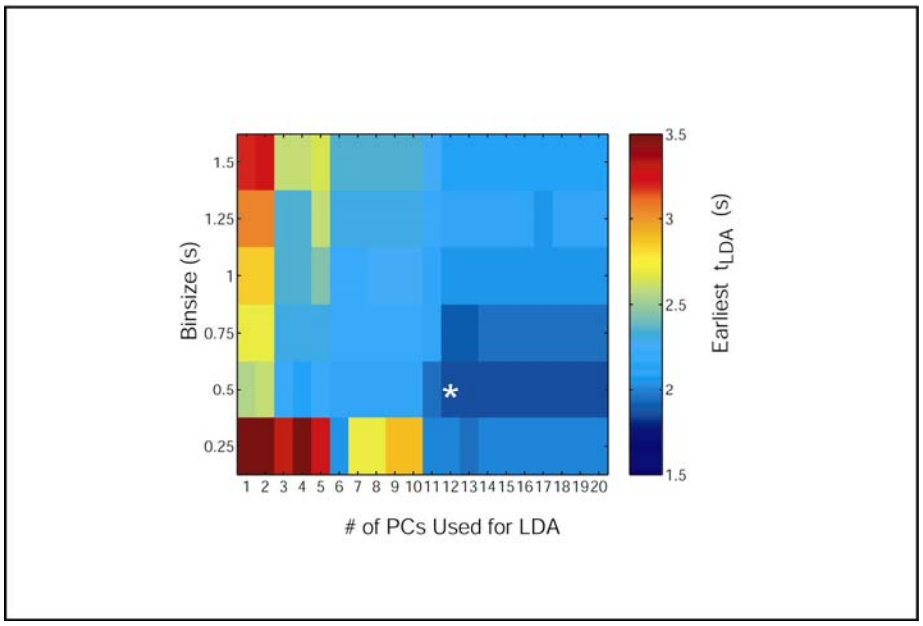


Figure S2

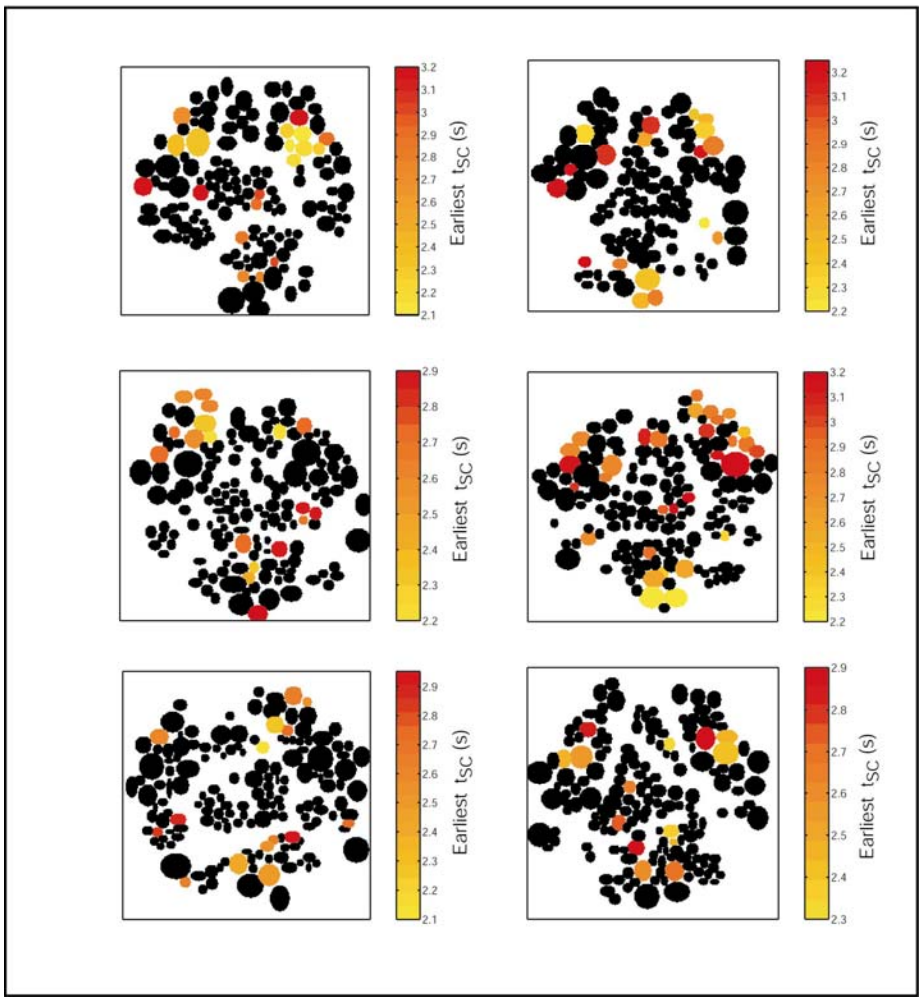
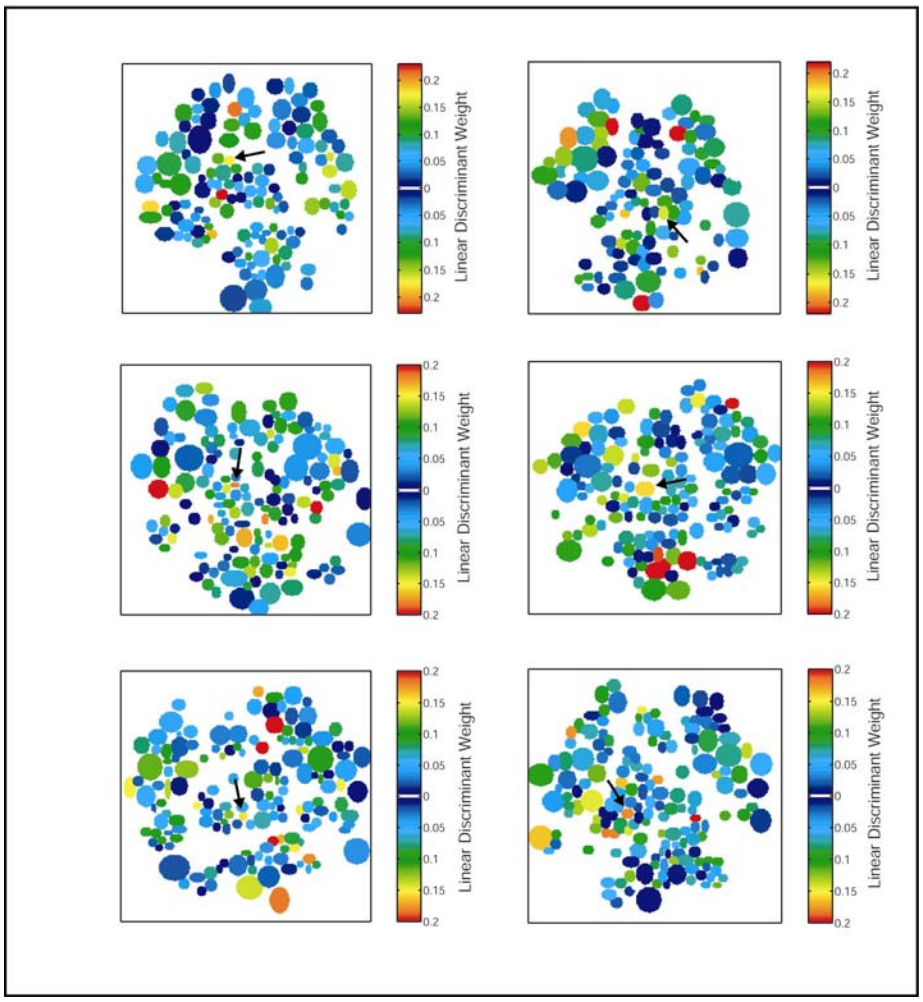

Earliest t_{SC} (s)
3.2
3.1
3
2.9
2.8
2.7
2.6
2.5
2.4
2.3
2.2
2.1
Earliest t_{SC} (s)
3.2
3.1
3
2.9
2.8
2.7
2.6
2.5
2.4
2.3
2.2
Earliest t_{SC} (s)
2.9
2.8
2.7
2.6
2.5
2.4
2.3
2.2
Earliest t_{SC} (s)
3.2
3.1
3
2.9
2.8
2.7
2.6
2.5
2.4
2.3
2.2
Earliest t_{SC} (s)
2.9
2.8
2.7
2.6
2.5
2.4
2.3
2.2
2.1
Earliest t_{SC} (s)
2.9
2.8
2.7
2.6
2.5
2.4
2.3

Figure S3



Supporting Online Material - Supplement Figure Legends

Fig S1. Linear discriminant analysis discrimination time. Linear discriminants were estimated by moving a sliding window along the data in the PC space. We varied the time bin width and number of PCs. At each time bin, we performed an ANOVA of the grouped data within the time bin, projected onto the linear discriminant, and derived a p value. The earliest time at which the p value became significant ($p < 10^{-6}$) is termed t_{LDA} and plotted for each combination of bin width and number of PCs. We used the linear discriminant with the earliest t_{LDA} for subsequent analysis. For the experiment shown, the linear discriminant was estimated using a 500 ms window and 12 PCs (white asterisk) as this combination led to the earliest t_{LDA}. We randomly shuffled trials to confirm that using up to 20 PCs with at least 10 trials per experiment did not lead to overfitting (i.e., the shuffled data projected onto the estimated linear discriminants was not significantly separable into two groups at any point in time).

Fig S2. Single cell discrimination ganglion maps. The ganglion maps from 6 experiments are shown. Neurons with single cell discrimination times (t_{SC}) that occurred prior to t_{NERVE} for each experiment are colored (mean +/- SD = 17 +/- 6 cells for 6 experiments). Yellow indicates early discrimination times and red indicates later discrimination times.

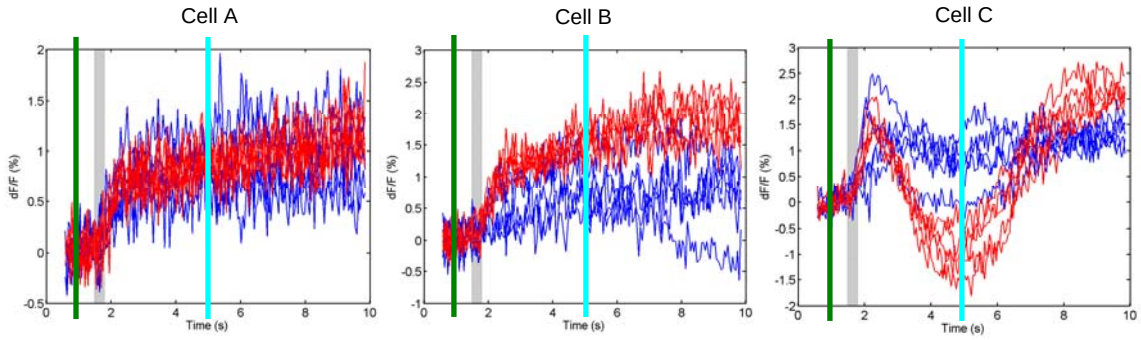
Fig. S3. Linear discriminant analysis ganglion maps. The ganglion maps from 6 experiments are shown. The maps are from the same experiments as in Fig. S2. Cells are color-coded based on the magnitude of the contribution to the linear discriminant direction. Red and yellow represent large magnitude contributions, blue represents small contributions. The black arrows indicate cell 208 in each experiment.

Supporting Online Material – LDA & PCA

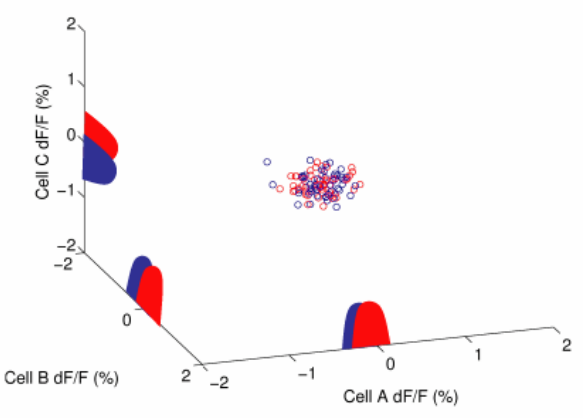
For readers unfamiliar with Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), we wish to demonstrate a hopefully intuitive example. As described in the paper, our datasets are N-dimensional, where N is the number of neurons from which we record. First, we will show an example in which N = 3 to demonstrate LDA. Then we will show a 2D example of PCA. The goal is to develop an intuitive feel for what LDA and PCA do in a low dimensional example. We hope the reader can then imagine how the technique can be extended to higher dimensional datasets that are more difficult to visualize.

LDA

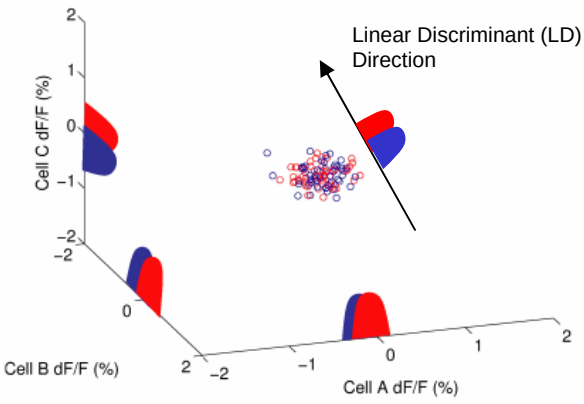
We begin by examining the raw signals from three neurons measured across 10 trials; 5 swimming trials (blue traces) and 5 crawling trials (red traces).



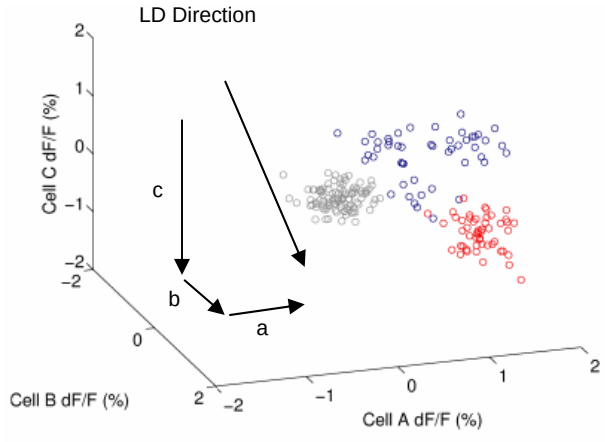
We are interested in when the trajectories for swimming and crawling trials have diverged in the 3D joint space of these three neurons. The 3D space is just a scatter plot of the optical data points from the neurons. First, we look at the data in a 500 ms window prior to the stimulus (highlighted by the green bars in the 3 plots above).



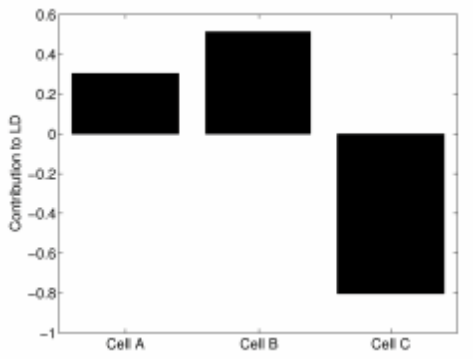
The distributions of swimming versus crawling points are shown along each axis. Since the window we chose is prior to the stimulus, none of these distributions are significantly different. The idea of the LDA is to find a line such that, when the data are projected onto it, the distributions are maximally separated. For example:



Again, the distributions along this line are not significantly different at this point in time. However, if we pick a later time window (the cyan colored bars in the 3 raw data plots), the behaviors are clearly distinguishable:



The linear discriminant (LD) direction tells us the optimal linear combination of the neurons to achieve maximal discrimination. This linear combination is defined by the slope (vectors a, b, and c above) of the LD:

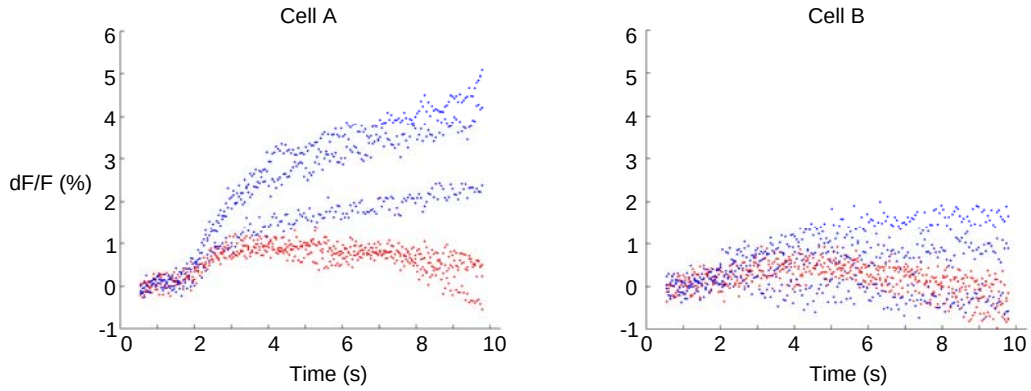


Since cell C does a good job of discriminating the two behaviors, it has a relatively large magnitude contribution to the slope of the LD. In contrast, cell A does not discriminate well, so it has a smaller contribution. Our basic technique then is to slide a window in time and estimate a LD at each time. We then perform an ANOVA on the data projected onto each line and ask when the distributions become significantly different. One of the main findings from the paper is that linear combinations of neurons can separate the swimming and crawling distributions earlier than single neurons.

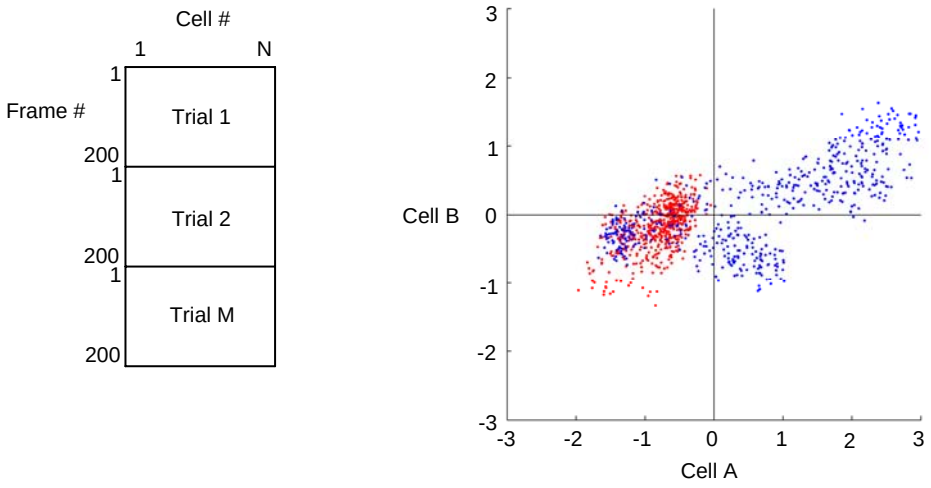
PCA

The LDA described above works well in low-dimensional spaces, but is susceptible to overfitting. The problem is that as more dimensions (neurons) are taken into consideration, more data points are needed to adequately sample the higher dimensional space. Since we were limited in the amount of data we could collect, we used PCA to reduce the dimensionality of our datasets prior to the LDA.

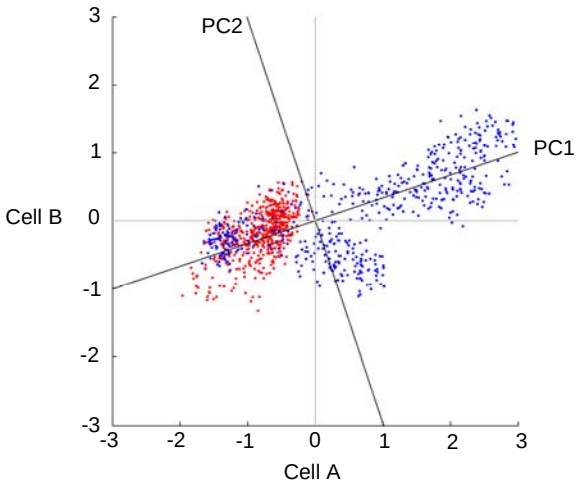
For this example, we will consider two neurons measured across 6 trials; 3 swimming trials (blue traces) and 3 crawling trials (red traces):



We organize the data into a matrix in which each column represents a neuron and the rows contain the time points for each trial. As mentioned above, these 2 neurons are considered to be dimensions (or axes) in a 2-D space. A scatter plot of the columns of the data matrix demonstrates this.

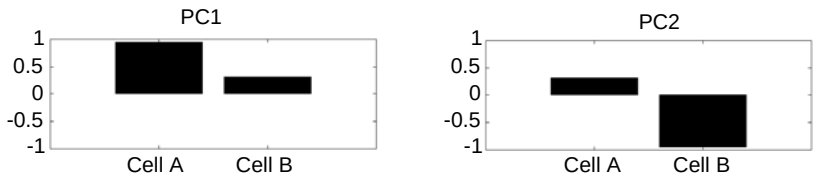


Notice that the activity of the 2 neurons is correlated. If the 2 neurons were uncorrelated, one would expect a circular cloud of data. PCA rotates the original axes to point in new directions. It chooses the new directions to lie along the directions of maximal covariance. These new directions are called the principal components (PCs). By rotating the original axes, it is hoped that the new PCs will provide a more parsimonious description of the data (i.e., we will ignore directions that do not contribute much to explaining the variance). The rotation that PCA found is shown by the two directions PC1 and PC2.

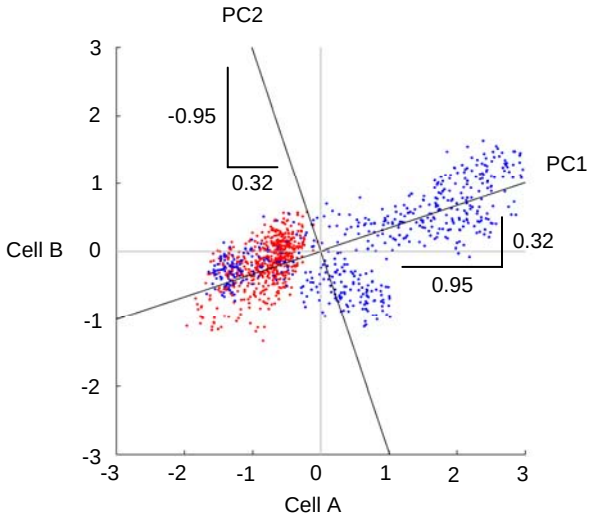


PC1 is always aligned along the direction of maximal covariance. PC2 is aligned along the direction of next maximal covariance, with the important constraint that it remain orthogonal to PC1. (The scatter plot was centered about the origin before the rotation).

What we call the PCs are actually the slopes of the new directions relative to the original axes. We choose to display them as bar graphs.



Each value in the bar graph is the contribution of each of the original axes to the new directions. When these values are combined, they are exactly the slopes of the new directions.

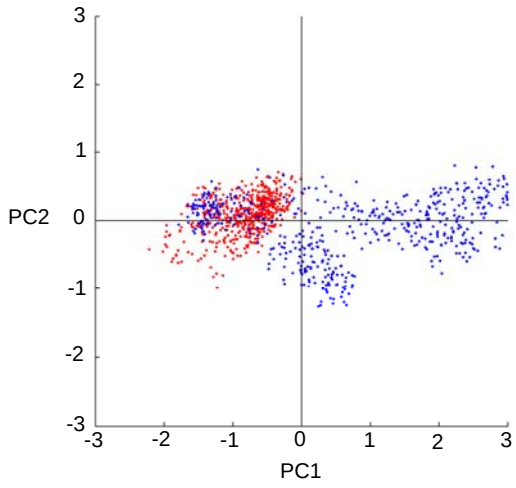


Note how PC1 has a larger value for cell A than cell B. This is because the direction of PC1 is oriented closer to the direction of cell A than cell B. This is what we mean when we refer to neurons with large magnitude contributions to a PC. Another way to express

this is that the PCs are linear combinations of the original directions. A neuron that does not influence the slope would then have zero contribution to the PC.

We hope that the reader can imagine how this generalizes to N dimensions. Each PC from an N-dimensional dataset contains the contributions of each of the N neurons. Also, by definition, PCA returns N PCs (N = 2 in the case above). When the datasets are of high dimensionality, the first few PCs can capture most of the overall variance. Suppose we record from 100 neurons (N = 100). PCA returns 100 PCs. The first 3 PCs (PC1, PC2, and PC3) might account for 90% of the variance in the 100-dimensional scatter plot. The remaining 10% of the variance is spread out over the remaining 97 PCs. Since each of these PCs explain little variance, they are discarded. Going back to our 2-dimensional example, PC1 explained 90% of the overall variance and PC2 explained 10%.

One important result from our analysis is not only that the first several PCs explain most of the variance, but that the new directions help separate the swimming from the crawling trials. We can see this by looking at the data in the rotated coordinate system.



Notice how PC1 separates the swimming and crawling data points. PC2, on the other hand, does not separate the behaviors well. By discarding PC2, we can reduce the dimensionality of the space, but still retain most of the interesting dynamics along PC1.

We now return to the idea of the LDA. Instead of performing the LDA on the raw data, we estimate LDs in a reduced PC space (to avoid overfitting). The slopes of these lines describe the relative contribution of each of the PCs to the discrimination. So the LD is a linear combination of PCs, each of which is a linear combination of neurons. Therefore, neurons contributing strongly to PCs that help separate swimming trials from crawling trials are candidate decision-making neurons. The issue of how many PCs and the bin width we use to estimate the LDs is addressed in Fig. S1.

We have attempted to graphically demonstrate the mechanics of LDA and PCA. There are many good texts that cover the mathematical background (*S1, S2*). Also, an excellent tutorial on PCA can be found at (*S3*).

Supplemental References

S1. Mardia K, Kent J, Bibby J, Multivariate Analysis (Academic Press, San Diego, CA, 1980).
S2. Duda RO, Hart PE, Stork DG, Pattern Classification (Wiley-Interscience, New York, ed. 2, 2000).
S3. Shlens J (2001) A Tutorial on Principal Component Analysis: Derivation, Discussion and Singular Value Decomposition, (http://www.snl.salk.edu/~shlens/pub/notes/pca.pdf)